Contents

An application of global bifurcation to the existence of nonnegative solutions of nonlinear Sturm-Liouville problem

Jacek Gulgowski
Institute of Mathematics
Uniwersity of Gdańsk
ul. Wita Stwosza 57, 80-952 Gdańsk

Abstract

In this paper we apply the global bifuraction theorem to the proof of the existence of nonnegative solutions of boundary value problems

$$(*) \qquad \begin{cases} u'' + \varphi(t, u(t), u'(t)) = 0 & \text{for } t \in (a,b) \\ l(u) = 0, \end{cases}$$

where $\varphi : [a,b] \times \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ and $l : C^1([a,b], \mathbb{R}^k) \to \mathbb{R}^k \times \mathbb{R}^k$ are continuous, and l represents Sturm-Liouville boundary conditions.

With a problem (*) we will associate the map

$$f : (0, +\infty) \times C^1([a,b], \mathbb{R}^k) \to C^1([a,b], \mathbb{R}^k)$$

such that if $f(1, u) = 0$ then u is solution of $(*)$.

The existence of $(1, u) \in (0, +\infty) \times C^1([a,b], \mathbb{R}^k)$, such that $f(1, u) = 0$, will be shown by proving the existence of connected component C of the set of nontrivial zeros of f, such that its projection to the interval $(0, +\infty)$ contains numbers $\lambda_1 \in (0,1)$ and $\lambda_2 \in (1, +\infty)$.

In the last section we will give an example referring to well known existence theorems for second order boundary value problems, for ordinary differential equations of second order.

1 Existence theorems

In this paper we will need the following notations. For $x = (x_1, ..., x_k) \in \mathbb{R}^k$ we write $|x| = \sum_{i=1}^k |x_i|$. We call x nonnegative (and write $x \geq 0$) when $x_j \geq 0$ for $j = 1, ..., k$. Let $p : \mathbb{R}^k \to \mathbb{R}^k$ be given by $p(x_1, ..., x_k) = (|x_1|, ..., |x_k|)$. Let $\| \cdot \|_0$ be the supremum norm in $C[a,b]$ and $\| \cdot \|_k$ be the norm in $C^1([a,b], \mathbb{R}^k)$ given by $\|u\|_k = \sum_{i=1}^k (\|u_i\|_0 + \|u_i'\|_0)$ for $u = (u_1, ..., u_k) \in C^1([a,b], \mathbb{R}^k)$.

In this section we will give the sufficient conditions for the existence of the nonnegative solution of the boundary value problems

$$(1.1) \qquad \begin{cases} u''(t) + \varphi(t, u(t), u'(t)) = 0 & \text{for } t \in (a, b) \\ l(u) = 0, \end{cases}$$

where $\varphi : [a, b] \times \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ is continuous and $l : C^1([a, b], \mathbb{R}^k) \to \mathbb{R}^k \times \mathbb{R}^k$ is given by $l(u_1, ..., u_k) = (l_1(u_1), ..., l_k(u_k))$, where

$$l_j(u_j) = (u_j(a) \sin \alpha_j - u_j'(a) \cos \alpha_j, u_j(b) \sin \beta_j + u_j'(b) \cos \beta_j),$$

and $\alpha_j, \beta_j \in [0, \frac{\pi}{2}], (j = 1, ..., k)$.

Before we state the existence theorems, we must refer to some of the spectral properties of the linear problem

$$(1.2) \qquad \begin{cases} u''(t) + \lambda u(t) = 0 & \text{for } t \in (a, b) \\ l(u) = 0. \end{cases}$$

It is obvious that $\mu \in \mathbb{R}$ is the eigenvalue of (1.2) if and only if there exists $j \in \{1, ..., k\}$ such that μ is the eigenvalue of the scalar problem

$$(1.2)_j \qquad \begin{cases} u_j''(t) + \lambda u_j(t) = 0 & \text{for } t \in (a, b) \\ l_j(u_j) = 0. \end{cases}$$

It is well known (cf [H, CL]), that there exists exactly one eigenvalue $\mu_j \in \mathbb{R}$ of $(1.2)_j$, for which there exists the eigenvector v_{μ_j}, such that $v_{\mu_j}(t) > 0$ for $t \in (a, b)$, and then $\mu_j \geq 0$. Let us observe that then $u_{\mu_j} = (0, ..., v_{\mu_j}, ...0)$ is the eigenvector of (1.2) associated with eigenvalue μ_j.

The set of eigenvalues of (1.2), for which there exists nonnegative eigenvector, is nonempty and contains at most k-elements. Let us denote those eigenvalues by $\mu_1 < ... < \mu_N$, $N \leq k$.

Before we state the theorems let us assume that continuous map $\varphi : [a, b] \times \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ satisfies two conditions

$$(1.3) \qquad \exists_{w_0, w_1, w_2 \in \mathbf{R}} \forall_{t \in [a,b]} \forall_{x \in \mathbf{R}^k} \forall_{y \in \mathbf{R}^k} |\varphi(t, x, y)| \leq w_0 + w_1|x| + w_2|y|;$$

$$(1.4) \qquad \forall_{t \in [a,b]} \forall_{y \in \mathbf{R}^k} \forall_{x \in \partial[0,+\infty)^k} \varphi(t, x, y) \geq 0;$$

Theorem 1 *Let $A, B \in \mathbb{R}$ be real constants satisfying $A < \mu_1 < \mu_N < B$ and $w \in \mathbb{R}^k$. Let continuous $\varphi : [a, b] \times \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ satisfies* (1.3), (1.4) *and*

$$(1.5) \qquad \exists_{r_0 > 0} \forall_{t \in [a,b]} \forall_{x \in [0,+\infty)^k} \forall_{y \in \mathbf{R}^k} |x| + |y| \leq r_0 \Rightarrow \varphi(t, x, y) \leq Ax;$$

$$(1.6) \qquad \exists_{R_0 > 0} \forall_{t \in [a,b]} \forall_{x \in [0,+\infty)^k} \forall_{y \in \mathbf{R}^k} |x| + |y| \geq R_0 \Rightarrow Bx + w \leq \varphi(t, x, y).$$

Then there exists nonnegative solution of Sturm-Liouville problem (1.1).

Theorem 2 *Let $A, B \in \mathbb{R}$ be real constants satisfying $A < \mu_1 < \mu_N < B$ and $w \in \mathbb{R}^k$. Let continuous $\varphi : [a,b] \times \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ satisfies* (1.3), (1.4) *and*

$$\exists_{r_0>0}\forall_{t\in[a,b]}\forall_{x\in[0,+\infty)^k}\forall_{y\in\mathbf{R}^k}|x|+|y|\leq r_0 \Rightarrow \varphi(t,x,y)\geq Bx; \tag{1.7}$$

$$\exists_{R_0>0}\forall_{t\in[a,b]}\forall_{x\in[0,+\infty)^k}\forall_{y\in\mathbf{R}^k}|x|+|y|\geq R_0 \Rightarrow \varphi(t,x,y)\leq w+Ax. \tag{1.8}$$

Then there exists the nonnegative solution of Sturm-Liouville problem (1.1).

The main tool used in this paper is the global bifurcation theorem given below; which is the consequence of the generalization of the Rabinowitz global bifurcation alternative (cf. [R], [CH], [G]). We need some notation to state the theorem. Let $F : (0,+\infty) \times C^1([a,b],\mathbb{R}^k) \to C^1([a,b],\mathbb{R}^k)$ be a completely continuous map such that $F(\cdot,0) = 0$ and let $f : (0,+\infty) \times C^1([a,b],\mathbb{R}^k) \to C^1([a,b],\mathbb{R}^k)$ be given by $f(\lambda,u) = u - F(\lambda,u)$.

The point $(\lambda_0, 0) \in (0,+\infty) \times C^1([a,b],\mathbb{R}^k)$ is a *bifurcation point* of the map f if for all open $U \subset (0,+\infty) \times C^1([a,b],\mathbb{R}^k)$ satisfying $(\lambda_0,0) \in U$ there exists $(\lambda,u) \in U$, such that $u \neq 0$ and $f(\lambda,u) = 0$. We will denote the set of all bifurcation points of f by $\mathcal{B}_f$. Let $\mathcal{R}_f \subset (0,+\infty) \times C^1([a,b],\mathbb{R}^k)$ be a closure of the *set of nontrivial solutions of the equation* $f(\lambda,u) = 0$, i.e.

$$\mathcal{R}_f = \overline{\{(\lambda,u) \in (0,+\infty) \times C^1([a,b],\mathbb{R}^k) : f(\lambda,u) = 0 \wedge u \neq 0\}}.$$

For each $(\mu_0, 0) \in \mathcal{B}_f$ being the isolated point in the set $\mathcal{B}_f$, we may define *bifurcation index of f in μ_0* by

$$s[f,\mu_0] = \deg(f(\mu_0+\varepsilon_0,\cdot),K(0,r),0) - \deg(f(\mu_0-\varepsilon_0,\cdot),K(0,r),0),$$

for $\varepsilon_0 > 0$ and $r > 0$ sufficiently small.

Theorem A *If $\mathcal{B}_f \subset \bigcup_{j=1}^{N}\{(\mu_j,0)\}$, $s[f,\mu_1] \neq 0$ and $s[f,\mu_j] = 0$ for $j = 2,...,N$, then there exists a connected component $C \subset \mathcal{R}_f$ such that $(\mu_1,0) \in C$ and C is not compact.* $\square$

2 Global bifurcation of nonlinear boundary value problems

It is well known that with the boundary value problem

$$\begin{cases} u''(t) + \mu u(t) + h(t) = 0 \\ l(u) = 0, \end{cases} \tag{2.1}$$

where μ is not the eigenvalue of the linear problem (1.2), we may associate the continuous map $T_\mu : C([a,b],\mathbb{R}^k) \to C^1([a,b],\mathbb{R}^k)$, such that $T_\mu(h) = u$ iff $u \in C^2([a,b],\mathbb{R}^k)$ is solution of (2.1). This map has the following properties (cf. [H, CL, P])

(2.2) For continuous $\Phi : (0,+\infty) \times C^1([a,b],\mathbb{R}^k) \to C([a,b],\mathbb{R}^k)$ such that for bounded sets $B \subset (0,+\infty) \times C^1([a,b],\mathbb{R}^k)$ the set $\Phi(B) \subset C([a,b],\mathbb{R}^k)$ is bounded, the superposition $T_\mu \circ \Phi : (0,+\infty) \times C^1([a,b],\mathbb{R}^k) \to C^1([a,b],\mathbb{R}^k)$ is completely continuous.

(2.3) For $u,v \in C([a,b],\mathbb{R}^k)$ we have $\langle T_\mu u, v\rangle = \langle u, T_\mu v\rangle$, where
$\langle u,v\rangle = \int\limits_a^b \sum\limits_{i=1}^k u_i(t)v_i(t)dt.$

(2.4) (maximum principle) Let $\mu \le 0$ and $u,h \in C([a,b],\mathbb{R}^k)$ satisfy $u = T_\mu h$ and $h \ge 0$. Then we have $u \ge 0$.

Now let us move to the calculation of bifurcation index for the specific map

Lemma 2.5 *Assume that $m > 0$, $\mu \le 0$, $\mu < \alpha_j^2 + \beta_j^2$ for $j = 1,...,k$. Let $\Phi : (0,+\infty) \times C^1([a,b],\mathbb{R}^k) \to C([a,b],\mathbb{R}^k)$ be continuous and there exists $r_0 > 0$, such that $\Phi(\lambda,u) = \lambda m p(u)$ for $\lambda \in (0,+\infty)$ and $\|u\|_k \le r_0$. Assume that $f : (0,+\infty) \times C^1([a,b],\mathbb{R}^k) \to C^1([a,b],\mathbb{R}^k)$ is given by $f(\lambda,u) = u - T_\mu\Phi(\lambda,u)$. Then $\mathcal{B}_f = \{(\frac{\mu_j-\mu}{m},0) | j = 1,...,N\}$ and $s[f,\frac{\mu_1-\mu}{m}] = -1$, $s[f,\frac{\mu_j-\mu}{m}] = 0$ for $j = 2,...,N$.*

Proof. Let us observe that if $f(\lambda,u) = 0$ for $\|u\|_k \le r_0$, then by (2.4) we have $u \ge 0$ and $p(u) = u$. Hence

$$\begin{cases} u''(t) + \mu u(t) + \lambda m u(t) = 0 & \text{for } t \in (a,b) \\ l(u) = 0, \end{cases}$$

and if $u \ne 0$, then there exists $j \in \{1,...,N\}$, such that $m\lambda + \mu = \mu_j$. This implies that $\mathcal{B}_f \subset \{(\frac{\mu_j-\mu}{m},0) | j = 1,...,N\}$. On the other hand $(\frac{\mu_j-\mu}{m},0) \in \mathcal{B}_f$ because for $r \in (0,r_0]$ the pair $(\frac{\mu_j-\mu}{m}, r\frac{u_{\mu_j}}{\|u_{\mu_j}\|_k})$ is the solution of $f(\lambda,u) = 0$.

Let us now calculate the bifurcation indices $s[f,\frac{\mu_j-\mu}{m}]$, $j = 1,...,N$. Let $\lambda_0 \in (0,\frac{\mu_1-\mu}{m})$ be fixed. We will show that $f(\lambda_0,\cdot) : \overline{K(0,r_0)} \to C^1([a,b],\mathbb{R}^k)$ may be joined by homotopy with the identity map.

Let the homotopy be given by $h(t,u) = u - m\lambda_0 t T_\mu p(u)$. Let us observe that $(\lambda_0 t, 0) \notin \mathcal{B}_f$ for $t \in [0,1]$. That is why we have no nontrivial zeros of $h(t,u) = 0$. Hence by homotopy property of topological degree we have $\deg(f(\lambda_0,\cdot), K(0,r_0), 0) = 1$.

Assume now that $\lambda_0 > \frac{\mu_1-\mu}{m}$ and $\lambda_0 \ne \frac{\mu_j-\mu}{m}$ for $j = 2,...,N$. We will show that for such λ_0 the map $f(\lambda_0,\cdot)$ may be joined by homotopy on $\overline{K(0,r_0)}$ with $f_1 : \overline{K(0,r_0)} \to C^1([a,b],\mathbb{R}^k)$ given by

$$f_1(u) = f(\lambda_0,u) - u_{\mu_1},$$

where (μ_1, u_{μ_1}) is the solution of (1.2), such that $u_{\mu_1} \geq 0$, $u_{\mu_1} \neq 0$. Let us first observe that $T_\mu u_{\mu_1} = \frac{1}{\mu_1 - \mu} u_{\mu_1}$.

The homotopy $h : [0,1] \times \overline{K(0, r_0)} \to C^1([a,b], \mathbb{R}^k)$ is given by

$$h(\tau, u) = f(\lambda_0, u) - \tau u_{\mu_1}.$$

Assume now that for $\|u\|_k \leq r_0$ and $\tau \in (0,1]$ there is $h(\tau, u) = 0$, and

$$u - \lambda_0 m T_\mu p(u) - \tau u_{\mu_1} = 0$$

$$u - T_\mu(\lambda_0 m p(u) + \tau(\mu_1 - \mu) u_{\mu_1}) = 0$$

Because $\lambda_0 m p(u) + \tau(\mu_1 - \mu) u_{\mu_1} \geq 0$, by (2.4) there is $u \geq 0$, and $p(u) = u$. Since we have $u = \lambda_0 m T_\mu u + \tau u_{\mu_1}$ and also

$$\langle u, u_{\mu_1} \rangle = \lambda_0 m \langle T_\mu u, u_{\mu_1} \rangle + \tau \langle u_{\mu_1}, u_{\mu_1} \rangle = \lambda_0 m \langle u, T_\mu u_{\mu_1} \rangle + \tau \langle u_{\mu_1}, u_{\mu_1} \rangle =$$

$$= \frac{\lambda_0 m}{\mu_1 - \mu} \langle u, T_\mu u_{\mu_1} \rangle + \tau \langle u_{\mu_1}, u_{\mu_1} \rangle.$$

That is why

$$\frac{\mu_1 - \mu - m\lambda_0}{\mu_1 - \mu} \langle u, u_{\mu_1} \rangle = \tau \langle u_{\mu_1}, u_{\mu_1} \rangle > 0.$$

Because $u_{\mu_1} \geq 0$ and $u \geq 0$, it must be also $\mu_1 - \mu > m\lambda_0$, what contradicts the assumption $\lambda_0 > \frac{\mu_1 - \mu}{m}$.

If $\tau = 0$, then $h(\tau, \cdot) = f(\lambda_0, \cdot)$, and

$$h(0, u) = 0 \Leftrightarrow f(\lambda_0, u) = 0.$$

Because $\lambda_0 \neq \frac{\mu_j - \mu}{m}$ for $j = 1, ..., N$,

$$f(\lambda_0, u) = 0 \Rightarrow u = 0.$$

Hence the homotopy h nas no nontrivial zeroes. Also $h(1, \cdot)$ has no zeroes at all and that is why, by the homotopy property of topological degree,

$$\deg(f(\lambda_0, \cdot), K(0, r_0), 0) = 0.$$

Hence we have $s[f, \frac{\mu_1 - \mu}{m}] = -1$, and $s[f, \frac{\mu_j - \mu}{m}] = 0$ for $j = 2, ..., N$. $\square$

Below we have the consequence of the lemma 2.5 and theorem A.

Corollary 2.6 *Let $f : (0, +\infty) \times C^1([a,b], \mathbb{R}^k) \to C^1([a,b], \mathbb{R}^k)$ satisfies assumptions of lemma* 2.5. *Then there exists noncompact component $C \subset \mathcal{R}_f$ such that* $(\frac{\mu_1 - \mu}{m}, 0) \in C$. $\square$

At the end of this section we will prove a few lemmas, which will be useful in the next two sections.

First let us make the observation that for any $v \in C^1([a,b],\mathbb{R}^k)$ such that $v \geq 0, v \neq 0$ there exists such eigenvalue μ_j $(j = 1,...,N)$ and corresponding eigenvector $u_{\mu_j} \geq 0$ that $\langle v, u_{\mu_j} \rangle > 0$.

Lemma 2.7 *Let $M > 0$ be fixed. Assume that $\varphi : [a,b] \times \mathbb{R}^k \times \mathbb{R}^k \times (0,M) \to \mathbb{R}^k$ is continuous and satsifes* (1.3) *uniformly with respect to $\lambda \in (0,M)$, and $\Phi : (0,+\infty) \times C^1([a,b],\mathbb{R}^k) \to C([a,b],\mathbb{R}^k)$ is Niemytskii operator associated with ϕ, given by $[\Phi(\lambda,u)](t) = \varphi(t,u(t),u'(t),\lambda)$.*

Assume additionally that $\{(\lambda_n,u_n)\} \subset (0,M) \times C^1([a,b],\mathbb{R}^k)$ is the sequence of solutions of equation $u - T_\mu\Phi(\lambda,u) = 0$, such that $\|u_n\|_k \to +\infty$, $u_n \geq 0$ and $\lambda_n \to \lambda_0 \in \mathbb{R}$. Then we have two theses:

(2.7.1) *If there exist constants $R_0 > 0$, $B > 0$, $\mu \leq 0$ and $w \in \mathbb{R}^k$ such that*

$$\forall_{t\in[a,b]}\forall_{x\in[0,+\infty)^k}\forall_{y\in\mathbf{R}^k}\forall_{\lambda\in(0,M)}|x| + |y| \geq R_0 \Rightarrow \varphi(t,x,y,\lambda) \geq \lambda Bx + w,$$

then

$$\lambda_0 \in \left[0, \frac{\mu_N - \mu}{B}\right].$$

(2.7.2) *If there exist constants $R_0 > 0$, $A > 0$, $\mu \leq 0$ and $w \in \mathbb{R}^k$ such that*

$$\forall_{t\in[a,b]}\forall_{x\in[0,+\infty)^k}\forall_{y\in\mathbf{R}^k}\forall_{\lambda\in(0,M)}|x| + |y| \geq R_0 \Rightarrow \varphi(t,x,y,\lambda) \leq \lambda Ax + w,$$

then

$$\lambda_0 \in \left[\frac{\mu_1 - \mu}{A}, +\infty\right).$$

Proof. Assume that $\|u_n\|_k \geq R_0$. Then

$$\frac{|\varphi(t,u_n(t),u_n'(t),\lambda)|}{\|u_n\|_k} \leq \frac{w_0}{R_0} + w_1 + w_2 \text{ for } t \in [a,b].$$

for $n \in \mathbb{N}$, hence by (2.2) the sequence $\left\{T_\mu\left(\frac{\Phi(\lambda,u_n)}{\|u_n\|_k}\right)\right\}$ contains convergent subsequence. Because $v_n = T_\mu\left(\frac{\Phi(\lambda_n,u_n)}{\|u_n\|_k}\right)$, where $v_n = \frac{u_n}{\|u_n\|_k}$, there exists the subsequence $\{v_{\psi(n)}\} \subset \{v_n\}$ converging to $v_0 \in C^1([a,b],\mathbb{R}^k)$, $\|v_0\|_k = 1$, $v_0 \geq 0$.

There is no loss in generality in assuming that $v_n \to v_0$

Proof 2.7.1. Let us observe that there exists constant $\gamma \in \mathbb{R}^k$ such that for $u \in C^1([a,b],\mathbb{R}^k)$, $u \geq 0$, $\lambda \in (0,M)$ there is

$$\varphi(t,u(t),u'(t),\lambda) \geq \lambda Bu(t) + \gamma,$$

for $t \in [a,b]$.

Then for solution (μ_j, u_{μ_j}) of (1.2), such that $\langle v_0, u_{\mu_j}\rangle > 0$ we have

$$\langle v_n, u_{\mu_j}\rangle = \langle T_\mu \frac{\Phi(\lambda_n, u_n)}{\|u_n\|_k}, u_{\mu_j}\rangle = \langle \frac{\Phi(\lambda_n, u_n)}{\|u_n\|_k}, T_\mu u_{\mu_j}\rangle \geq$$

$$\geq \frac{\lambda_n B}{\mu_j - \mu}\langle v_n, u_{\mu_j}\rangle + \frac{\gamma}{\|u_n\|_k}\int_a^b u_{\mu_j}(t)dt.$$

Letting $n \to +\infty$ we have

$$\langle v_0, u_{\mu_j}\rangle \geq \frac{\lambda_0 B}{\mu_j - \mu}\langle v_0, u_{\mu_j}\rangle \geq \frac{\lambda_0 B}{\mu_N - \mu}\langle v_0, u_{\mu_j}\rangle.$$

Because $(\mu_N - \mu - \lambda_0 B)\langle v_0, u_{\mu_j}\rangle \geq 0$ and $\langle v_0, u_{\mu_j}\rangle > 0$ there is $\lambda_0 \leq \frac{\mu_N - \mu}{B}$, what finishes the proof.

Proof 2.7.2. Let us observe that there exists a constant $\gamma \in \mathbb{R}^k$, such that for $u \geq 0$, $u \in C^1([a,b], \mathbb{R}^k)$, $\lambda \in (0, M)$ we have

$$\varphi(t, u(t), u'(t), \lambda) \leq \lambda A u(t) + \gamma$$

for $t \in [a,b]$.

For such solution (μ_j, u_{μ_j}) of (1.2), that $\langle v_0, u_{\mu_j}\rangle > 0$ we have

$$\langle v_n, u_{\mu_j}\rangle = \langle T_\mu \frac{\Phi(\lambda_n, u_n)}{\|u_n\|_k}, u_{\mu_j}\rangle = \langle \frac{\Phi(\lambda_n, u_n)}{\|u_n\|_k}, T_\mu u_{\mu_j}\rangle \leq$$

$$\leq \frac{\lambda_n A}{\mu_j - \mu}\langle v_n, u_{\mu_j}\rangle + \frac{\gamma}{\|u_n\|_k}\int_a^b u_{\mu_j}(t)dt.$$

Letting $n \to +\infty$ we have

$$\langle v_0, u_{\mu_j}\rangle \leq \frac{\lambda_0 A}{\mu_j - \mu}\langle v_0, u_{\mu_j}\rangle.$$

Because $(\mu_j - \mu - \lambda_0 A)\langle v_0, u_{\mu_j}\rangle \leq 0$ and $\langle v_0, v_{\mu_j}\rangle > 0$, we have $\lambda_0 \geq \frac{\mu_j - \mu}{A} \geq \frac{\mu_1 - \mu}{A}$, what finishes the proof. □

Lemma 2.8 *Suppose that continuous* $\Phi : (0, +\infty) \times C^1([a,b], \mathbb{R}^k) \to C([a,b], \mathbb{R}^k)$ *satisfies condition*

$$\forall_{\lambda \in (0,+\infty)} \forall_{u \in C^1([a,b], \mathbf{R}^k)} \forall_{t \in [a,b]} \forall_{j \in \{1,\dots,k\}} u_j(t) < 0 \Rightarrow [\Phi(\lambda, u)]_j(t) \geq 0. \tag{2.9}$$

Then

$$\forall_{u \in C^1([a,b], \mathbf{R}^k)} u = T_\mu \Phi(\lambda, u) \Rightarrow u \geq 0.$$

Proof. Let $u \in C^1([a,b],\mathbb{R}^k)$, $u = (u_1,...,u_k)$. Let us fix any $j \in \{1,...,k\}$. Let $t_0 \in [a,b]$ be a number such that $u_j(t_0) = \inf_{t\in[a,b]} u_j(t)$. Assume contrary to our claim that $u_j(t_0) < 0$. Because u_j is continuous there exists an interval $[c_0,d_0] \subset [a,b]$, such that for $t \in [c_0,d_0]$ $u_j(t) \le 0$. Let $[c,b] \subset [a,b]$ be the maximal interval having this property.

Let us observe that $u_j''(t) = -\mu u_j(t) - [\Phi(\lambda,u)]_j(t) \le 0$ for $t \in [c,d]$. This is why $u_j'(t) \ge u_j'(t_0) = 0$ for $t \in [c,t_0]$ and $u_j'(t) \le u_j'(t_0) = 0$ for $t \in [t_0,d]$. This implies $u_j(c) \le u_j(t_0)$ and $u_j(d) \le u_j(t_0)$.

Hence $u_j(c) < 0$ and from the selection of c it must be $c = a$ and, similarly, $b = d$.

That is why $[\Phi(\lambda,u)]_j \ge 0$, and $u_j''(t) = -[\Phi(\lambda,u)]_j(t) \le 0$, $l(u) = 0$ and by (2.4) we have $u_j \ge 0$ what contradicts the assumption $u_j(t_0) < 0$. □

3 Proof of theorem 1

Let $H : \mathbb{R}^1 \to \mathbb{R}^1$ be given by

$$H(x) = \begin{cases} x & \text{for } x \ge 0 \\ 0 & \text{for } x < 0. \end{cases}$$

Let $r_k : \mathbb{R}^k \to [0,+\infty)^k$ be a continuous retraction given by

$$r_k(x_1,...,x_k) = (H(x_1),...,H(x_2)).$$

Let $A, B \in \mathbb{R}$ be constants satisfying assumptions of theorem 1. Let us fix $\mu < \min\{0, A\}$ and denote $A_\mu = A - \mu > 0$ and $B_\mu = B - \mu$.

For continuous map $\varphi : [a,b] \times \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ we will define the continuous $\tilde{\varphi} : [a,b] \times \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ by

$$\tilde{\varphi}(t,x,y) = \begin{cases} \varphi(t,x,y) - \mu x & \text{for } x \in [0,+\infty)^k \\ A_\mu r_k(-x) + \varphi(t,r_k(x),y) - \mu r_k(x) & \text{for } x \notin [0,+\infty)^k. \end{cases}$$

Let us observe that $\tilde{\varphi}(t,x,y) \ge 0$ for $x \in \mathbb{R}^k \setminus (0,+\infty)^k$.

Let $r_0 > 0$ be a constant given in (1.5) and let

$$U_1 = K(0,r_0) \qquad \text{and} \qquad U_2 = C^1([a,b],\mathbb{R}^k) \setminus \overline{K(0,\frac{r_0}{2})}.$$

Sets U_1, U_2 form the open cover of $C^1([a,b],\mathbb{R}^k)$. Let $\{\eta_1,\eta_2\}$ be the continuous partition of unity associated with cover $\{U_1,U_2\}$.

Let ν_1,ν_2 be real positive numbers such that $\frac{\mu_N - \mu}{A_\mu} < \nu_1 < \nu_2$. Let ψ_1,ψ_2 be a continuous partition of unity associated with open cover $\{(0,\nu_2),(\nu_1,+\infty)\}$ of interval $(0,+\infty)$.

The maps $\Phi_1, \Phi_2, \Phi : (0, +\infty) \times C^1([a, b], \mathbb{R}^k) \to C([a, b], \mathbb{R}^k)$ are given by:

$$\Phi_1(\lambda, u) = \lambda\eta_1(u)A_\mu p(u) + \lambda\eta_2(u)\tilde{\Phi}(u),$$

where $\tilde{\Phi} : C^1([a, b], \mathbb{R}^k) \to C([a, b], \mathbb{R}^k)$ is Niemytskii operator for $\tilde{\varphi}$;

$$\Phi_2(\lambda, u) = \lambda\eta_1(u)A_\mu p(u) + \lambda\eta_2(u)B_\mu p(u);$$

$$\Phi(\lambda, u) = \psi_1(\lambda)\Phi_1(\lambda, u) + \psi_2(\lambda)\Phi_2(\lambda, u).$$

Let us observe that for $\lambda > \nu_2$ there is $\Phi(\lambda, u) = \Phi_2(\lambda, u)$, and for $\lambda < \nu_1$ we have $\Phi(\lambda, u) = \Phi_1(\lambda, u)$. It may be seen also that for $\|u\|_k \leq \frac{r_0}{2}$ and any $\lambda > 0$ it is $\Phi(\lambda, u) = \lambda A_\mu p(u)$.

Let $f : (0, +\infty) \times C^1([a, b], \mathbb{R}^k) \to C^1([a, b], \mathbb{R}^k)$ be given by

$$f(\lambda, u) = u - T_\mu(\Phi(\lambda, u)).$$

By (2.2) the map $[T_\mu \circ \Phi] : (0, +\infty) \times C^1([a, b], \mathbb{R}^k) \to C^1([a, b], \mathbb{R}^k)$ is completely continuous. By corollary 2.6 there exists a component $C \subset (0, +\infty) \times C^1([a, b], \mathbb{R}^k)$ of $\mathcal{R}_f$, such that $(\frac{\mu_1 - \mu}{A_\mu}, 0) \in C$ and C is not compact.

Let us observe that the map Φ satsifies (2.9). It is obvious considering that for $u_j(t_0) \notin [0, +\infty)^k$ there is $\tilde{\varphi}(t_0, u(t_0), u'(t_0)) \geq 0$. Hence by lemma 2.8 if $f(\lambda, u) = 0$, then $u \geq 0$.

The map f is such that, if $f(\lambda, u) = 0$, $\|u\|_k \geq r_0$ and $\lambda \in (0, \nu_1]$, then (λ, u) is the solution of boudary value problem

$$\text{(3.1)} \qquad \begin{cases} u''(t) + \mu u(t) + \lambda(\varphi(t, u(t), u'(t)) - \mu u(t)) = 0 & \text{for } t \in [a, b] \\ l(u) = 0 \end{cases}$$

such that $u \geq 0$.

We will show that there exists no nontrivial solution of $f(\lambda, u) = 0$ for $\lambda > \nu_2$. First observe that for $\lambda > \nu_2$ we have

$$\Phi(\lambda, u) = \Phi_2(\lambda, u) \geq \lambda A_\mu p(u).$$

Let us now assume, contrary to our claim, that for $(\lambda, u) \in (\nu_2, +\infty) \times C^1([a, b], \mathbb{R}^k)$ such that $f(\lambda, u) = 0$ we have, $u \geq 0$ and $u \neq 0$. For (μ_j, u_{μ_j}) being the solution of (1.2) and such that $\langle u, u_{\mu_j} \rangle > 0$, we have

$$\langle u, u_{\mu_j} \rangle = \langle T_\mu \Phi(\lambda, u), u_{\mu_j} \rangle = \langle \Phi(\lambda, u), T_\mu u_{\mu_j} \rangle =$$
$$= \frac{1}{\mu_j - \mu} \langle \Phi(\lambda, u), u_{\mu_j} \rangle \geq \frac{\lambda A_\mu}{\mu_j - \mu} \langle u, u_{\mu_j} \rangle$$

and

$$(\mu_j - \mu - \lambda A_\mu)\langle u, u_{\mu_j} \rangle \geq 0,$$

implying $\lambda \leq \frac{\mu_j - \mu}{A_\mu} < \nu_2$.

This contradiction proves that $C \subset (0, \nu_2] \times C^1([a,b], \mathbb{R}^k)$.

Now we are going to show that there exists $u \in C^1([a,b], \mathbb{R}^k)$, such that $(1, u) \in C$. Because $(\frac{\mu_1 - \mu}{A_\mu}, 0) \in C$, $\frac{\mu_1 - \mu}{A_\mu} > 1$, and C is connected it is enough to show that there exists $(\lambda^*, u) \in C$ such that $\lambda^* < 1$.

The component C is not compact hence there exists the sequence $\{(\lambda_n, u_n)\} \subset C$ such that $\lambda_n \to 0$ or $\|u_n\|_k \to +\infty$.

If $\lambda_n \to 0$ then of course there exists $(\lambda^*, u) \in C$ such that $\lambda^* < 1$. Thus it is enough to investigate the case of $\|u_n\|_k \to +\infty$. Without loss of generality we may assume that $\lambda_n \to \lambda_0 \in [0, \nu_2]$. Taking in the lemma 2.7 the constant $M = \nu_2$ we have, by (2.7.1),

$$\lambda_0 \in [0, \frac{\mu_N - \mu}{B_\mu}] \subset [0, 1),$$

That is why there exists $(\lambda^*, u) \in C$ such that $\lambda^* < 1$.

We have just shown that there exists $u \in C^1([a,b], \mathbb{R}^k)$, such that $(1, u) \in C$. We are going to prove that then we have $\|u\|_k > r_0$. Because $1 < \nu_1$ it is $\Phi(1, u) = \Phi_1(1, u)$.

Assume contrary to our claim that $\|u\|_k \leq r_0$. Then $\tilde{\Phi}(u) \leq A_\mu u$ and

$$\Phi_1(1, u)(t) = \eta_1(u) A_\mu u(t) + \eta_2(u) \tilde{\Phi}(u)(t) \leq A_\mu u(t)$$

and for appropriate (μ_j, u_{μ_j}) we have $\langle u, u_{\mu_j} \rangle > 0$ and

$$\langle u, u_{\mu_j} \rangle = \langle T_\mu \Phi(1, u), u_{\mu_j} \rangle = \frac{1}{\mu_j - \mu} \langle \Phi(1, u), u_{\mu_j} \rangle \leq \frac{A_\mu}{\mu_j - \mu} \langle u, u_{\mu_j} \rangle.$$

Then $A_\mu \geq \mu_j - \mu \geq \mu_1 - \mu$ what contradicts our assumption.

So we have $f(1, u) = 0$ and $\|u\|_k \geq r_0$. By (3.1) the function u is nonnegative solution of (1.1). $\square$

4 Proof of theorem 2

Let us define continuous map $\tilde{\varphi} : [a,b] \times \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ by

$$\tilde{\varphi}(t, x, y) = \begin{cases} \varphi(t, x, y) - \mu x & \text{for } x \in [0, +\infty)^k \\ B_\mu r_k(-x) + \varphi(t, r_k(x), y) - \mu r_k(x) & \text{for } x \notin [0, +\infty)^k. \end{cases}$$

Then $\tilde{\varphi}(t, x, y) \geq 0$ for $x \in \mathbb{R}^k \setminus (0, +\infty)^k$.

Let $r_0 > 0$ be a constant given in (1.7) and let

$$U_1 = K(0, r_0) \qquad \text{and} \qquad U_2 = C^1([a,b], \mathbb{R}^k) \setminus \overline{K(0, \frac{r_0}{2})}.$$

Sets U_1, U_2 form the open cover of the $C^1([a,b], \mathbb{R}^k)$. Let $\{\eta_1, \eta_2\}$ be a cotinuous partition of unity associtaed with cover $\{U_1, U_2\}$.

Let $\Phi : (0, +\infty) \times C^1([a, b], \mathbb{R}^k) \to C([a, b], \mathbb{R}^k)$ be given by

$$\Phi(\lambda, u) = \lambda\eta_1(u)B_\mu p(u) + \lambda\eta_2(u)\tilde{\Phi}(u),$$

where $\tilde{\Phi} : C^1([a, b], \mathbb{R}^k) \to C([a, b], \mathbb{R}^k)$ is Niemytskii operator for $\tilde{\varphi}$.

Let $f : (0, +\infty) \times C^1([a, b], \mathbb{R}^k) \to C^1([a, b], \mathbb{R}^k)$ be given by

$$f(\lambda, u) = u - T_\mu(\Phi(\lambda, u)).$$

Let us observe that the map $[T_\mu \circ \Phi] : (0, +\infty) \times C^1([a, b], \mathbb{R}^k) \to C^1([a, b], \mathbb{R}^k)$ is completely continuous.

By corollary 2.6 there exists connected component $C \subset (0, +\infty) \times C^1([a, b], \mathbb{R}^k)$ of $\mathcal{R}_f$, such that $(\frac{\mu_1-\mu}{B_\mu}, 0) \in C$ and C is not compact.

We will prove that the component C is unbounded. If, assuming the opposite, C was bounded, then because it is not compact, there would exist the sequence $\{(\lambda_n, u_n)\} \subset C$ such that $\lambda_n \to 0$ and $u_n \neq 0$. Then $u_n = T_\mu\Phi(\lambda_n, u_n)$ and u_n contains the subsequence converging to $u_0 \in C^1([a, b], \mathbb{R}^k)$. Letting $n \to +\infty$ we have $\Phi(\lambda_n, u_n) \to 0$ and $u_0 = 0$. Thus for almost all $n \in \mathbb{N}$ $\eta_2(u_n) = 0$ and $\eta_1(u_n) = 1$ and $u_n = \lambda_n B_\mu T_\mu p(u_n)$. If $u_n \neq 0$, then it must be $\lambda_n \in \{\frac{\mu_j-\mu}{B_\mu} : j \in 1, ..., N\}$. This contradicts $\lambda_n \to 0$ and proves that, the component C is unbounded.

Let us observe that Φ satisfies (2.9). Hence if $f(\lambda, u) = 0$, then, by lemma 2.8, $u \geq 0$.

The map f is such that, if $f(\lambda, u) = 0$, $\|u\|_k \geq r_0$, then (λ, u) satsifies

$$\begin{cases} u''(t) + \mu u(t) + \lambda(\varphi(t, u(t), u'(t)) - \mu u(t)) = 0 & \text{for } t \in [a, b] \\ l(u) = 0 \end{cases} \tag{4.1}$$

and $u \geq 0$.

Now we are going to show that there exists $u \in C^1([a, b], \mathbb{R}^k)$, such that $(1, u) \in C$.

Because $(\frac{\mu_1-\mu}{B_\mu}, 0) \in C$ and $\frac{\mu_1-\mu}{B_\mu} < 1$ and C is connected, it is enough to show that there exists $(\lambda^*, u) \in C$ such that $\lambda^* > 1$.

Because C is not bounded then there exists the sequence $\{(\lambda_n, u_n)\} \subset C$ such that $\lambda_n \to +\infty$ or $\|u_n\|_k \to +\infty$.

If $\lambda_n \to +\infty$, then there exists $(\lambda^*, u) \in C$ satisfying $\lambda^* > 1$. So it is enough to investigate the case of $\|u_n\|_k \to +\infty$ assuming, contrary to our claim, that $\{\lambda_n\} \subset (0, 1]$. We may assume that $\lambda_n \to \lambda_0 \in [0, 1]$. Taking in lemma 2.7 constant $M = 1$ we have, by (2.7.2),

$$\lambda_0 \in [\frac{\mu_1 - \mu}{A_\mu}, +\infty) \subset (1, +\infty),$$

The contradiction implies, that there exists $(\lambda^*, u) \in C$ such that $\lambda^* > 1$.

So we have shown that there exists $u \in C^1([a, b], \mathbb{R}^k)$, such that $(1, u) \in C$. Now we will prove that $\|u\|_k > r_0$.

Assume contrary to our claim that $\|u\|_k \leq r_0$. Then

$$\Phi(1,u)(t) \geq \eta_1(u)B_\mu u(t) + \eta_2(u)\tilde{\Phi}(u)(t) = B_\mu u(t)$$

and for appropriate (μ_j, u_{μ_j}) we have $\langle u, u_{\mu_j}\rangle > 0$. Then

$$\langle u, u_{\mu_j}\rangle = \langle T_\mu \Phi(1,u), u_{\mu_j}\rangle = \frac{1}{\mu_j - \mu}\langle \Phi(1,u), u_{\mu_j}\rangle \geq \frac{B_\mu}{\mu_j - \mu}\langle u, u_{\mu_j}\rangle.$$

So $B_\mu \leq \mu_j - \mu \leq \mu_N - \mu$, what contradicts our assumption.

Hence $f(1,u) = 0$ and $\|u\|_1 \geq r_0$. By (4.1) u is nonegative solution of (1.1). $\square$

5 The example

In [GGL] authors studied nonlinear boundary value problem for systems of differential equations of second order. Below we are going to give two of such results. In the sequel we denote $\langle x, y\rangle_k = \sum_{j=1}^k x_j y_j$, where $x = (x_1, ..., x_k), y = (y_1, ..., y_k) \in \mathbb{R}^k$.

Theorem 5.1 *Let $f : [0,1] \times \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ be continuous and assume*

(5.2) *There exists $M \geq 0$ such that $\langle x, f(t,x,y)\rangle_k \geq 0$, for all $|x| > M$ and all $x, y \in \mathbb{R}^k$ such that $\langle x, y\rangle_k = 0$.*

(5.3) *The equation can be listed in such a way that for the j-th equation there exist functions $A_j(t,x,y_1,...,y_{j-1}), B_j(t,x,y_1,...,y_{j-1}) \geq 0$ and bounded for $(t,x,y_1,...,y_{j-1})$ in bounded subsets of $[0,1] \times \{y \in \mathbb{R}^k : |y| \leq M\} \times \mathbb{R}^{j-1}$, such that*

$$|f_j(t,x,y)| \leq A_j(t,x,y_1,...,y_{j-1})y_j^2 + B_j(t,x,y_1,...,y_{j-1}).$$

Then there exists at least one solution of the problem

$$\begin{cases} u''(t) = f(t,u(t),u'(t)) & \text{for } t \in (0,1) \\ u(0) = u(1) = 0. \end{cases} \tag{5.4}$$

Theorem 5.5 *Suppose $f : [0,1] \times \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ is given by $f(t,x,y) = f_1(t,x,y) + f_2(t,x,y)$, where $f_1, f_2 : [0,1] \times \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ are continuous. Then there exists at least one solution to the Picard problem* (5.4) *provided:*

(5.6) $|f_2(t,x,y)| \leq B(1+|x|^\alpha+|y|^\beta)$, *where $B \in (0,+\infty)$, $0 \leq \alpha, \beta < 1$ are constants;*

(5.7) $\langle x, f_1(t,x,y)\rangle_k \geq 0$*;*

(5.8) *for (t,x) in bounded subsets of $[0,1] \times \mathbb{R}^k$ the function $f_1(t,x,y)$ is bounded.*

Below we will give the example of the Picard problem, which does not satisfy (5.2). Then we will show that for this example the righthand side of the equation may not be decomposed into the sum of functions satisfying (5.6) and (5.7).

Before we state the example, let us observe that the only eigenvalue of the Dirichlet problem

$$\begin{cases} u''(t) + \lambda u(t) = 0 \\ u(0) = u(1) = 0, \end{cases}$$

for which there exists the nonegative eigenvector, is $\mu_0 = \pi^2$.

Example 5.9 Let $\varphi : [0,1] \times \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$\varphi(t, x_1, x_2, y_1, y_2) =$$

$$= (B_{11}x_1 + B_{12}x_2, B_{21}x_1 + B_{22}x_2)p(|x| + |y|) + (A_1x_1, A_2x_2)q(|x| + |y|),$$

where $B_{11}, B_{22} > \pi^2$, $B_{12}, B_{21} \geq 0$, $0 \leq A_1, A_2 < \pi^2$; $p, q : [0, +\infty) \to [0, +\infty)$ are continuous and such that $p(0) = 1$, $\lim_{s\to+\infty} sp(s) = 0$, $q(0) = 0$ and $\lim_{s\to+\infty} q(s) = 1$.

Let us observe that φ satsifies the assumption (1.3). Also for $(x_1, x_2) \in [0, +\infty)^2$ we have $\varphi(t, x_1, x_2, y_1, y_2) \geq 0$; so (1.4) is satsified as well. Now we will show that φ satsifies (1.7) and (1.8), too.

For $|x| + |y|$ small enough we have

$$(B_{11}x_1 + B_{12}x_2)p(|x| + |y|) + A_1x_1q(|x| + |y|) \geq B_{11}x_1p(|x| + |y|) \geq \beta x_1 > \pi^2 x_1$$

$$(B_{21}x_1 + B_{22}x_2)p(|x| + |y|) + A_2x_2q(|x| + |y|) \geq B_{22}x_2p(|x| + |y|) \geq \beta x_2 > \pi^2 x_2$$

for the constant $\beta = \frac{1}{2}(\pi^2 + \min\{B_{11}, B_{22}\})$.

On the other hand, because functions $(B_{11}x_1 + B_{12}x_2)p(|x| + |y|)$ and $(B_{21}x_1 + B_{22}x_2)p(|x| + |y|)$ are limited, we have for $|x| + |y|$ large

$$(B_{11}x_1 + B_{12}x_2)p(|x| + |y|) + A_1x_1q(|x| + |y|) \leq \alpha x_1 + w_1 < \pi^2 x_1 + w_1$$

$$(B_{21}x_1 + B_{22}x_2)p(|x| + |y|) + A_2x_2q(|x| + |y|) \leq \alpha x_2 + w_1 < \pi^2 x_1 + w_2,$$

for the constant $\alpha = \frac{1}{2}(\pi^2 + \max\{A_1, A_2\})$.

That is why φ satsifies all assumptions of theorem 2. Now let $f : [0,1] \times \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ be given by $f(t, x, y) = -\varphi(t, x, y)$. Let us observe that if $x = (x_1, x_2) \in [0, +\infty)^2$, then for the large $|x|$ there is $q(|x| + |y|) > 0$ and

$$\langle x, f(t, x, y)\rangle_2 =$$

$$= -\Big(B_{11}x_1^2 + (B_{12} + B_{21})x_1x_2 + B_{22}x_2^2\Big)p(|x| + |y|) - (A_1x_1^2 + A_2x_2^2)q(|x| + |y|) < 0.$$

So condition (5.2) is not satsified for any $M \geq 0$.

Now let us take any decomposition $f(t, x, y) = f_1(t, x, y) + f_2(t, x, y)$ such that f_2 satsifies (5.6). Then $f_1 = f - f_2$ and let us take any $x_1 \in [0, +\infty)$, $x = (x_1, 0)$ and $y = (0, 0)$. So we have

$$\langle x, f_1(t, x, y)\rangle_2 = -B_{11}x_1^2p(|x_1|) - A_1x_1^2q(|x_1|) - \langle x, f_2(t, x, y)\rangle_2$$

Let us observe that

$$A_1 x_1^2 q(|x_1|) + \langle x, f_2(t,x,y)\rangle_2 \geq A_1 x_1^2 q(|x_1|) - x_1 B(1 + |x_1|^\alpha) > 0$$

for large x_1. That is why $\langle x, f_1(t,x,y)\rangle_2 < 0$ for some $(x,y) \in \mathbb{R}^2 \times \mathbb{R}^2$, so condition (5.7) is not satisfed.

As a conclusion let us observe that, by the theorem 2, there exists the nonnegative solution of (5.4), where $f = -\varphi$, whereas assumptions of theorems 5.1 and 5.5 are not satsified.

6 Bibliography

[CH] N-S. Chow, J.K. Hale, Methods of bifurcation theory, Springer Verlag, 1982.

[CL] E.Coddington, N. Levinson, Theory of Ordinary Differential Equations, McGraw--Hill, New York, 1955.

[G] J.Gulgowski, A global bifurcation theorem with applications to nonlinear Picard problems, *Nonlinear Analysis*, 41 (2000), 787-801.

[GGL] A. Granas, R.B. Guenther, J.W. Lee, Nonlinear boundary value problems for ordinary differential equations, *Dissertationes Mathematicae*, 244, Warszawa 1985

[H] P. Hartman, Ordinary differential equations, Birkhauser, Boston-Basel-Stuttgart, 1982

[PW] M.H. Protter, H.F.Weinberger, Maximum Principles in Differential Equations, Springer-Verlag, New York, 1984

[R] P. Rabinowitz, Some global results for nonlinear eigenvalue problems, *Journal of Functional Analysis*, **7**, 487-513, (1971).

SOLVABILITY OF QUASILINEAR ELLIPTIC SECOND ORDER DIFFERENTIAL EQUATIONS IN $\mathbb{R}^n$ WITHOUT CONDITION AT INFINITY

GENNADY I. LAPTEV

ABSTRACT. Solvability conditions for the equation

$$-\sum_{i=1}^{n} D_i A_i(x, Du) + A_0(x, u) = f(x), \qquad x \in \mathbf{R}^n$$

are considered in the whole space $\mathbf{R}^n$, $n \geq 2$. A solution $u(x)$ and the functions $f(x)$ and $A_i(x, \xi)$ for $i = 1, \dots, n$ and $A_0(x, u)$ may grow arbitrary as $|x| \to \infty$. These functions satisfy the standard conditions of the theory of monotone operators on the arguments $\xi \in \mathbf{R}^n$ and $u \in \mathbf{R}^1$. The method of monotone operators is developed and an existence theorem is proved for the solutions $u \in W^{1,p}_{\mathrm{loc}(\mathbf{R}^n) \cap L^q_{\mathrm{loc}(\mathbb{R}^n)}}$, where $q > p > 1$.

1. Posing the problem and statement of the result

The paper deals with the existence of solutions to quasilinear elliptic second order differential equations in the whole space $\mathbb{R}^n$, $n \geq 2$. The method of monotone operators is used in combination with the method of compact operators. These methods were developed in the 1960s by many authors (see [18, 12, 3]); they made it possible to study wide classes of higher-order partial differential equations of elliptic type in bounded domains or in unbounded domains under the condition that the solution belongs to an appropriate Sobolev space $W^{m,p}(\Omega)$. The aplications of these methods were summarized in monographs by Lions [11] and Skrypnik [16]. The development of monotone operators method is countinued at the present time (see for example [15, 6, 10]). The variation method for the equation $\Delta u = f(x, u)$ in the whole space $\mathbb{R}^n$ was considered in the monograph [9]. A number of articles are devoted to qualitative theory of positive solutions for elliptic equations in unbounded domains and in the whole $\mathbb{R}^n$ ($n \geq 2$), see [17, 13, 7, 8]. Existence theory of parabolic coercive equations in unbounded domains without conditions at infinity has been widely studied by many authors, see [1, 14, 5].

There are a few works for elliptic coercive equations in unbounded domains without conditions at infinity. Brezis in [2] studied semilinear equations of the form $-\Delta u + |u|^{q-1}u = f(x)$, $q > 0$, $x \in \mathbb{R}^n$. Gladkov in [4]considered the Dirichlet problem for the equation

$$-\sum_{i=1}^{n} (|u_{x_i}|^{\alpha} u_{x_i})_{x_i} + c(x)u = f(x), \quad x \in \Omega,$$

in unbounded domain Ω with a smooth compact boundary $\partial\Omega$. Here $\alpha > 0$, $c(x) \in L_{\mathrm{loc}^\infty(\mathbf{R}^n)}$, $c(x) \geq 0$, $f(x) \in L_{\mathrm{loc}^2(\mathbf{R}^n)}$.

This paper is devoted to the study of the solvability of equation

$$-\sum_{i=1}^{n} D_i A_i(x, Du) + A_0(x, u) = f(x), \quad x \in \mathbb{R}^n. \tag{1.1}$$

Here $D_i = \partial/\partial x_i$ $(i = 1, \dots, n)$, $Du = (D_1 u, \dots, D_n u)$.

Let us list our assumptions concerning the functions involved in (1.1). We assume that the functions $A_i(x, \xi)$ for $i = 1, \dots, n$ and $A_0(x, u)$ are defined for $x \in \mathbb{R}^n$, $\xi \in \mathbb{R}^n$, $u \in \mathbb{R}^1$ and satisfy the Carathéodory condition, that is, they are measurable with respect to $x \in \mathbb{R}^n$ for all $\xi \in \mathbb{R}^n$, $u \in \mathbb{R}^1$ and continuous in ξ, u for almost all $x \in \mathbb{R}^n$. Moreover, they are subject to the following constraints.

(1A) *Growth conditions.* For $i = 1, \dots, n$ we have

$$|A_i(x, \xi)| \leq a_1(x)|\xi|^{p-1} + b_1(x) \qquad (x, \xi \in \mathbb{R}^n),$$

where $p > 1$, $p + p' = pp'$, $a_1(x) \in L_{\mathrm{loc}^\infty(\mathbf{R}^n)}$, $b_1(x) \in L_{\mathrm{loc}^{p'}(\mathbf{R}^n)}$; futhermore

$$|A_0(x, u)| \leq a_0(x)|u|^{q-1} + b_0(x) \qquad (x \in \mathbb{R}^n, \; u \in \mathbb{R}^1),$$

where $q > p$, $q + q' = qq'$, $a_0(x) \in L_{\mathrm{loc}^\infty(\mathbf{R}^n)}$, $b_0(x) \in L_{\mathrm{loc}^{q'}(\mathbf{R}^n)}$.

(2A) *Monotonicity condition.* For almost all $x \in \mathbb{R}^n$ and all $\xi, \eta \in \mathbb{R}^n$ we have

$$\sum_{i=1}^{n} [A_i(x, \xi) - A_i(x, \eta)](\xi_i - \eta_i) > 0 \qquad (\xi \neq \eta),$$

where $\xi = (\xi_1, \dots, \xi_n)$ and $\eta = (\eta_1, \dots, \eta_n)$.

(3A) *Coercivity condition.* For almost all $x \in \mathbb{R}^n$ and all $\xi \in \mathbb{R}^n$ and $u \in \mathbb{R}^1$ we have

$$\sum_{i=1}^{n} A_i(x, \xi)\xi_i + A_0(x, u)u \geq a(x)|\xi|^p + b(x)|u|^q - g(x),$$

where $a(x), b(x) \in L_{\mathrm{loc}^\infty(\mathbf{R}^n)}$; $a(x), b(x) \geq c(R) > 0$ for $|x| \leq R$; $g(x) \in L_{\mathrm{loc}^1(\mathbf{R}^n)}$.

(1f) There exist functions $f_i(x) \in L_{\mathrm{loc}^{p'}(\mathbf{R}^n)}$ for $i = 1, \dots, n$ and $f_0(x) \in L_{\mathrm{loc}^{q'}(\mathbf{R}^n)}$ such that

$$f(x) = -\sum_{i=1}^{n} D_i f_i(x) + f_0(x)$$

in the sense of distributions.

Hereafter we use the traditional notations for spaces of summable functions. Let us recall them. Let Ω be a bounded domain in $\mathbb{R}^n$. All measurable functions $u(x)$, $x \in \Omega$, with finite norm

$$\|u\|^p_{L^p(\Omega)} = \int_\Omega |u(x)|^p \, dx, \qquad 1 \leq p < \infty,$$

form the Banach space $L^p(\Omega)$. All measurable functions $u(x)$, $x \in \mathbb{R}^n$, with finite norm $\|u\|_{L^p(\Omega)}$ for any bounded domain $\Omega \subset \mathbb{R}^n$ form the space $L_{\mathrm{loc}^p(\mathbf{R}^n)}$ which is not Banach one. The Sobolev space $W^{1,p}(\Omega)$ consists of all measurable functions $u(x)$, $x \in \Omega$, which have measurable in Ω partial derivatives $Du(x)$ with finite norm

$$\|u\|^p_{W^{1,p}(\Omega)} = \int_\Omega |Du|^p\,dx + \int_\Omega |u|^p\,dx. \tag{1.2}$$

We say that $u(x) \in W^{1,p}_{\mathrm{loc}(\mathbf{R}^n)}$ if the function $u(x)$ is defined for almost all $x \in \mathbb{R}^n$ and $u(x) \in W^{1,p}(\Omega)$ for any bounded domain $\Omega \subset \mathbb{R}^n$. Let $C_0^\infty(\Omega)$ be a set of infinitely differentiable functions with compact suppot in Ω. Closure of $C_0^\infty(\Omega)$ with respect to the norm (1.2) forms the Banach space $W_0^{1,p}(\Omega)$. The norm in $W_0^{1,p}(\Omega)$ is equivalent to the next one

$$\|u\|_{W_0^{1,p}(\Omega)} = \int_\Omega |Du|^p\,dx.$$

The duality between the space $L^p(\Omega)$ and its dual $L^{p'}(\Omega)$ we denote by $(f,u) = \int_\Omega f(x)u(x)\,dx$, where $f \in L^{p'}(\Omega)$, $u \in L^p(\Omega)$. The same notion is used for the duality between the space $W_0^{1,p}(\Omega)$ and its dual $(W_0^{1,p}(\Omega))^* = W^{-1,p'}(\Omega)$.

Definition 1.1. Let $u(x) \in W^{1,p}_{\mathrm{loc}(\mathbf{R}^n)\cap L_{\mathrm{loc}^q(\mathbf{R}^n)}}$. The function $u(x)$ is called a solution to the equation (1.1) if for any bounded domain $\Omega \subset \mathbb{R}^n$ with smooth boundary and any function $\psi(x) \in W_0^{1,p}(\Omega) \cap L^q(\Omega)$ the following equality holds

$$\int_\Omega \left(\sum_{i=1}^n A_i(x,Du)D_i\psi + A_0(x,u)\psi \right) dx = \int_\Omega \left(\sum_{i=1}^n f_i D_i\psi + f_0\psi \right) dx.$$

The main result of this paper is as the follow.

Theorem 1.1. *Assume that conditions (1A)–(3A) and (1f) are satisfied. Then the equation* (1.1) *has a solution in the sense of Definition 1.1.*

2. Approximation by bounded domains

We denote by B_R an open ball of the radius $R > 0$ centered at the origin of $\mathbb{R}^n$, i.e. $B_R = \{x \in \mathbb{R}^n : |x| < R\}$, with boundary $\partial B_R = \{x \in \mathbb{R}^n : |x| = R\}$. We approximate the equation (1.1) by the Dirichlet problems

$$\begin{gathered} A_R u \equiv -\sum_{i=1}^n D_i A_i(x,Du) + A_0(x,u) = f(x), \quad x \in B_R, \\ u(x) = 0, \quad x \in \partial B_R. \end{gathered} \tag{2.1}$$

Conditions on the functions $A_i(x,\xi)$ for $i = 1,\dots,n$, $A_0(x,u)$ and $f(x)$ are the same as in (1A)–(3A) and (1f) but now all these functions we consider on the ball B_R only. So the functions $A_i(x,\xi)$ for $i = 1,\dots,n$ and $A_0(x,u)$ are defined for $x \in B_R$, $\xi \in \mathbb{R}^n$, $u \in \mathbb{R}^1$ and satisfy the Carathéodory condition. They are subject to the following constraints.

$(1A_R)$ There exist constants C_{1R} and C_{0R} such that the following inequalities hold for $x \in B_R$, $\xi \in \mathbb{R}^n$ and $u \in \mathbb{R}^1$:

$$|A_i(x,\xi)| \le C_{1R}|\xi|^{p-1} + b_1(x), \qquad b_1(x) \in L^{p'}(B_R),$$
$$|A_0(x,u)| \le C_{0R}|u|^{q-1} + b_0(x), \qquad b_0(x) \in L^{q'}(B_R).$$

$(2A_R)$ For almost all $x \in B_R$ and all $\xi, \eta \in \mathbb{R}^n$ we have

$$\sum_{i=1}^n [A_i(x,\xi) - A_i(x,\eta)](\xi_i - \eta_i) > 0 \qquad (\xi \ne \eta).$$

$(3A_R)$ There exists a constant $c_R > 0$ such that for all $x \in B_R$, $\xi \in \mathbb{R}^n$, $u \in \mathbb{R}^1$ we have

$$\sum_{i=1}^n A_i(x,\xi)\xi_i + A_0(x,u)u \ge c_R(|\xi|^p + |u|^q) - g(x),$$

where $g(x) \in L^1(B_R)$.

$(1f_R)$ There exist functions $f_i(x) \in L^{p'}(B_R)$ for $i = 1, \ldots, n$ and $f_0(x) \in L^{q'}(B_R)$ such that we have $f(x) = -\sum_{i=1}^n D_i f_i(x) + f_0(x)$.

Throughout, we denote positive and, in general, different constants depending only on the parameters of the problem under consideration by c and C.

Let us introduce the space $X_R = W_0^{1,p}(B_R) \cap L^q(B_R)$ with the norm

$$\|u\|_{X_R} = \left(\int_{B_R} |Du|^p \, dx\right)^{1/p} + \left(\int_{B_R} |u|^q \, dx\right)^{1/q}. \tag{2.2}$$

It is well known that X_R is a separable reflexive Banach space with a dual space $X_R^* = W^{-1,p'}(B_R) + L^{q'}(B_R)$. We define an operator $A_R : X_R \to X_R^*$ by the formula for $u, v \in X_R$:

$$(A_R u, v) = \int_{B_R} \sum_{i=1}^n A_i(x, Du) D_i v \, dx + \int_{B_R} A_0(x,u) v \, dx. \tag{2.3}$$

It is known that the operator $A_R : X_R \to X_R^*$ under the conditions $(1A_R)$–$(3A_R)$ is bounded coercive continuous and that a function $f(x)$ under condition $(1f_R)$ is an element of the space X_R^*. The equation (2.1) may be written in the operator form $A_R u = f$, where $f \in X_R^*$. We are looking for a function $u \in X$ such that the identity $A_R u = f$ holds in the space X^*.

Recall the following definition. Let X be a real reflexive Banach space and let X^* denotes its dual space with a duality (f, u), where $f \in X^*$, $u \in X$. A mapping $A : X \to X^*$ possesses (M)-property (or it is type M) if for each sequence $\{u_k\}$ converging weakly to $u \in X$ from the conditions as $k \to \infty$

$$Au_k \rightharpoonup f \quad (\text{in } X^*), \qquad \limsup (Au_k, u_k) \le (f, u) \tag{2.4}$$

we have the equality $Au = f$.

We may write the conditions (2.4) in the form

$$\limsup_{k\to\infty}(Au_k, u_k - u) = \limsup_{k\to\infty}(Au_k, u_k) - \lim_{k\to\infty}(Au_k, u) = \limsup_{k\to\infty}(Au_k, u_k) - (f, u) \le 0.$$

It is easy to see that an operator $A : X \to X^*$ possesses (M)-property if from the conditions as $k \to \infty$

$$u_k \rightharpoonup u \quad (\text{in } X), \qquad \limsup(Au_k, u_k - u) \le 0 \tag{2.5}$$

we have that $Au_k \rightharpoonup Au$ in X^*.

The main assertion of this section is as follow.

Theorem 2.1. *The operator $A_R : X_R \to X_R^*$ defined by* (2.3) *under conditions* ($1A_R$)-($3A_R$) *possesses (M)-property.*

Proof of the theorem consists of a few lemmas. We suppose that a sequence $\{u_k\}$ convergering weakly to $u \in X_R$ is given.

Lemma 2.1. *There exists a subsequence of the sequence $\{u_k\}$ which we denote also by $\{u_k\}$, such that, for some $w_i \in L^{p'}(B_R)$ ($i = 1, \dots, n$) and $w_0 \in L^{q'}(B_R)$ we have the following convergence relations as $k \to \infty$:*

$$a) \quad A_i(x, Du_k) \rightharpoonup w_i \text{ in } L^{p'}(B_R) \text{ for } i = 1, \dots, n,$$
$$b) \quad A_0(x, u_k) \rightharpoonup w_0 \text{ in } L^{q'}(B_R).$$

Proof. If a sequence $\{u_k\}$ converges weakly to $u \in X_R$ then it is bounded uniformly: $\|u_k\|_{X_R} \le C$ with a unique constant C for all $k \in \mathbb{N}$. It follows from (2.2) that

$$\int_{B_R} |Du_k|^p\, dx + \int_{B_R} |u_k|^q\, dx \le C \qquad (k \in \mathbb{N}). \tag{2.6}$$

By growth condition ($1A_R$), for $i = 1, \dots, n$ we have

$$|A_i(x, Du_k)| \le C_{1R}|Du_k|^{p-1} + b_1(x).$$

Raising to the power $p' = \frac{p}{p-1}$ and integrating over B_R we obtain

$$\int_{B_R} |A_i(x, Du_k)|^{p'}\, dx \le c\left(C_{1R}^{p'} \int_{B_R} |Du_k|^p\, dx + \int_{B_R} b_1^{p'}(x)\, dx\right) \le C,$$

where C is independent of $k \in \mathbb{N}$.

Analogously, by growth condition ($1A_R$) we have

$$|A_0(x, u_k)| \le C_{0R}|u_k|^{q-1} + b_0(x).$$

Rasing to the power q' and integrating over B_R we obtain

$$\int_{B_R} |A_0(x, u_k)|^{q'}\, dx \le c\left(C_{0R}^{q'} \int_{B_R} |u|^q\, dx + \int_{B_R} b_0^{q'}(x)\, dx\right) \le C,$$

where C is independent of $k \in \mathbb{N}$.

Now the assertions of the lemma are consequence of the reflexivity of the spaces $L^{p'}(B_R)$ and $L^{q'}(B_R)$. □

Lemma 2.2. *There exists a subsequence of the sequence $\{u_k\}$, which we denote also by $\{u_k\}$, such that the following limit relations hold as $k \to \infty$:*

$$
\begin{aligned}
&(a) \quad u_k \to u \text{ in } L^p(B_R);\\
&(b) \quad u_k(x) \to u(x) \text{ a.e. in } B_R;\\
&(c) \quad A_0(x, u_k(x)) \to A_0(x, u(x)) \text{ a.e. in } B_R;\\
&(d) \quad A_0(x, u_k(x))u_k(x) \to A_0(x, u(x))u(x) \text{ a.e. in } B_R.
\end{aligned}
$$

Proof. Let us write the inequality (2.6) in the form

$$\|u_k\|_{W_0^{1,p}(B_R)} + \|u_k\|_{L^q(B_R)} \le C \quad (k \in \mathbb{N}).$$

We see that the sequence $\{u_k\}$ is uniformly bounded in the Sobolev space $W_0^{1,p}(B_R)$, which is compactly embedded in $L^p(B_R)$. Hence there exists a subsequence of the sequence $\{u_k\}$, which we denote also by $\{u_k\}$, such that $u_k \to u$ in $L^p(B_R)$ and $u_k(x) \to u(x)$ a.e. in B_R. This proves two first assertions of the Lemma.

The rest assertions are the consequences of two first ones which completes the proof. □

Now, we suppose that the last inequality in (2.5) for the operator A_R holds:

$$\limsup_{k\to\infty}(A_R u_k, u_k - u) \le 0,$$

or in the full form

$$\limsup_{k\to\infty} \int_{B_R} \left(\sum_{i=1}^n A_i(x, Du_k)(D_i u_k - D_i u) + A_0(x, u_k)(u_k - u) \right) dx \le 0. \tag{2.7}$$

Let us introduce the functions defined for $x \in B_R$:

$$
\begin{aligned}
F_k(x) &= \sum_{i=1}^n (A_i(x, Du_k) - A_i(x, Du))(D_i u_k - D_i u);\\
G_k(x) &= \sum_{i=1}^n A_i(x, Du)(D_i u_k - D_i u);\\
H_k(x) &= A_0(x, u_k)(u_k - u).
\end{aligned}
$$

Then we may write the condition (2.7) in the form

$$\limsup_{k\to\infty} \left(\int_{B_R} F_k(x)\,dx + \int_{B_R} G_k(x)\,dx + \int_{B_R} H_k(x)\,dx \right) \le 0. \tag{2.8}$$

We find the limit of each term as $k \to \infty$.

Lemma 2.3. *If $u_k \rightharpoonup u$ in X_R as $k \to \infty$, then*

$$\lim_{k\to\infty} \int_{B_k} G_k(x)\,dx = 0.$$

Proof. As $u_k \rightharpoonup u$ in X_R $(k \to \infty)$, so $Du_k \rightharpoonup Du$ in $L^p(B_R)$ and $Du \in L^p(B_R)$. By growth condition $(1A_R)$ then $A_i(x, Du) \in L^{p'}(B_R)$ for $i = 1, \dots, n$. Therefore as $k \to \infty$

$$\int_{B_R} A_i(x, Du)(D_i u_k - D_i u)\, dx \to 0 \qquad (i = 1, \dots, n).$$

This is the required result. □

Lemma 2.4. *If $u_k \rightharpoonup u$ in X_R as $k \to \infty$, then*

$$\liminf_{k\to\infty} \int_{B_k} H_k(x)\, dx \geq 0.$$

Proof. We know from Lemma 2.1(b) and Lemma 2.2(c) that $A_0(x, u_k) \rightharpoonup w_0$ in $L^{q'}(B_R)$ and $A_0(x, u_k(x)) \to A_0(x, u(x))$ a.e. in B_R. Comparing these convergences we see that $w_0 = A_0(x, u)$ and

$$A_0(x, u_k) \rightharpoonup A_0(x, u) \text{ in } L^{q'}(B_R). \tag{2.9}$$

In particular

$$\lim_{k\to\infty} \int_{B_R} A_0(x, u_k(x))u(x)\, dx = \int_{B_R} A_0(x, u(x))u(x)\, dx. \tag{2.10}$$

Suppose that $\xi = 0$ in coercive condition $(3A_R)$, then we have for all $x \in B_R$ and $u \in \mathbb{R}^1$:

$$A_0(x, u)u \geq c_R|u|^q - g(x) \qquad (c_R > 0;\ g(x) \in L^1(B_R)).$$

Therefore the functions $z_k(x) = A_0(x, u_k(x))u_k(x) + g(x)$ are nonnegative for all $x \in B_R$. By Fatue's lemma and in view of Lemma 2.2(d) we have

$$\liminf_{k\to\infty} \int_{B_R} A_0(x, u_k(x))u_k(x)\, dx \geq \int_{B_R} A_0(x, u(x))u(x)\, dx. \tag{2.11}$$

Subtracting (2.10) from (2.11) we get

$$\liminf_{k\to\infty} \int_{B_R} A_0(x, u_k(x))(u_k(x) - u(x))\, dx \geq 0.$$

This is the assertion of the lemma. □

Lemma 2.5. *Under conditions* (2.5) *for the operator $A_R : X_R \to X_R^*$ we have for almost all $x \in B_R$:*

$$\lim_{k\to\infty} \sum_{i=1}^{n} \big(A_i(x, Du_k(x)) - A_i(x, Du(x))\big)\big(D_i u_k(x) - D_i u(x)\big) = 0.$$

Proof. Conditions (2.5) are equivalent to the inequality (2.8). Taking Lemmas 2.3 and 2.4 into account, we obtain from (2.8)

$$\limsup_{k\to\infty} \int_{B_k} F_k(x)\, dx \leq 0.$$

However, $F_k(x) \geq 0$ a.e. in B_R by monotonicity condition $(2A_R)$. This shows that there exists $\lim_{k\to\infty} F_k(x) = 0$ a.e. in B_R, which in a more extended form can be written as the assertion of the lemma. □

Lemma 2.6. *Under conditions* (2.5) *for the operator* $A_R : X_R \to X_R^*$ *we have that* $Du_k(x) \to Du(x)$ *as* $k \to \infty$ *for almost all* $x \in B_R$.

Proof. We choose a point $x \in B_R$ such that: $Du(x)$ and $g(x)$ are well defined and finite; the functions $A_i(x, \xi)$ are continuous in $\xi \in \mathbb{R}^n$ for $i = 1, \dots, n$. Each of these properties holds a.e. in B_R, therefore their combination must also hold for almost all $x \in B_R$. For such a point we set $\xi_k = Du_k(x)$, and $\eta = Du(x)$. We consider also the number sequence

$$\varphi_k = \sum_{i=1}^{n} (A_i(x, \xi_k) - A_i(x, \eta))(\xi_{ki} - \eta_i).$$

In view of Lemma 2.5 and the above notation we can assume that $\varphi_k \to 0$ as $k \to \infty$. We are now in position to discuss the properties of the sequence $\xi_k = Du_k(x)$. Assume that it is unbounded, that is, there exists a subsequence ξ_{k_m} such that $|\xi_{k_m}| \to \infty$ as $m \to \infty$. Using the coercivity condition $(3A_R)$ for $u = 0$ we obtain

$$\varphi_k = \sum_{i=1}^{n} (A_i(x, \xi_k) - A_i(x, \eta))(\xi_{k_i} - \eta_i)$$
$$\geq c_R |\xi_k|^p - g(x) - \sum_{i=1}^{n} A_i(x, \eta)(\xi_{k_i} - \eta_i) - \sum_{i=1}^{n} A_i(x, \xi_k)\eta_i.$$

As $c_R > 0$, η is fixed and $p > 1$ it is evident that $\varphi_{k_m} \to \infty$ $(m \to \infty)$. This contradicts the assumption $\varphi_k \to 0$ as $k \to \infty$, therefore the sequence ξ_k is uniformly bounded. Let ξ be a limit point of it, that is, assume that $\xi_{k_m} \to \xi$ as $m \to \infty$. The functions $A_i(x, \xi)$ are continuous in the last arguments, therefore

$$\lim_{m \to \infty} \varphi_{k_m} = \sum_{i=1}^{n} (A_i(x, \xi) - A_i(x, \eta))(\xi_i - \eta_i) = 0.$$

By monotonicity conditions $(2A_R)$ this is possible only for $\xi = \eta$. Hence each limit point of the sequence $\{\xi_k\}$ coincides with $\eta = Du(x)$, which is equivalent to the limit relation $Du_k(x) \to Du(x)$ as $k \to \infty$. By construction, this holds for almost all $x \in B_R$, as required. □

Lemma 2.7. *Under conditions* (2.5) *for the operator* $A_R : X_R \to X_R^*$ *we have as* $k \to \infty$ *for* $i = 1, \dots, n$:

$$A_i(x, Du_k) \rightharpoonup A_i(x, Du) \text{ in } L^{p'}(B_R).$$

Proof. We know from Lemma 2.1(a) that $A_i(x, Du_k) \rightharpoonup w_i$ in $L^{p'}(B_R)$. It follows from Lemma 2.6 that $A_i(x, Du_k(x)) \to A_i(x, Du(x))$ for almost all $x \in B_R$. Comparing these convergences we see that $w_i = A_i(x, u(x))$ and that the assertion of the lemma holds. □

Proof of the Theorem 2.1. Let the sequence $\{u_k\}$ with the conditions (2.5) for the operator $A_R : X_R \to X_R^*$ is given. From Lemma 2.7 we have for any $v \in X_R$ as

$k \to \infty$:

$$\int_{B_R} \sum_{i=1}^{n} A_i(x, Du_k) D_i v \, dx \to \int_{B_R} \sum_{i=1}^{n} A_i(x, Du) D_i v \, dx. \tag{2.12}$$

From (2.9) if follows that as $k \to \infty$

$$\int_{B_R} A_0(x, u_k) v \, dx \to \int_{B_R} A_0(x, u) v \, dx. \tag{2.13}$$

Summarizing (2.12) and (2.13) we get as $k \to \infty$

$$\int_{B_R} \left(\sum_{i=1}^{n} A_i(x, Du_k) D_i v + A_0(x, u_k) v \right) dx \to$$

$$\to \int_{B_R} \left(\sum_{i=1}^{n} A_i(x, Du) D_i v + A_0(x, u) v \right) dx,$$

or in the short form: $(A_R u_k, v) \to (A_R u, v)$, $v \in X_R$. In other words, if the conditions (2.5) for the operator A_R hold then $A_R u_k \rightharpoonup A_R u$ as $k \to \infty$. That is, the operator $A_R : X_R \to X_R^*$ possesses (M)-property which proves Theorem 2.1. $\square$

Now we know that the operator $A_R : X_R \to X_R^*$ defined by (2.3) under conditions ($1A_R$)–($3A_R$) is bounded continuous coercive and possesses (M)-property. This is enough for the following existence theorem which is found in [11], Theorem 2.1: for each $f \in X_R^*$ there exists $u \in X$ such that $A_R u = f$. We get the next assertion.

Theorem 2.2. *The Dirichlet problem* (2.1) *has a solution $u \in X_R$ for any $f \in X_R^*$ as an equation $A_R u = f$ with the operator A_R defined by* (2.3) *under conditions* ($1A_R$)*–*($3A_R$) *and* ($1f_R$)*.*

3. Proof of Theorem 1.1

In vew of Theorem 2.2 for each $R > 0$ there exists a solution $u_R \in X_R = W_0^{1,p}(B_R) \cap L^q(B_R)$ of the boundary value problem (2.1), that is,

$$\begin{gathered} -\sum_{i=1}^{n} D_i A_i(x, Du_R) + A_0(x, u_R) = -\sum_{i=1}^{n} D_i f_i(x) + f_0(x), \quad |x| < R; \\ u_R(x) = 0, \quad |x| = R. \end{gathered} \tag{3.1}$$

Let $\{u_N\}$ be a sequence of solutions of the problem (3.1) for $R = N$, $N \in \mathbb{N}$, so that for each $N \in \mathbb{N}$ we have

$$\int_{B_N} \sum_{i=1}^{n} A_i(x, Du_N) D_i \psi \, dx + \int_{B_N} A_0(x, u_N) \psi \, dx = \int_{B_N} \sum_{i=1}^{n} f D_i \psi \, dx + \int_{B_N} f_0 \psi \, dx, \tag{3.2}$$

where $\psi \in X_R = W_0^{1,p}(B_R) \cap L^q(B_R)$, $q > p > 1$.

We fix a nonnegative function $\varphi(x) \in C_0^1(\mathbb{R}^n)$ and a number $m \in \mathbb{N}$ such that $\operatorname{supp} \varphi \subset B_m$. For all $N > m$ we define the functions $\psi = u_N \varphi^s$, where $s = \frac{qp}{q-p}$. It is evident that $D_i\psi = (D_i u_N)\varphi^s + s\varphi^{s-1}u_N D_i\varphi$. Using the condition $\operatorname{supp}\psi \subset B_m$ and therefore $\psi(x) \equiv 0$ for $|x| \geq m$ we obtain from (3.2) the identity for $N > m$:

$$\int_{B_m} \sum_{i=1}^n A_i(x, Du_N)(D_i u_N)\varphi^s\,dx + s\int_{B_m} \sum_{i=1}^n A_i(x, Du_N)u_N\varphi^{s-1}D_i\varphi\,dx$$
$$+ \int_{B_m} A_0(x, u_N)u_N\varphi^s\,dx = \int_{B_m} \sum_{i=1}^n f_i(D_i u_N)\varphi^s\,dx$$
$$+ s\int_{B_m} \sum_{i=1}^n f_i u_N\varphi^{s-1}D_i\varphi\,dx + \int_{B_m} f_0 u_N\varphi^s\,dx,$$

which we agree to write as

$$J_1 + J_2 + J_3 = F_1 + F_2 + F_3. \tag{3.3}$$

Here we estimate each term. Notice the following inportant fact. We evaluate all integrals over the set B_m where m is fixed, therefore we need conditions (1A)-(3A) and (1f) for $x \in B_m$ only, that is, for $|x| < m$. In other words, we may use the conditions $(1A_R)$–$(3A_R)$ from section 2 for $R = m$. For example, using coercivity condition $(3A_R)$ for $R = m$ we have for $N > m$

$$J_1 + J_3 = \int_{B_m} \left(\sum_{i=1}^n A_i(x, Du_N)D_i u_N + A_0(x, u_N)u_N\right)\varphi^s\,dx$$
$$\geq c_m \int_{B_m} \left(|Du_N|^p + |u_N|^q\right)\varphi^s\,dx - C_m,$$

where constants $c_m > 0$ and $C_m = \int_{B_m} g(x)\varphi^s(x)\,dx$ do not depend on $N \in \mathbb{N}$ for $N > m$.

Now let us estimate the integral J_2:

$$|J_2| \leq s\sum_{i=1}^n \int_{B_m} |A_i(x, Du_N)|\varphi^{s/p'}|u_N|\varphi^{s/q}|D_i\varphi|\,dx.$$

By Hölder's inequality for the exponents p', q and $s = \frac{qp}{q-p}$ we obtain

$$|J_2| \leq s\sum_{i=1}^n \left(\int_{B_m} |A_i(x, Du_N)|^{p'}\varphi^s\,dx\right)^{1/p'} \left(\int_{B_m} |u_N|^q\varphi^s\,dx\right)^{1/q} \left(\int_{B_m} |D_i\varphi|^s\,dx\right)^{1/s}.$$

Finally, using Young's inequality we have for any $\varepsilon > 0$

$$|J_2| \leq \varepsilon\sum_{i=1}^n \int_{B_m} |A_i(x, Du_N)|^{p'}\varphi^s\,dx + \varepsilon\int_{B_m} |u_N|^q\varphi^s\,dx + C(\varepsilon).$$

Here $C(\varepsilon)$ is a continuous function defined for all $\varepsilon > 0$. Using the growth condition $(1A_R)$ for $R = m$ we get the inequality

$$|J_2| \leq \varepsilon c\int_{B_m} \left(C_{1m}^{p'}|Du_N|^p + b_1^{p'}(x)\right)\varphi^s\,dx + \varepsilon\int_{B_m} |u_N|^q\varphi^s\,dx + C(\varepsilon),$$

where all constants are independent of N for $N > m$.

Next estimates are rather evident:

$$|F_1| \le \int_{B_m} \sum_{i=1}^{n} |f_i \varphi^{s/p'}| \cdot |D_i u_N| \varphi^{s/p} dx$$

$$\le \varepsilon \int_{B_m} |Du_N|^p \varphi^s \, dx + C(\varepsilon) \sum_{i=1}^{n} \int_{B_m} |f_i|^{p'} \varphi^s \, dx;$$

$$|F_2| \le s \int_{B_m} \sum_{i=1}^{n} |f_i \varphi^{s/p'}| \cdot |u_N| \varphi^{s/q} |D_i \varphi| \, dx$$

$$\le \varepsilon \int_{B_m} |u_N|^q \varphi^s \, dx + C(\varepsilon) \sum_{i=1}^{n} \left(\int_{B_m} |f_i|^{p'} \varphi^s \, dx + \int_{B_m} |D_i \varphi|^s \, dx \right);$$

$$|F_3| \le \int_{B_m} |f_0 \varphi^{s/q'}| \cdot |u_N \varphi^{s/q}| \, dx \le \varepsilon \int_{B_m} |u_N|^q \varphi^s \, dx + C(\varepsilon) \int_{B_m} |f_0|^{q'} \varphi^s \, dx.$$

Combining the above estimates of the individual terms in identity (3.3) we arrive at the inequality:

$$\left(c_m - \varepsilon(c C_{1m}^{q'} + 1)\right) \int_{B_m} |Du_N|^p \varphi^s \, dx + (c_m - \varepsilon c) \int_{B_m} |u_N|^q \varphi^s \, dx$$

$$\le C_m + \varepsilon c \int_{B_m} b_1^{p'} \varphi^s \, dx + C(\varepsilon) \int_{B_m} \left(\sum_{i=1}^{n} |f_i|^{p'} \varphi^s + |f_0|^{q'} \varphi^s + |D\varphi|^s \right) dx.$$

Clearly, if we choose sufficiently small $\varepsilon > 0$, then we get the following estimate for all $N > m$:

$$\int_{B_m} \left(|Du_N|^p + |u_N|^q \right) \varphi^s \, dx \le C(\varepsilon, m). \tag{3.4}$$

Here the constant $C(\varepsilon, m)$ depends on $\varepsilon > 0$ and $m \in \mathbb{N}$, but it does not depend on $N \in \mathbb{N}$ if $N > m$.

We take a special function $\varphi \in C_0^1(\mathbb{R}^n)$ so that $\varphi(x) \equiv 1$ for $|x| \le m - 1$ and $\varphi(x) \equiv 0$ for $|x| \ge m - \frac{1}{2}$. Then we may write the estimate (3.4) in the form

$$\int_{|x| \le m-1} \left(|Du_N|^p + |u_N|^q \right) dx \le C(m-1), \qquad N > m. \tag{3.5}$$

For the sake of simplicity let us change $m - 1$ on m. Using also the growth conditions $(1A_R)$ for $R = m$ we arrive at the following conclusion.

Lemma 3.1. *The solutions u_N of the boundary value problems* (3.2) *satisfy the estimates for each fixes $m \in \mathbb{N}$ and all $N > m + 1$:*

$$\|Du_N\|_{L^p(B_m)} \le C(m); \qquad \|u_N\|_{L^q(B_m)} \le C(m);$$
$$\|A_i(x, Du_N)\|_{L^{p'}(B_m)} \le C(m); \qquad \|A_0(x, u_N)\|_{L^{q'}(B_m)} \le C(m).$$

Here constant $C(m)$ does not depend on N for $N > m + 1$.

Now we can construct a solution of the equation (1.1) step by step.

Step 1. Let $m = 1$. We introduce the Banach space $Y_1 = W^{1,p}(B_1) \cap L^q(B_1)$ with the norm

$$\|u\|_{Y_1} = \|Du\|_{L^p(B_1)} + \|u\|_{L^q(B_1)}.$$

From Lemma 3.1 we have the estimate $\|u_N\|_{Y_1} \le C(1)$ for $N > 2$. The space Y_1 is reflexive for $p, q > 1$ so there exists a subset $N(1) \subset \mathbb{N}$ of natural numbers and a function $u^{(1)}(x)$ defined in the ball B_1 such that $u_k \rightharpoonup u^{(1)}$ in Y_1 for $k \in N(1)$ as $k \to \infty$. The space Y_1 for $q > p$ is compactly embedded in $L^p(B_1)$ so we may suppose that for $k \in N(1)$ as $k \to \infty$:

(3.7)

$u_k(x) \to u^{(1)}(x)$ a.e. in B_1.

Step 2. Let $m = 2$. From Lemma 3.1 we have the estimate $\|u_N\|_{Y_2} \le C(2)$ for $N > 3$ where $Y_2 = W^{1,p}(B_2) \cap L^q(B_2)$. This estimate we consider on the set $N(1)$ only, that is,

$$\|u_k\|_{Y_2} \le C(2), \qquad k \in N(1).$$

The space Y_2 is reflexive and compactly embedded in $L^p(B_2)$ so there exists a subset $N(2) \subset N(1)$ and a function $u^{(2)}(x)$ defined in the ball B_2 such that for $k \in N(2)$ as $k \to \infty$ we have:

$$\begin{gathered} u_k \rightharpoonup u^{(2)} \text{ in } Y_2; \qquad u_k(x) \to u^{(2)} \text{ in } L^p(B_2); \\ u_k(x) \to u^{(2)}(x) \text{ a.e. in } B_2. \end{gathered} \tag{3.8}$$

The set $N(2)$ is a subset of $N(1)$, therefore the convergences (**??**) are correct for $k \in N(2)$. Comparing (**??**) and (3.8) we conclude that $u^{(2)}(x) = u^{(1)}(x)$ for $x \in B_1$.

Step m. For any $m \in \mathbb{N}$ we construct a space $Y_m = W^{1,p}(B_m) \cap L^q(B_m)$, a set $N(m) \subset N(m-1)$ and a function $u^{(m)}(x)$ defined in the ball B_m such that for $k \in N(m)$ as $k \to \infty$ we have:

(3.11)

$u_k(x) \to u^{(m)}(x)$ a.e. in B_m;

Continuing the process we obtain a sequence of functions $u^{(1)}(x)$, $u^{(2)}(x)$, $u^{(3)}(x)$, … with the properties:

$$u^{(m)}(x) \in W^{1,p}(B_m) \cap L^q(B_m);$$
$$u^{(m)}(x) = u^{(m-1)}(x) \text{ for } x \in B_{m-1}.$$

The last properties shows that we have a unique function $u(x)$ defined for all $x \in \mathbb{R}^n$ and such that

$$u(x) = u^{(m)}(x) \text{ for } x \in B_m. \tag{3.12}$$

In particular, we have

$$u(x) \in W^{1,p}_{\mathrm{loc}(\mathbf{R}^n) \cap L^q_{\mathrm{loc}(\mathbf{R}^n).}}$$

Next we show that the function $u(x)$ defined by (3.12) satisfies the equality

$$\int_\Omega \left(\sum_{i=1}^n A_i(x, Du) D_i \psi + A_0(x, u)\psi \right) dx = (f, \psi)$$

for any test function $\psi(x) \in W_0^{1,p}(\Omega) \cap L^q(\Omega)$ with a compact support $\Omega \subset \mathbb{R}^n$. We suppose that $\psi(x)$ is extended by zero outside Ω and fix such a function $\psi(x)$ and a natural number $m \in \mathbb{N}$ with the property $\overline{\Omega} \subset B_m$, so that

$$\psi(x) \equiv 0 \text{ for } |x| \geq m, \tag{3.13}$$

and $\psi(x) \in W_0^{1,p}(B_m) \cap L^q(B_m)$.

We consider the identity (3.2) for $N > m$. In view of (3.13) we may integrate in (3.2) for $x \in B_m$ only, so we have for $N > m$

$$\int_{B_m} \left(\sum_{i=1}^n A_i(x, Du_N) D_i \psi + A_0(x, u_N)\psi \right) dx = (f, \psi). \tag{3.14}$$

For a fixed $m \in \mathbb{N}$ we introduce the subset $N(m) \subset \mathbb{N}$ from the construction of the function $u(x)$ and consider the identities (3.14) for $k \in \mathbb{N}(m)$ only:

$$\int_{B_m} \sum_{i=1}^n A_i(x, Du_k) D_i \psi \, dx + \int_{B_m} A_0(x, u_k)\psi \, dx = (f, \psi). \tag{3.15}$$

We know from (**??**) and (3.12) that for $k \in N(m)$ as $k \to \infty$ we have

$$u_k(x) \to u^{(m)}(x) = u(x) \text{ a.e. in } B_m$$

and therefore

$$A_0(x, u_k(x)) \to A_0(x, u(x)) \text{ a.e. in } B_m. \tag{3.16}$$

Comparing (3.16) with the estimate $\|A_0(x, u_k)\|_{L^{q'}(B_m)} \leq C(m)$, from Lemma 3.1 we conclude that for $k \in N(m)$ as $k \to \infty$

$$A_0(x, u_k) \rightharpoonup A_0(x, u) \text{ in } L^{q'}(B_m). \tag{3.17}$$

Let us take a function $\varphi(x) \in C_0^1(B_m)$ and consider the functions $\psi = (u_k - u)\varphi$ in identities (3.15). We get for $k \in N(m)$

$$\int_{B_m} \sum_{i=1}^n A_i(a, Du_k)(D_i u_k - D_i u)\varphi\, dx + \int_{B_m} A_0(x, u_k)(u_k - u)\varphi\, dx$$
$$= (f, (u_k - u)\varphi) - \int_{B_m} \sum_{i=1}^n A_i(x, Du_k)(u_k - u) D_i\varphi\, dx. \quad (3.18)$$

We know that $u_k \rightharpoonup u$ in $Y_m = W^{1,p}(B_m) \cap L^q(B_m)$, so for $k \in N(m)$

$$\lim_{k\to\infty} (f, (u_k - u)\varphi) = 0. \quad (3.19)$$

Now let us estimate last integral in (3.18).

Lemma 3.2. *We have for* $k \in N(m)$

$$\lim_{k\to\infty} \int_{B_m} \sum_{i=1}^n A_i(x, Du_k)(u_k - u) D_i\varphi\, dx = 0.$$

Proof. Using Hölder's inequality we have for $k \in N(m)$ as $k \to \infty$

$$\left| \int_{B_m} A_i(x, Du_k)(u_k - u) D_i\varphi\, dx \right|$$
$$\le \max_{|x|\le m} |D\varphi| \cdot \|A_i(x, Du_k)\|_{L^{p'}(B_m)} \|u_k - u\|_{L^p(B_m)} \to 0,$$

as it follows from Lemma 3.1 and relations (**??**). $\square$

Using (3.19) and Lemma 3.2 we get from (3.18) for $k \in N(m)$

$$\lim_{k\to\infty} \left(\int_{B_m} \sum_{i=1}^n A_i(x, Du_k)(D_i u_k - D_i u)\varphi\, dx + \int_{B_m} A_0(x, u_k)(u_k - u)\varphi\, dx \right) = 0. \quad (3.20)$$

The relation (3.20) is analogous to the inequality (2.7). We represent (3.20) as follows for $k \in N(m)$:

$$\lim_{k\to\infty} \Bigg[\int_{B_m} \sum_{i=1}^n (A_i(x, Du_k) - A_i(x, Du))(D_i u_k - D_i u)\varphi\, dx$$
$$+ \int_{B_m} \sum_{i=1}^n A_i(x, Du)(D_i u_k - D_i u)\varphi\, dx + \int_{B_m} A_0(x, u_k)(u_k - u)\varphi\, dx \Bigg] = 0.$$

It is evident that the factor $\varphi(x)$ does not play an essential role for estimates, hence proceeding as in Lemmas 2.3–2.5 we obtain the next assertions.

Lemma 3.3. *We have for* $k \in N(m)$

$$\lim_{k\to\infty} \int_{B_m} \sum_{i=1}^n A_i(x, Du)(D_i u_k - D_i u)\varphi\, dx = 0.$$

Lemma 3.4. *We have for $k \in N(m)$ as $k \to \infty$*

$$\liminf \int_{B_m} A_0(x, u_k)(u_k - u)\varphi(x)\, dx \geq 0.$$

Lemma 3.5. *We have for $k \in N(m)$ and almost all $x \in B_m$*

$$\lim_{k\to\infty} \sum_{i=1}^{n} \left[A_i(x, Du_k(x)) - A_i(x, Du(x))\right](D_i u_k(x) - D_i u(x))\varphi(x) = 0.$$

Now we choose a special function $\varphi(x) \in C_0^1(B_m)$ with a property: $\varphi(x) \equiv 1$ for $x \in \overline{\Omega}$, where $\overline{\Omega} = \operatorname{supp}\psi$ and $\psi(x)$ is a given function such that $\overline{\Omega} \subset B_m$. Then we have from Lemma 3.5 that for $k \in N(m)$ and almost all $x \in \overline{\Omega}$

$$\lim_{k\to\infty} \sum_{i=1}^{n} \left[A_i(x, Du_k(x)) - A_i(x, Du(x))\right](D_i u_k(x) - D_i u(x))\varphi(x) = 0. \tag{3.21}$$

Using (3.21) and repeating the proof of Lemma 2.6 we obtain the next assertion.

Lemma 3.6. *We have $Du_k(x) \to Du(x)$ for $k \in N(m)$ as $k \to \infty$ for almost all $x \in \overline{\Omega}$.*

Lemma 3.7. *We have for $k \in N(m)$ as $k \to \infty$ for $i = 1, \dots, n$:*

$$A_i(x, Du_k) \rightharpoonup A_i(x, Du) \text{ in } L^{p'}(\Omega).$$

Proof. We know from Lemma 3.1 that $\|A_i(x, Du_k)\|_{L^{p'}(\Omega)} \leq C(m)$. It follows from Lemma 3.6 that $A_i(x, Du_k) \to A_i(x, Du)$ for almost all $x \in \Omega$ as $k \to \infty$, $k \in N(m)$. This is enough for the assertion of the Lemma. □

Now we can finish the proof of Theorem 1.1. From Lemma 3.7 for a given test function $\psi(x) \in W_0^{1,p}(\Omega) \cap L^q(\Omega)$ we have as $k \to \infty$, $k \in N(m)$:

$$\int_\Omega \sum_{i=1}^{n} A_i(x, Du_k) D_i\psi\, dx \to \int_\Omega \sum_{i=1}^{n} A_i(x, Du) D_i\psi\, dx. \tag{3.22}$$

From (3.17) we have for $k \in N(m)$ as $k \to \infty$:

$$\int_\Omega A_0(x, u_k)\psi\, dx \to \int_\Omega A_0(x, u)\psi\, dx. \tag{3.23}$$

Using limit relations (3.22) and (3.23) in the identities (3.15) we get for $k \in N(m)$ as $k \to \infty$:

$$\int_\Omega \left(\sum_{i=1}^{n} A_i(x, Du) D_i\psi + A_0(x, u)\psi \right) dx = (f, \psi).$$

This equality means that the function $u(x)$ is a solution of the equation (1.1) in the sense of Definition 1.1, which completes the proof of Theorem 1.1.

4. Examples

In conclusion we present examples of the equations satisfying all the conditions of the solvability Theorem 1.1. Consider the equation

$$-\sum_{i=1}^{n} D_i\left(\alpha_i(x)|D_iu|^{p-2}D_iu\right)+\alpha_0(x)|u|^{q-2}u=f(x). \tag{4.1}$$

Here $1<p<q$ are arbitrary exponents, the function $f(x)$ is supposed to satisfy the condition (1f) from section 1. The functions $\alpha_i(x)>0$ for $i=0,1,\dots,n$ are defined for almost all $x\in\mathbb{R}^n$ and such that

$$\alpha_i(x),\ \alpha_i^{-1}(x)\in L_{\mathrm{loc}^{\infty}(\mathbf{R}^n)}. \tag{4.2}$$

It is easy to check all the conditions (1A)-(3A). So the equation (4.1) under the conditions (4.2) has a solution $u\in W^{1,p}_{\mathrm{loc}(\mathbf{R}^n)\cap L^q_{\mathrm{loc}(\mathbf{R}^n)}}$ for any $f(x)=-\sum_{i=1}^n D_if_i(x)+f_0(x)$, where $f_i\in L_{\mathrm{loc}^{p'}(\mathbf{R}^n)}$ and $f_0\in L_{\mathrm{loc}^{q'}(\mathbf{R}^n)}$.

For example the simplest case is the next one:

$$-\sum_{i=1}^{n} D_i(|D_iu|^{p-2}D_iu)+|u|^{q-2}u=f(x).$$

The functions $\alpha_i(x)$ (in (4.1)) for $i=0,1,\dots,n$ may grow arbitrary as $|x|\to\infty$. For example we can solve the equation

$$-\sum_{i=1}^{n} D_i(e^{x_1+\dots+x_n}|D_iu|^{p-2}D_iu)+e^{-|x|^2}|u|^{q-2}u=f(x),\quad x\in\mathbb{R}^n.$$

5. Acknowledgments

Author was supported in part by RFBR 01-01-00884 and 02-01-00069.

References

[1] E. Di Benedetto, M.A. Herrero, *On the Cauchy problem and initial traces for a degenerate parabolic equation*, Trans. Amer. Math. Soc. 314 (1989), 187–224.
[2] H. Brezis, *Semilinear equation in $\mathbf{R}^N$ without condition at infinity*, Appl. Math. Optim. 12 (1984), 271–282.
[3] F.E. Browder, *Nonlinear monotone operators and convex sets in Banach spaces*, Bull. Amer. Math. Soc. 71 (1965), 780–785.
[4] A.L. Gladkov, *Dirichlet problem for a certain degenerate elliptic equations in unbounded domains*, Differenc. Uravn. 29 (1993), 267–273.
[5] A.L. Gladkov, *Evolution in time of solutions of nonlinear parabolic equations*, Nonlinear boundary value problems 8 (1998), 108-113.
[6] N. Hungerbühler, *Quasilinear elliptic systems in divergence form with weak monotonicity*, New York J. Math. 5 (1999), 83–90.
[7] A.A. Kon'kov, *Behavior of solutions of nonlinear second-order elliptic inequalities*, Nonlinear Anal. TMA. 42 (2000), 1253–1270.
[8] A. Kartsatos, V.V. Kurta, *Nonexistence theorems for weak solutions of quasilinear elliptic equations*, Abstr. Appl. Anal. 6 (2001), 163–189.
[9] I. Kuzin, S. Pohozaev, *Entire solutions of semilinear elliptic equations*, Birkhäuser, 1997.

[10] G.I. Laptev, *Evolution equations with monotone operator and functional nonlinearity at the time derivative*, Mat. Sb. 191 (2000), 43–64.
[11] J.L. Lions, *Quelques Methods de Resolution des Problemes aux Limites Non-lineares*, Dunod, 1969.
[12] G.J. Minti, *Monotone (nonlinear) operators in Hilbert space*, Duke Math. J. 29 (1962), 341–346.
[13] E. Mitidieri, S.I. Pohozaev, *A priori Estimates and Nonexistence of Solutions to Nonlinear Partial Differential Equations and Inequalities*, Moscow, Nauka, 2001. (Proc. Steklov Inst. Math., Vol.234).
[14] A.E. Shishkov, *Support evolution of solution with unbounded energy for higher order quasilinear degenerate parabolic equations*, Mat. Sb. 186 (1995), 151–172.
[15] R.E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, Providence: AMS, 1997.
[16] I.V. Skrypnik, *Methods of Analysis of Nonlinear Elliptic Boundary Value Problems*, Translations of Math. Monographs, A.M.S. Providence, V.139 (1994).
[17] L. Veron, *Singularities of solutions of second order quasilinear euqations*, Addison Wesley Longman, 1996.
[18] M.I. Vishik, *Quasi-linear strongly elliptic systems of differential equations of divergence form (Russian)*, Trudy Moskov. Math. Obsch. 12 (1963), 125–184.

Some Recent Results and Problems for Set-Valued Mappings

Hong-Kun Xu
Department of Mathematics
University of Durban-Westville
Private Bag X54001, Durban 4000
South Africa
E-mail: hkxu@pixie.udw.ac.za

Abstract

We present some recent results for the existence of fixed points of set-valued contractions and nonexpansive mappings in metric and Banach spaces. The existence of Caristi selections is discussed and some open problems are raised.

Keywords: Set-valued mapping, contraction, directional contraction, nonexpansive mapping, fixed point, selection, inward set.

2000 Mathematics Subject Classification. Primary 47H04, 47H09, 47H10; Secondary 54C65.

1 Introduction

Since Nadler [29] extended Banach's contraction principle to set-valued contractions in 1969, fixed point theory for set-valued mappings has been developed rapidly. Like single-valued contractions, a lot of papers have been contributed to set-valued contractions in which the constant contraction coefficient in Nadler's theorem is relaxed, but the existence of fixed points is still guaranteed. The relaxed coefficient is a function defined in the underlying space. Due to the complexity of the images which the map takes, there are still no effective ways to prove the existence of a fixed point of a set-valued map of contractive type. For example, it is not easy to treat a set-valued mapping of contractive type if the mapping takes closed bounded (not necessarily compact) values (see Reich's open problem in section 6).

Martin [27] first proved that a nonexpansive compact-valued self-mapping of a closed bounded convex subset of a Hilbert space has a fixed point. This result was extended to a Banach space framework by Lami Dozo [22] in the case where the space satisfies the Opial property [30] (e.g., l^p for $1 < p < \infty$) and by Lim [23] in the case where the space is uniformly convex (e.g., both l^p and L^p for $1 < p < \infty$). But it still remains an open question whether a nonexpansive set-valued self-mapping of a closed bounded convex subset of a Hilbert space with closed bounded values has a fixed point.

The technique of asymptotic centers introduced by Edelstein [10] has been shown to be an effective tool in the existence theory for both single- and set-valued nonexpansive mappings. In a unifomly convex Banach space, the asymptotic center of a bounded sequence with respect to a closed convex set consists of exactly one point. Kirk and Massa [21] (partially) extended Lim's fixed point theorem [23] to the case where the underlying Banach space has the following property: The asymptotic center of a bounded sequence with respect to a closed convex set is nonempty and compact. This applies to a wider class of Banach spaces than the class of uniformly convex Banach spaces (e.g., the class of Banach spaces which are $k-$uniformly rotund in the sense of Sullivan [34]).

Another direction of the fixed point theory for set-valued mappings is the existence theory for nonself-mappings. Certain boundary conditions must then be satisfied. Among these are the inward and the weakly inward conditions (See Section 4 for definitions). Assad and Kirk [1] considered those nonself-contractions which keep the boundary of the domain in the domain. Lim [25] extended his own theorem [23] to those compact-valued nonexpansive mappings which are weakly inward. This result was independently regained by Xu [39] via using inequality techniques in uniformly convex Banach spaces (Xu [37]). Kirk-Massa's fixed point theorem is extended to nonself-mappings recently by Xu ([38, 39]). One must also mention a recent result by Lim [26] who proves that a weakly inward contraction which maps a closed set of a Banach space into the collection of nonempty closed subsets of the space has a fixed point. This result removes the assumptions that the values of the map are required to be either compact or proximinal in previous works (Deimling [8] and Xu [38, 39]).

The purpose of this paper is to give some recent results in the existence theory for set-valued contractions and nonexpansive mappings in either metric or Banach spaces. Some new results are also obtained and open problems are raised.

The paper is organized as follows. In Section 2 we introduce the notion of Caristi's selections and show that both contractions and generalized contractions admit Caristi selections. In Section 3 we show that a directional set-valued contraction has a fixed point. Section 4 is devoted to nonexpansive set-valued mappings. A new fixed point theorem for nonself-mappings is proved. In Section 5 we include Lim's recent fixed point theorem for nonself set-valued contractions. Finally in Section 6 we present several open problems for set-valued contractions and nonexpansive mappings.

2 Caristi's Selection

Let (M,d) be a complete metric space and let $f : M \to M$ be a mapping. We say that f is a *Caristi* mapping if there exists a lower semicontinuous function $\varphi : M \to \mathbf{R}$ such that φ is bounded below and satisfies

$$d(x, f(x)) \le \varphi(x) - \varphi(f(x)), \quad x \in M.$$

Lemma 2.1. (Caristi [4]) *Any Caristi mapping of a complete metric space admits a fixed point.*

We introduce some notation. We denote by 2^M the power set of M, by $CB(M)$ the family of nonempty closed bounded subsets of M, by $K(M)$ the family of nonempty compact subsets of M, and by H the *Hausdorff metric* on $CB(M)$. Thus we have for $A, B \in CB(M)$,

$$H(A,B) = \max\left\{\sup_{a\in A} d(a,B), \sup_{b\in B} d(b,A)\right\},$$

where $d(x,D) := \inf\{d(x,y) : y \in D\}$ is the distance from a point x in M to a subset D of M.

Let now $T : M \to CB(M)$ be a set-valued mapping. By a *selection* of T we mean a (single-valued) mapping $f : M \to M$ such that

$$f(x) \in Tx, \quad x \in M.$$

A selection f for T is said to be a *Caristi* selection for T if it is also a Caristi mapping. Note that a Caristi selection may not be continuous. It is also immediately clear that a Caristi set-valued mapping always admits a fixed point. The converse is however not true. Namely, the existence of a fixed point of a set-valued mapping does not guarantee the existence of a Caristi selection for the map. The following is a counterexample.

Example. Let $M = [0,\infty)$ be equipped with the usual distance and let $k > 1$ be a number. Let $T : M \to CB(M)$ be defined by

$$Tx := [kx, (k+1)x], \quad x \in M.$$

It is seen that 0 is the only fixed point of T and $T(0) = \{0\}$. We now show that T fails to admit a Caristi selection. Suppose the contrary that f is a Caristi selection for T. Then we have

$$|x - f(x)| \le \varphi(x) - \varphi(f(x)), \quad x \in M, \tag{2.1}$$

where $\varphi : M \to \mathbf{R}$ is a lower semicontinuous function bounded below. Since $kx \le f(x) \le (k+1)x$ for $x \ge 0$, we have by (2.1)

$$(k-1)x \le \varphi(x) - \varphi(f(x)), \quad x > 0. \tag{2.2}$$

Fix $x > 0$ and substitute $f(x)$ for x in (2.2) to get

$$(k-1)f(x) \leq \varphi(f(x)) - \varphi(f^2(x)).$$

Continue this way to obtain

$$(k-1)f^n(x) \leq \varphi(f^n(x)) - \varphi(f^{n+1}(x)), \quad n \geq 0, \tag{2.3}$$

where f^n is the n-th composite of f. Summing up (2.3) yields (since φ is bounded below)

$$\sum_{n=0}^{\infty} f^n(x) < \infty.$$

In particular,

$$\lim_{n\to\infty} f^n(x) = 0. \tag{2.4}$$

On the other hand, however, it is easily seen that the sequence $\{f^n(x)\}$ is strictly increasing and so (2.4) can not hold. This contradiction shows that f can not be a Caristi selection for T.

Since a set-valued mapping does not always have a continuous selection ([29], [19]), one would turn his attention to look for a Caristi selection. It is however obvious that such a selection does not exist if the set-valued mapping is fixed point free.

The next result shows that a set-valued contraction has a Caristi selection. Recall that a set-valued mapping $T : M \to CB(M)$ is said to be a *contraction* if there exists a number $k \in [0, 1)$ such that

$$H(Tx, Ty) \leq kd(x, y), \quad x, y \in M.$$

Theorem 2.2. (cf. [18]) *Let (M, d) be a complete metric space and $T : M \to CB(M)$ a contraction. Then T admits a Caristi selection.*

Proof. Let $\varepsilon > 0$ be small enouough so that $k + \varepsilon < 1$ and define φ by

$$\varphi(x) := \frac{1}{\varepsilon} d(x, Tx), \quad x \in M.$$

It is immediately clear that φ is continuous. For any $x \in M$ we can find some $f(x) \in Tx$ satsifying

$$d(x, f(x)) \leq \frac{1}{k+\varepsilon} d(x, Tx). \tag{2.5}$$

Obviously f is a selection for T. It remains to show that

$$d(x, f(x)) \leq \varphi(x) - \varphi(f(x)), \quad x \in M.$$

Note that

$$d(f(x), Tf(x)) \le H(Tx, Tf(x)) \le kd(x, f(x)). \tag{2.6}$$

We calculate

$$\begin{aligned} d(x, f(x)) &= \frac{1}{\varepsilon}[(k+\varepsilon)d(x, f(x)) - kd(x, f(x))] \\ &\le \frac{1}{\varepsilon}[d(x, T(x)) - d(f(x), Tf(x))] \qquad \text{(by (2.5)-(2.6))} \\ &= \varphi(x) - \varphi(f(x)). \end{aligned}$$

■

Recall now that a set-valued mapping $T : M \to CB(M)$ is said to be a *generalized contraction* if there exists a function $k : M \to [0, 1)$ such that

$$H(Tx, Ty) \le k(x)d(x, y), \quad x, y \in M. \tag{2.7}$$

Theorem 2.3. *Let (M, d) be a complete metric space and $T : M \to CB(M)$ a generalized contraction. If the function k in (2.7) is continuous, then T admits a Caristi selection.*

Proof. Let

$$\varepsilon(x) := \frac{1 - k(x)}{2}, \quad x \in M.$$

Then $k(x) + \varepsilon(x) = (k(x) + 1)/2 < 1$ for $x \in M$. Define φ by

$$\varphi(x) := \frac{1}{\varepsilon(x)} d(x, Tx), \quad x \in M.$$

Since k is continuous, φ is continuous and bounded below (by 0). For any $x \in M$ we can find some $f(x) \in Tx$ satsifying

$$d(x, f(x)) \le \frac{1}{k(x) + \varepsilon(x)} d(x, Tx). \tag{2.8}$$

This f is a selection of T. In order to show that f is a Caristi selection for T, it remains to show that

$$d(x, f(x)) \le \varphi(x) - \varphi(f(x)), \quad x \in M.$$

Note that

$$d(f(x), Tf(x)) \le H(Tx, Tf(x)) \le k(x)d(x, f(x)). \tag{2.9}$$

We calculate

$$\begin{aligned} d(x, f(x)) &= \frac{1}{\varepsilon(x)}[(k(x)+\varepsilon(x))d(x, f(x)) - k(x)d(x, f(x))] \\ &\le \frac{1}{\varepsilon(x)}[d(x, T(x)) - d(f(x), Tf(x))] \qquad \text{(by } (2.8) - (2.9)\text{)} \\ &= \varphi(x) - \varphi(f(x)). \end{aligned}$$

■

3 Directional Set-Valued Contractions

Let (M,d) be a metric space. Given points $x, y \in M$, the *open segment* (x,y) defined by x and y is the set of points z (if any) in M distinct from x and y and satisfying

$$d(x,z) + d(z,y) = d(x,y). \tag{3.1}$$

The notion of directional (single-valued) contractions was introduced by Clarke [6].

Definition 3.1. A (single-valued) map $f : M \to M$ is said to be a *directional contraction* provided f is continuous and there exists a number $\sigma \in (0,1)$ with the following property: whenever $v \in M$ is such that $f(v) \neq v$, there exists $w \in (v, f(v))$ such that

$$d(f(v), f(w)) \leq \sigma d(v,w).$$

We next extend this notion to set-valued mappings.

Definition 3.2. A set-valued map $T : M \to CB(M)$ is said to be a *directional contraction* provided T is upper semicontinuous with respect to the Hausdorff distance H and there exists a number $\sigma \in (0,1)$ with the following property: whenever $v \in M$ is such that $v \notin T(v)$ and $u \in T(v)$, there exists $w \in (v,u)$ such that

$$H(T(v), T(w)) \leq \sigma d(v,w). \tag{3.2}$$

Clarke [6] shows that every directional single-valued contraction of a complete metric space has a fixed point. We next extend Clarke's result to the set-valued case. We need Ekeland's ε-Variational Principle [11, 12].

Lemma 3.3. (Ekeland's ε-Variational Principle) *Let (M,d) be a complete metric space and let $F : M \to \mathbf{R} \cup \{+\infty\}$ be a lower semicontinuous function which is bounded below. If u is a point in M satisfying*

$$F(u) < \inf_M F + \varepsilon$$

for some $\varepsilon > 0$, then, for every $\lambda > 0$ there exists a point v in M such that

(i) $F(v) \leq F(u)$.

(ii) $d(u,v) \leq \lambda$.

(iii) For all $w \neq v$ in M, one has

$$F(v) < F(w) + \frac{\varepsilon}{\lambda} d(v,w).$$

Theorem 3.4. *Let (M,d) be a complete metric space and $T : M \to K(M)$ be a directional set-valued contraction. Then T has a fixed point.*

Proof. Let $\varphi : M \to \mathbf{R}$ be defined by

$$\varphi(x) = d(x, Tx).$$

Since T is upper semicontinuous, φ is lower semicontinuous (and bounded below obviously). By Lemma 3.3 (with $\varepsilon = (1-\sigma)/2$ and $\lambda = 1$), we have a point v in M such that, for all $w \in M$ one has

$$d(v, Tv) \leq d(w, Tw) + \frac{1-\sigma}{2} d(w, v). \tag{3.3}$$

If $v \in Tv$, we are done. So suppose the contrary that $v \notin Tv$. Noting that Tv is compact, we can find a point u in Tv such that

$$d(v, u) = d(v, Tv). \tag{3.4}$$

Since T is a directional contraction, we can also find a point $w \in (u, v)$ such that

$$H(Tu, Tv) \leq \sigma d(v, w). \tag{3.5}$$

Note that there holds

$$d(v, w) + d(w, u) = d(v, u). \tag{3.6}$$

Note also the triangle inequality

$$d(w, Tw) \leq d(w, Tu) + H(Tv, Tw). \tag{3.7}$$

We calculate

$$\begin{aligned} 0 &\leq \sigma d(v, w) - H(Tv, Tw) \qquad \text{(by (3.5))} \\ &\leq \sigma d(v, w) - d(w, Tw) + d(w, Tv) \qquad \text{(by (3.7))} \\ &\leq \sigma d(v, w) - d(w, Tw) + d(w, u) \qquad (u \in Tv) \\ &= (\sigma - 1) d(v, w) - d(w, Tw) + d(v, u) \qquad \text{(by (3.6))} \\ &= (\sigma - 1) d(v, w) - d(w, Tw) + d(v, Tv) \qquad \text{(by (3.4))} \\ &\leq \frac{\sigma - 1}{2} d(v, w) \qquad \text{(by (3.3))}. \end{aligned}$$

Since $\sigma < 1$, this implies $w = v$. This contradiction proves the theorem. ■

4 Nonexpansive Set-Valued Mappings

Let X be a Banach space and E a nonempty closed convex subset of X. Recall that $K(E)$ is the family of nonempty compact subsets of E. We shall use $K_c(E)$ to denote the family of nonempty compact convex subsets of E. A set-valued mapping $T : E \to K(X)$ is said to be *nonexpansive* if, for all $x, y \in E$,

$$H(Tx, Ty) \leq \|x - y\|.$$

Given a bounded sequence (x_n) in X, the *asymptotic radius* and *center* of (x_n) with respect to E are defined respectively by

$$r_E(x_n) := \inf \left\{ \limsup_{n\to\infty} \|x_n - x\| : x \in E \right\}$$

and

$$A_E(x_n) := \left\{ x \in E : \limsup_{n\to\infty} \|x_n - x\| = r_E(x_n) \right\}.$$

(The notion of asymptotic center was introduced by Edelstein [10].) Recall that the *inward set* to E at a point x in E is defined as

$$I_E(x) := \{x + \lambda(y - x) : \lambda \geq 1, y \in E\}.$$

Recall also that a set-valued map $T : E \to K(X)$ is said to be *inward* provided, for each x in E one has

$$Tx \subset I_E(x).$$

T is *weakly inward* if, for each x in E one has

$$Tx \subset \overline{I_E(x)},$$

where $\overline{I_E(x)}$ is the closure of $I_E(x)$.

Recall that a bounded sequence (x_n) in X is called *regular* with respect to E if $r_E(x_n) = r_E(x_{n_i})$ for all subsequences (x_{n_i}) of (x_n); while (x_n) is called *asymptotically uniform* if $A_E(x_n) = A_E(x_{n_i})$ for all subsequences (x_{n_i}) of (x_n).

Lemma 4.1. ([24], [13]) *Let (x_n) and E be as above.*

1. *There always exists a subsequence of (x_n) which is regular with respect to E;*
2. *If E is separable, then (x_n) contains a subsequence which is asymptotically uniform with respect to E.*

One of the fundamental results in the fixed point theory for set-valued mappings is the following theorem proved by Lim in 1974. The random version was proved by Xu [36] in 1993.

Theorem 4.2. (Lim [23]) *If X is a uniformly convex Banach space and if E is a nonempty closed bounded convex subset of X, then every nonexpansive set-valued mapping $T : E \to K(E)$ has a fixed point.*

One of the features of a uniformly convex Banach space X is that the asymptotic center of any bounded sequence with respect to a closed convex subset of X consists of one and only one point. Kirk and Massa [21] extended Theorem 4.2 in the sense that the asymptotic center is allowed to contain more than one point (a compact set, to be precise). However, T is in addition assumed to take convex values.

Theorem 4.3. (Kirk-Massa [21]) *Let E be a nonempty closed bounded convex subset of a Banach space X and $T : E \to K_c(E)$ a nonexpansive mapping. Suppose that the asymptotic center in E of each bounded sequence of X is nonempty and compact. Then T has a fixed point.*

Both Theorems 4.2 and 4.3 have been extended to nonself mappings.

Theorem 4.4. (Lim [25] and Xu [38, 39]) *Assume X is a uniformly convex Banach space, E is a closed bounded convex subset of X, and $T : E \to K(X)$ is a nonexpansive weakly inward set-valued mapping. Then T has a fixed point.*

Theorem 4.5. (Xu [38, 39]) *Let E be a nonempty closed bounded convex subset of a Banach space X and $T : E \to K_c(X)$ a nonexpansive inward set-valued mapping. Suppose that the asymptotic center in E of each bounded sequence of X is nonempty and compact. Then T has a fixed point.*

Now recall that the modulus of noncompact convexity of a Banach space X is defined as a function $\Delta_X : [0, 2] \to [0, 1]$ by

$$\Delta_X(\varepsilon) := 1 - \sup\left\{\inf_{x\in A} \|x\| : A \subset B_X \text{ convex}, \alpha(A) \geq \varepsilon\right\},$$

where B_X is the closed unit ball of X and $\alpha(A)$ is the *Kuratowski measure of noncompactness* of A; that is,

$$\begin{aligned}\alpha(A) \quad := \quad & \inf\{r > 0 : A \text{ can be covered with a finite family} \\ & \qquad\qquad \text{of subsets of diameters less than } r\}.\end{aligned}$$

Let

$$\varepsilon_X := \sup\{\varepsilon > 0 : \Delta_X(\varepsilon) = 0\}.$$

It is known that the space X is *nearly uniformly convex* ([17], [14]) if and only if $\varepsilon_X = 0$.

Recall also that the *Chebyshev radius* of a bounded set E with respect to another set G is defined by

$$\tilde{r}_G(E) := \inf \left\{ \sup_{y \in E} \|x - y\| : x \in G \right\}.$$

The following lemma establishes a connection, through the modulus Δ_X of X, of the asymptotic radius of a bounded sequence (x_n) with respect to E and the Chebyshev radius of the asymptotic center of the sequence (x_n) with respect to E.

Lemma 4.6. (cf. Goebel-Kirk [14]) *Let X be a reflexive Banach space, E a closed convex subset of X, and (x_n) a bounded sequence in E. Then*

$$\tilde{r}_E(A_E(x_n)) \leq [1 - \Delta_X(1^-)]\, r_E(x_n),$$

where $\Delta_X(1^-) = \lim_{\varepsilon \to 1^-} \Delta_X(\varepsilon)$.

We next recall the *Hausdorff measure of noncompactness* of a bounded subset A of a Banach space X is defined by

$$\begin{aligned} \chi(A) \quad := \quad & \inf\{r > 0 : A \text{ can be covered with a finite family} \\ & \qquad \text{of balls of radius less than } r\}. \end{aligned}$$

Let γ be either α or χ. A set-valued map $T : E \to CB(X)$ is said to be *γ-condensing* if, whenever $A \subset E$ is such that $\gamma(A) > 0$, one has

$$\gamma(T(A)) < \gamma(A), \tag{4.1}$$

where $T(A) = \cup\{Tx : x \in A\}$. If the nonstrict inequality $\leq$ holds in (4.1), then T is said to be *1-γ-contractive.*

Lemma 4.7. (cf [8]) *Let X be a Banach space and E a closed bounded convex subset of X. Let $F : E \to 2^X$ be a set-valued mapping with nonempty closed convex values. Then F has a fixed point if either one of the following conditions is satisfied:*

1. *F is a weakly inward contraction;*

2. *F is an upper semicontinuous γ-condensing mapping and satisfies the property: $F \cap \overline{I_E(x)} \neq \emptyset$ for every $x \in E$.*

Recently Dominguez Benavides and Lorenzo Ramirez proved the following interesting result.

Theorem 4.8. ([9]) *Let X be a Banach space such that $\varepsilon_X < 1$, E a nonempty closed bounded convex subset of X, and $T : E \to K_c(E)$ a nonexpansive set-valued mapping. Assume T is also 1-χ-contractive. Then T has a fixed point.*

The next result extends Theorem 4.8 to the nonself-mapping case.

Theorem 4.9. *Let X be a Banach space such that $\varepsilon_X < 1$, E a nonempty closed bounded convex subset of X, and $T : E \to K_c(X)$ a nonexpansive inward set-valued mapping. Assume T is also 1-χ-contractive. Then T has a fixed point.*

Proof. Fix an $x_0 \in E$ and define for each integer $n \geq 1$ the contraction $T_n : E \to K_c(X)$ by

$$T_n(x) := \frac{1}{n}x_0 + \left(1 - \frac{1}{n}\right)Tx, \quad x \in E.$$

Then T_n satisfies the inwardness condition, i.e., $T_n x \subset I_E(x)$ for all $x \in E$. Thus by Lemma 4.7, T_n has a fixed point $x_n \in E$. By Lemma 4.1, we may assume that (x_n) is regular. Let $y_n \in Tx_n$ be such that $\|x_n - y_n\|$=dist$(x_n, Tx_n) \to 0$ as $n \to \infty$. Let (n_α) be a universal subnet of (n) (see [20] for more details about universal nets) and define a function g by

$$g(x) := \lim_\alpha \|x_{n_\alpha} - x\|, \quad x \in E.$$

Let

$$A := \{x \in E : g(x) = r\},$$

where $r = \inf_{x \in E} g(x)$. Then by assumption (see Proposition 6 of [21]), A is nonempty and compact. The key to the proof is that the inwardness of T on E implies that

$$Tx \cap I_A(x) \neq \emptyset, \quad x \in A. \tag{4.2}$$

Indeed, if $x \in A$, by compactness, we have for each $n \geq 1$, some $z_n \in Tx$ such that

$$\|y_n - z_n\| = d(y_n, Tx) \leq H(Tx_n, Tx) \leq \|x_n - x\|. \tag{4.3}$$

Let $z = \lim_\alpha z_{n_\alpha}$. Then $z \in Tx$. So in order to show (4.2), we only need to show $z \in I_A(x)$. We notice that (4.3) implies

$$g(z) = \lim_\alpha \|x_{n_\alpha} - z\| = \lim_\alpha \|y_{n_\alpha} - z_{n_\alpha}\| \leq \lim_\alpha \|x_{n_\alpha} - x\|.$$

Hence

$$g(z) \leq g(x) = r. \tag{4.4}$$

Furthermore, as $Tx \subset I_E(x)$, we have some $\lambda \geq 0$ and $v \in E$ such that

$$z = x + \lambda(v - x).$$

What we need to show is that $v \in A$. If $\lambda \leq 1$, then by the convexity of E, $z \in E$ and hence by (4.4), $z \in A \subset I_A(x)$ and we are done. So assume $\lambda > 1$. Then we can write

$$v = \mu z + (1 - \mu)x \quad \text{with } \mu = \frac{1}{\lambda} \in (0, 1).$$

By the convexity of g, we have by (4.4),

$$g(v) \leq \mu g(z) + (1-\mu)g(x) \leq r.$$

Since $v \in E$, this implies $v \in A$. Hence $z = x + \lambda(v - x) \in I_A(x)$. Now we have a nonexpansive mapping $T : A \to K_c(X)$ which satisfies the (boundary) condition (4.2). For each $n \geq 1$ we define another contraction $S_n : A \to K_c(X)$ by

$$S_n x := \frac{1}{n}x_0 + (1 - \frac{1}{n})Tx, \quad x \in A,$$

where $x_0 \in A$. Since (4.2) implies that S_n satisfies the same boundary condition; i.e.,

$$S_n x \cap I_A(x) \neq \emptyset, \quad x \in A.$$

Also by assumption we see that S_n is a continuous χ-condensing mapping. An application of Lemma 4.7 yields a fixed point $x_n^1 \in A$ of S_n. It is easy to see that

$$d(x_n^1, Tx_n^1) \leq \frac{\sigma}{n}, \quad n \geq 1,$$

where $\sigma > 0$ is a constant. Set $A^1 := A_E(x_n^1)$. Apply Lemma 4.6 to get

$$\tilde{r}_E(A^1) \leq \lambda r_E(x_n^1), \tag{4.5}$$

where $\lambda = 1 - \Delta_X(1^-) < 1$. Note that $r_E(x_n^1) \leq \tilde{r}_E(A)$ as $(x_n^1) \subset A^0 := A$. We obtain by (4.5)

$$\tilde{r}_E(A^1) \leq \lambda \tilde{r}_E(A^0). \tag{4.6}$$

Now by induction we can construct, for each integer $k \geq 1$, a sequence (x_n^k) and a nonempty subset A^k such that, for each $k \geq 1$,

(1) $(x_n^k) \subset A^{k-1}$;

(2) $A^k = A_E(x_n^k)$;

(3) $\lim_{n\to\infty} d(x_n^k, Tx_n^k) = 0$;

(4) $\tilde{r}_E(A^k) \leq \lambda^k \tilde{r}_E(A)$.

Pick a $v_k \in A^k$ for each $k \geq 1$ to get a sequence (v_k). Since

$$\begin{aligned}
\|v_{k+1} - v_k\| &\leq \limsup_{n\to\infty}(\|x_n^{k+1} - v_k\| + \|v_{k+1} - x_n^{k+1}\|) \\
&\leq \operatorname{diam} A^k + \limsup_{n\to\infty} \|x_n^{k+1} - v_{k+1}\| \\
&= \operatorname{diam} A^k + r_E(x_n^{k+1}) \\
&\leq \operatorname{diam} A^k + \tilde{r}_E(A^k) \\
&\leq 3\tilde{r}_E(A^k) \leq 3\lambda^k \tilde{r}_E(A).
\end{aligned}$$

Since $\lambda < 1$, it follows that (v_k) is strongly convergent. Let v be the limit of (v_k). Since we have

$$\begin{aligned} d(v_k, Tv_k) &\leq \|v_k - x_n^k\| + d(x_n^k, Tx_n^k) + H(Tx_n^k, Tv_k) \\ &\leq 2\|v_k - x_n^k\| + d(x_n^k, Tx_n^k), \end{aligned}$$

we obtain by taking the limsup as $n \to \infty$

$$d(v_k, Tv_k) \leq 2 \cdot \limsup_{n\to\infty} \|v_k - x_n^k\| \leq 2 \cdot \lambda^{k-1} \tilde{r}_E(A).$$

Letting $k \to \infty$ yields $d(v, Tv) = 0$ and hence $v \in Tv$. The proof is complete. ■

Recall that a Banach space X is said to satisfy the *nonstrict Opial property* (cf [22]) if, whenever a sequence (x_n) is weakly convergent to a point x_∞, one has

$$\limsup_{n\to\infty} \|x_n - x_\infty\| \leq \limsup_{n\to\infty} \|x_n - x\| \text{for all } x \in X.$$

If a Banach space X satisfies the nonstrict Opial property, then every nonexpansive set-valued mapping $T : E \to K(X)$ is 1-χ-contractive (see [9] for a proof). We therefore have the following result.

Corollary 4.10. *Let X be a Banach space such that $\varepsilon_X < 1$ and satisfy the nonstrict Opial property, E a nonempty closed bounded convex subset of X, and $T : E \to K_c(X)$ a nonexpansive inward set-valued mapping. Then T has a fixed point.*

5 Transfinite Induction and Lim's Theorem

Transfinite induction plays an important role in the existence theory for fixed points of set-valued mappings. As a matter of fact, Lim's theorem (Theorem 4.2) and Caristi's theorem (Lemma 2.1) were first proved by using the transfinite induction technique (cf. [23], [4]). The next theorem, again due to Lim, is another example of applications of the transfinite induction in the fixed point theory.

Theorem 5.1. ([26]) *Let E be a closed subset of a Banach space X and $T : E \to 2^X \setminus \{\emptyset\}$ be a contraction taking nonempty closed values. If T is weakly inward (i.e., $Tx \subset \overline{I_E(x)}$ for $x \in E$), then T has a fixed point.*

Proof. Let $k \in [0, 1)$ be the contraction constant of T. Pick l, $k < l < 1$ and $\varepsilon \in (0, 1)$ so that $b := \frac{1-\varepsilon}{1+\varepsilon} - l > 0$. Assume on the contrary that T does not have fixed points. Take $z_0 \in E$ and $y_0 \in Tz_0$ arbitrarily. Let Ω be the first noncountable ordinal and γ an ordinal $< \Omega$. Suppose z_α, y_α have been defined for all $\alpha < \gamma$ such that

(i) $y_\alpha \in Tz_\alpha$ for $\alpha < \gamma$,

(ii) $z_\alpha \neq z_{\alpha+1}$ for $\alpha < \alpha + 1 < \gamma$,

(iii) $b \max\{\|z_\beta - z_\alpha\|, \frac{1}{l}\|y_\beta - y_\alpha\|\} \leq \|y_\alpha - z_\alpha\| - \|y_\beta - z_\beta\|$ for $\alpha, \beta < \gamma$.

We next define z_γ, y_γ so that (i)-(iii) remain valid for all $\alpha, \beta < \gamma + 1$. We shall distinguish two cases.

Case 1. γ has a predecessor $\gamma - 1$. In this case, since $y_{\gamma-1} \in Tz_{\gamma-1}$ and T is fixed point free, we see $\|y_{\gamma-1} - z_{\gamma-1}\| > 0$. By the weak inwardness of T we have $z_\gamma \in E$ and $\lambda \geq 1$ such that

$$\|y_{\gamma-1} - (z_{\gamma-1} + \lambda_\gamma(z_\gamma - z_{\gamma-1}))\| \leq \varepsilon\|y_{\gamma-1} - z_{\gamma-1}\|.$$

This clearly implies that ${}_{\gamma-1} \neq z_\gamma$ and

$$\begin{aligned} \|z_\gamma - z_{\gamma-1}\| &\leq (1+\varepsilon)\mu_\gamma\|y_{\gamma-1} - z_{\gamma-1}\|, \\ \|z_\gamma - x_\gamma\| &\leq \varepsilon\mu_\gamma\|y_{\gamma-1} - z_{\gamma-1}\|, \end{aligned}$$

where $\mu_\gamma = \frac{1}{\lambda_\gamma}$ and $x_\gamma = \mu_\gamma y_{\gamma-1} + (1-\mu_\gamma)z_{\gamma-1}$. Since $H(Tz_\gamma, Tz_{\gamma-1}) \leq k\|z_\gamma - z_{\gamma-1}\|$, there is some $y_\gamma \in Tz_\gamma$ such that

$$\|y_\gamma - y_{\gamma-1}\| \leq l\|z_\gamma - z_{\gamma-1}\|. \tag{5.1}$$

Thus

$$\begin{aligned} \|y_\gamma - z_\gamma\| &\leq \|y_\gamma - y_{\gamma-1}\| + \|y_{\gamma-1} - x_\gamma\| + \|x_\gamma - z_\gamma\| \\ &\leq l\|z_\gamma - z_{\gamma-1}\| + (1-\mu_\gamma)\|y_{\gamma-1} - z_{\gamma-1}\| + \varepsilon\mu_\gamma\|y_{\gamma-1} - z_{\gamma-1}\| \\ &\leq l\|z_\gamma - z_{\gamma-1}\| + \|y_{\gamma-1} - z_{\gamma-1}\| - \frac{1-\varepsilon}{1+\varepsilon}\|z_\gamma - z_{\gamma-1}\|. \end{aligned}$$

It then follows that

$$b\|z_\gamma - z_{\gamma-1}\| \leq \|y_{\gamma-1} - z_{\gamma-1}\| - \|y_\gamma - z_\gamma\|$$

and from (5.1)

$$\frac{b}{l}\|y_\gamma - y_{\gamma-1}\| \leq \|y_{\gamma-1} - z_{\gamma-1}\| - \|y_\gamma - z_\gamma\|.$$

For any $\alpha < \gamma - 1$,

$$b\|z_\alpha - z_{\gamma-1}\| \leq \|y_\alpha - z_\alpha\| - \|y_{\gamma-1} - z_{\gamma-1}\| \qquad \text{(by (iii))}.$$

So

$$\begin{aligned} b\|z_\gamma - z_\alpha\| &\leq b(\|z_\gamma - z_{\gamma-1}\| + \|z_{\gamma-1} - z_\alpha\|) \\ &\leq \|y_\alpha - z_\alpha\| - \|y_\gamma - z_\gamma\|. \end{aligned}$$

Similarly
$$\frac{b}{l}\|y_\gamma - y_\alpha\| \le \|y_\alpha - z_\alpha\| - \|y_\gamma - z_\gamma\|.$$
So (i)-(iii) are valid for $\alpha, \beta < \gamma + 1$.

Case 2. γ is a limit ordinal. We then have a strictly increasing sequence (γ_n) that converges to γ. Set $r_n = \|y_{\gamma_n} - z_{\gamma_n}\|$. Condition (iii) then implies that (z_{γ_n}) and (y_{γ_n}) are both Cauchy and hence convergent. Let z_γ and y_γ be their respective limits. Since $y_{\gamma_n} \in Tz_{\gamma_n}$, we have a $w_n \in Tz_\gamma$ such that $\|w_n - y_{\gamma_n}\| \le l\|z_{\gamma_n} - z_\gamma\|$. Thus $w_n - y_{\gamma_n} \to 0$. As $y_{\gamma_n} \to y_\gamma$, we get $w_n \to y_\gamma$ and thus $y_\gamma \in Tz_\gamma$ for Tz_γ is closed. Now for $\alpha < \gamma$, we have $\gamma_n > \alpha$ for sufficiently large n, so
$$b\|z_{\gamma_n} - z_\alpha\| \le \|y_\alpha - z_\alpha\| - \|y_{\gamma_n} - z_{\gamma_n}\|$$
and upon taking limits,
$$b\|z_\gamma - z_\alpha\| \le \|y_\alpha - z_\alpha\| - \|y_\gamma - z_\gamma\|.$$
Similarly
$$\frac{b}{l}\|y_\gamma - y_\alpha\| \le \|y_\alpha - z_\alpha\| - \|y_\gamma - z_\gamma\|.$$
Therefore, (i)-(iii) remain valid for all $\alpha, \beta < \gamma+1$. If $\alpha < \alpha+1 < \gamma+1$, then $\alpha < \gamma$. Since γ is a limit ordinal, $\alpha + 1 < \gamma$. So (ii) is also valid for $\alpha < \alpha + 1 < \gamma + 1$.

By the transfinite induction, z_α, y_α for $\alpha < \Omega$ satisfying (i)–(iii) have been defined. Let $s_\alpha = \|y_\alpha - z_\alpha\|$. Since $(s_\alpha)_{\alpha<\Omega}$ is decreasing and bounded below by 0, it must eventually be constant. If $\gamma < \Omega$ is such that $s_\alpha = s_\beta$ for all $\alpha, \beta \ge \gamma$, then by (iii) $z_{\gamma+1} = z_\gamma$, contradicting (ii). Therefore, T must have a fixed point. ■

6 Some Open Problems

There are still a lot of basic problems in the fixed point theory for set-valued mappings which remain unsolved. Below is a partial list of some open problems. See also the open problems raised in Reich [33] and Xu [38].

Problem 1. Assume (M, d) is a complete metric space and $T : M \to CB(M)$ (or even $K(X)$) has a Caristi selection. Does there exist an equivalent metric on M under which T becomes a contraction?

Problem 2. Assume (M, d) is a complete metric space and $T : M \to K(X)$ is a directional set-valued contraction. Does T admit a Caristi selection?

Problem 3. Can one find a proof of Theorem 5.1 without using the transfinite induction?

Problem 4. Does a set-valued weakly inward (or even inward) contraction which maps a closed subset of a Banach space into the collection of nonempty compact (or even compact convex) subsets have a Caristi selection?

Problem 5. In Theorems 4.5 and 4.9, can the inwardness condition be weakened to the weak inwardness condition?

Problem 6. Assume E is a closed bounded convex subset of a uniformly convex Banach space X and assume $T : E \to K_c(X)$ is a set-valued contraction such that $Tx \cap \overline{I_E(x)} \neq \emptyset$ (or even $Tx \cap I_E(x) \neq \emptyset$) for all $x \in E$. Does T have a fixed point?

The answer to this problem is no if T is not assumed to take convex values. A counterexample has been constructed in the plane $\mathbf{R}^2$ (see [39], page 700).

Problem 7. In Theorem 4.9, can one remove the assumption that the mapping T is 1-β-contractive? In particular, can one remove the non-strict Opial assumption in Corollary 4.10?

Problem 8. (Reich's problem [32, 33]) Assume (M, d) is a complete metric space and assume $T : M \to CB(X)$ is a set-valued mapping satisfying the condition:

$$H(Tx, Ty) \leq k(d(x, y))d(x, y), \quad x, y \in X, \ x \neq y,$$

where $k : (0, \infty) \to (0, 1)$ is a function with the following property:

$$\limsup_{s \to t^+} k(s) < 1 \quad \text{for all } t > 0.$$

Does T have a fixed point?

Hu [16] claimed that the answer to Problem 8 is yes. But his proof contains a gap (see Reich [33] and Jachymski [19]). (Unfortunately this gap has been repeated in several papers, e.g. [2, 3].)

Though problem 8 remains unsolved, some partial answers have been obtained. Precisely, the asnwer to this problem is yes if any one of the following additional conditions is satisfied:

1. For each $x \in M$, Tx is a compact set. ([31], see also [38])
2. $\limsup_{s \to 0^+} k(s) < 1$. ([35], [28])
3. There are constants $a, t_0 > 0$ and $\sigma \in (0, 1)$ such that $k(t) \leq 1 - at^\sigma$ for all $t \in (0, t_0)$. ([15], [18], [7])
4. Whenever H is a closed subset of M such that $Tx \cap H \neq \emptyset$ for all $x \in H$, it follows that $d(x, Tx \cap H) = d(x, Tx)$ for all $x \in H$. ([5]; see also [38] for a remark on this condition.)

References

[1] N.A. Assad and W.A. Kirk, Fixed point theorems for set-valued mappings of contractive type, *Pacific J. Math.* 43(1972), 553-562.

[2] I. Beg and A. Azam, Fixed points of multivalued locally contractive mappings, *Bullettino U.M.I.* A (7) 4(1990), 227-233.

[3] I. Beg and A. Azam, Fixed points of asymptotically regular multivalued mappings, *J. Austral. Math. Sco. Ser. A* 53(1992), 313-326.

[4] J. Caristi, Fixed point theorems for mappings satisfying inwardness conditions, *Trans. Amer. Math. Soc.* 215(1976), 241–251.

[5] Y.Q. Chen, On a fixed point problem of Reich, *Proc. Amer. Math. Soc.* 124(1996), 3085–3088.

[6] F.H. Clarke, Optimization and Nonsmooth Analysis, John Wiley & Sons, 1983.

[7] P.D. Daffer, H. Kaneko and W. Li, On a conjecture of S. Reich, *Proc. Amer. Math. Soc.* 124(1996), 3159–3162.

[8] K. Deimling, "Multivalued Differential Equations," Walter de Gruyter, Berlin/New York, 1992.

[9] T. Dominguez Benavides and P. Lorenzo Ramirez , Deterministic and random fixed point for multivalued nonexpansive mappings without uniform convexity. (Preprint)

[10] M. Edelstein, The construction of an asymptotic center with a fixed point property, *Bull. Amer. Math. Soc.* 78(1972), 206–208.

[11] I. Ekeland, On the variational principle, *J. Math. Anal. Appl.* 47(1974), 324-353.

[12] I. Ekeland, Nonconvex minimization problems, *Bull. Amer. Math. Soc.* 1(1979), 443-474.

[13] K. Goebel, On a fixed point theorem for multivalued nonexpansive mappings, *Annales Univ. Maria Curie-Sklodowska* 29(1975), 70–72.

[14] K. Goebel and W.A. Kirk, "Topics in Metric Fixed Point Theory," Cambridge University Press, 1990.

[15] C.L. Guo, Fixed point theorems for singlevalued mappings and multivalued mappings on complete metric spaces, *Chinese. J. Math.* 19(1991), 31-53.

[16] T. Hu, Fixed point theorems for multi-valued mappings, *Canad. Math. Bull.* 23(1980), 193–197.

[17] R.E. Huff, Banach spaces which are nearly uniformly convex, *Rocky Mountain J. Math.* 10(1980), 743–749.

[18] J.R. Jachymski, Caristi's fixed point theorem and selections of set-valued contractions, *J. Math. Anal. Appl.* 227(1998), 55–67.

[19] J.R. Jachymski, On Reich's question cerncerning fixed points of multimaps, *Bullettino U.M.I.* (7) 9-A(1995), 453-460.

[20] J. L. Kelly, "General Topology," van Nostrand, Princeton, NJ, 1955.

[21] W.A. Kirk and S. Massa, Remarks on asymptotic and Chebyshev centers, *Houston J. Math.* 16(1990), 357–364.

[22] E. Lami Dozo, Multivalued nonexpansive mappings and Opial's condition, *Proc. Amer. Math. Soc.* 38(1973), 286–292.

[23] T.C. Lim, A fixed point theorem for multivalued nonexpansive mappings in a uniformly convex Banach space, *Bull. Amer. Math. Soc.* 80(1974), 1123–1126.

[24] T.C. Lim, Remarks on some fixed point theorems, *Proc. Amer. Math. Soc.* 60(1976), 179–182.

[25] T.C. Lim, On asymptotic centers and fixed points of nonexpansive mappings, *Can. J. Math.* 32(1980), 421-430.

[26] T.C. Lim, A fixed point theorem for weakly inward multivalued contractions, *J. Math. Anal. Appl.* 247(2000), 323-327.

[27] J.T. Martin, A fixed point theorem for set valued mappings, *Bull. Amer. Math. Soc.* 74(1968), 639-940.

[28] N. Mizoguchi and W. Takahashi, Fixed point theorems for multivalued mappings on complete metric spaces, *J. Math. Anal. Appl.* 141(1989), 177–188.

[29] S.B. Nadler, Jr., Multi-valued contraction mappings, *Pacific J. Math.* 30(1969), 475–488.

[30] Z. Opial, Weak convergence of the sequence of successive approximations for nonexpansive mappings, *Bull. Amer. Math. Soc.* 73(1967), 595–597.

[31] S. Reich, Fixed points of contractive functions, *Boll. Un. Mat. Ital.* 5(1972), 26–42.

[32] S. Reich, Some fixed point theorems, *Atti Accad. Naz. Lincei* 57(1974), 194–198.

[33] S. Reich, Some problems and results in fixed point theory, *Contemporary Math.* 21(1983), 178–187.

[34] F. Sullivan, A generalization of uniformly rotund Banach spaces, *Can. J. Math.* 31(1979), 628–636.

[35] H.K. Xu, Fixed point theorems for single-valued and set-valued mappings. Thesis, Zhejiang University, 1985.

[36] H.K. Xu, A random fixed point theorem for multivalued nonexpansive operators in a uniformly convex Banach space, *Proc. Amer. Math. Soc.* 117(1993), 1089–1092.

[37] H.K. Xu, Inequalities in Banach spaces with applications, *Nonlinear Analysis* 16(1991), 1127–1138.

[38] H.K. Xu, Metric fixed point theory for multivalued mappings, *Dissert. Math.* 389(2000).

[39] H.K. Xu, Multivalued nonexpansive mappings in Banach spaces, *Nonlinear Analysis* 43(2001), 693-706.

A Coupled Legendre Petrov–Galerkin and Collocation Method for the Generalized Korteweg–de Vries Equations

Heping Ma* Weiwei Sun†

Abstract

A coupled Legendre Petrov–Galerkin and collocation method for the generalized
Korteweg–de Vries equation with non-periodic boundary conditions is developed. The method is basically formulated in the Legendre Petrov–Galerkin form to keep good stability. But the nonlinear term is treated by collocation methods for ease of implementation. By choosing appropriate base functions, the scheme can be solved efficiently. It is shown that this non-symmetric approach is more suitable to the underlying $(2r+1)th$-order differential equations and enables us to derive an optimal rate of convergence in L^2-norm. A classical leapfrog–Crank–Nicolson scheme is adopted for the time discretization.

Keywords. generalized Korteweg–de Vries equation, coupled Legendre Petrov–Galerkin and collocation

1 Introduction

The generalized Korteweg–de Vries (KdV) equation covers many important cases such as the KdV equation, the modified KdV equation, and the nonlinear lattice wave equation arising in the study of a number of different physical problems of water wave, plasma physics and anharmonic lattices. The physical phenomena to be modeled by the fifth-order dispersive KdV equation are described briefly in [10]. The spectral/pseudospectral method provides a powerful technique for the numerical

*Department of Mathematics, Shanghai University, Shanghai 200436, People's Republic of China (`hpma@mail.shu.edu.cn`).

†Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong (`maweiw@math.cityu.edu.hk`).

solutions of such problems due to its high-order accuracy and has been studied by many authors in both theoretical and computational aspects [1, 12, 8, 15, 18, 21].

The one of the most important aspect in numerical analysis is to describe the error of the approximation method by the convergence order related to the smoothness of the exact solution. It is well know that under the assumption that $U(x) \in H^\sigma := W^{\sigma,2}$, which is the Sobolev space, the error in L^2-norm of the best approximation to $U(x)$ by algebraic polynomials (or trigonometrical polynomials in the case of the Fourier method) of the degree at most N is of the order $O(N^{-\sigma})$.

For the KdV equation with periodic boundary conditions, Maday and Quarteroni [21] presented a class of Fourier spectral and Fourier pseudospectral methods. They show that the error of the Fourier spectral method is of the order $O(N^{1-\sigma})$ in L^2-norm when the analytic solution is in H^σ. The error of the Fourier pseudospectral method in H^1-norm is of the order $O(N^{2-\sigma})$ in this case. No corresponding L^2-estimate for the Fourier pseudospectral method is known except some modification done by [18].

The KdV equation with non-periodic boundary conditions is also studied by many authors [24, 5, 7, 10]. In a recent work, Huang and Sloan [14] presented a new pseudospectral method for solving the linear third-order differential equation $(\partial_t U + \partial_x^3 U = f)$, which is based on the use of the zeros of $P_{N-2}^{(2,1)}(x)$ as the collocation points. Here we denote by $P_n^{(\alpha,\beta)}(x)$ $(\alpha, \beta > -1)$ the Jacobi polynomials orthogonal on $(-1, 1)$ with the weight functions $\omega_{\alpha,\beta}(x) = (1-x)^\alpha(1+x)^\beta$. Huang and Sloan [14] proved that the method for the linear problem is of "infinite" order accuracy under the assumption that the analytic solution $U(x,t) \in C^\infty$. Li et al [17] extended this method to the KdV equation with the nonlinear term being interpolated at the above collocation points and presented a more precise estimate $O(N^{2-\sigma})$ in L^2-norm. It is obvious that the estimate in [17] is not optimal and it seems unlikely that the optimal error estimate $O(N^{-\sigma})$ can be obtained for the method in [17]. Pavoni [24] first proposed single and multidomain Chebyshev collocation methods for the KdV equation with non-periodic boundary conditions but no error estimate has been provided. A Legendre Petrov–Galerkin (LPG) method for the KdV equation with collocation treatment for the nonlinear term was presented in our recent work [19, 20].

Spectral approximation to the generalized KdV equation with nonperiodic boundary conditions may be written in the following operator form: Find $u_N \in \mathbb{P}_N(I)$ (the space of all polynomials of degree $\leq N$ restricted to $I = (-1, 1)$), satisfying the

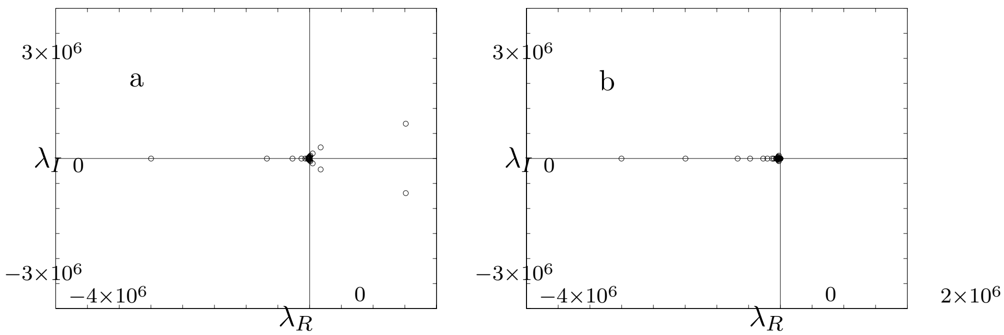

Figure 1: The eigenvalues, except some smaller real ones, of the spectral approximation to third derivative operator with $N = 42$ for the Legendre collocation method (a) and the LPG method (b). The largest real parts of the eigenvalues are 1.52×10^6 and -9.48, respectively

boundary conditions, such that

$$\partial_t u_N(t) + \partial_x \mathcal{P}_1 F(u_N(t)) + (-1)^{r+1} \mathcal{P}_2 \partial_x^{2r+1} u_N(t) = 0,$$

where $\mathcal{P}_1$ and $\mathcal{P}_2$ are some spectral Galerkin/collocation projection operators. The operators, $\mathcal{P}_1$ and $\mathcal{P}_2$, should be designed suitably based on the properties of the problem since classical Legendre/Chebyshev methods do not work well. For example, consider the spectral approximation to the linear third-order equation

$$\partial_t U(x,t) + \partial_x^3 U(x,t) = f(x,t)$$

with the boundary conditions $U(\pm 1, t) = \partial_x U(1,t) = 0$. When a classical Legendre/Chebyshev collocation method is applied, the scheme is unstable as shown in [22]. They find that the eigenvalue of spectral approximation to the third-order derivative operator having the largest positive real part for given N grows in proportion to N^6. Noting the nonsymmetry of the problem, a LPG method was proposed in [19] so that the test functions differ from the trial functions. The scheme reads: Find $u_N \in \mathbb{P}_N(I)$, satisfying the boundary conditions, such that

$$(\partial_t u_N(t), v) - (\partial_x^2 u_N(t), \partial_x v) = (f(t), v), \quad \forall v \in \mathbb{P}_{N-1}(I), v(\pm 1) = 0.$$

Eigenvalue analysis shows that unstable modes disappear in the LPG method. In Figure 1, the eigenvalues of the spectral approximation to third derivative operator with $N = 42$ for the Legendre collocation method and the LPG method are plotted in (a) (given in [22]) and (b), respectively.

For problems with high-order derivatives, the stability limit for an explicit temporal discretization is usually much severe. So the Crank–Nicolson scheme was adopted in [19]. The straightforward use of the method leads to solving a system of a full structure [24]. Appropriate base functions can be constructed to make the corresponding matrix sparse as in the case of the spectral method for the second-order problem [25]. For the third-order problem ($r = 1$), set $\phi_n := L_n - L_{n+2}$, where $\{L_n\}$ are the Legendre polynomials. Thus $\{\phi_n(x)\}_{n=0}^{n=N-3}$ are base functions of the space of test functions. Expanding u_N in terms of $(1-x)\phi_m$ and taking $v = \phi_n$ lead to solving the system:

$$\mathbf{A}\mathbf{a} = \mathbf{g},$$

where $\mathbf{A}$ is an octa-diagonal matrix. A numerical experiment shows that the condition number $Cond\,(\mathbf{A})$ grows like N^2.

As can be seen from above, the LPG method is an efficient approximation to the linear third-order problem. However, when it is applied to the nonlinear problem in the usually way the method becomes undesirable since integrals are involved. The key to this problem is to treat the linear part and nonlinear part in different ways, namely $\mathcal{P}_1 \neq \mathcal{P}_2$. A LPG and collocation method was presented in [20]. Basically, the scheme was formulated in the LPG way. But several collocation methods were considered to compute the nonlinear term as well as the Galerkin spectral method. Also, the nonlinear term was treated explicitly so that the resulting system to be solved was the same as for the linear problem while still keeping the good stability.

In this paper, we apply the LPG and collocation method to the following generalized KdV equation with non-periodic boundary conditions

$$(1.1)\qquad \begin{cases} \partial_t U + \partial_x F(U) + (-1)^{r+1}\partial_x^{2r+1} U = 0, & x \in I, & t \in (0,T], \\ \partial_x^l U(-1,t) = 0, & 0 \le l \le r-1, & t \in [0,T], \\ \partial_x^l U(1,t) = 0, & 0 \le l \le r, & t \in [0,T], \\ U(x,0) = U_0(x), & x \in I, & \end{cases}$$

where $r \geq 1$, $F(z)$ is a smooth function of z, and $I = (-1,1)$. Attentions will be paid on the efficient implementation of the scheme, good stability of time advancing, better condition number of algebraic system, and the theoretical analysis of optimal rate of convergence as discussed above.

2 The Legendre Petrov–Galerkin and collocation methods

For any non-negative integer σ, $H^\sigma(I) := W^{\sigma,2}(I)$ and $H_0^\sigma(I) := W_0^{\sigma,2}(I)$ are the Sobolev spaces with the norm $\|\cdot\|_\sigma$ and semi-norm $|\cdot|_\sigma$, respectively. For a positive weight $\omega(x)$ on I, the inner product and norm of $L^2_\omega(I)$ are denoted by $(\cdot,\cdot)_\omega$ and $\|\cdot\|_\omega$, respectively. We will drop the subscript ω whenever $\omega(x) \equiv 1$. Let

$$H_0^{r_1,r_2}(I) = \{v \in H^{\max\{r_1,r_2\}}(I) \mid \partial_x^l v(-1) = 0,\}$$

$0 \leq l \leq r_1 - 1;\ \partial_x^l v(1) = 0,\ 0 \leq l \leq r_2 - 1.$

A weak form of the problem (1.1) is to find $U(t) \in H_0^{r,r+1}(I)$ such that for any $v \in H_0^r(I)$,

$$\text{(2.1)} \qquad \begin{cases} (\partial_t U, v) + (\partial_x F(U), v) - (\partial_x^{r+1} U, \partial_x^r v) = 0, & t \in (0, T], \\ (U, v) = (U_0, v), & t = 0. \end{cases}$$

Let $\mathbb{P}_N(I)$ be the space of polynomials of degree at most N on the interval I,

$$\mathcal{V}_N = \mathbb{P}_N(I) \cap H_0^{r,r+1}(I), \qquad \mathcal{W}_{N-1} = \mathbb{P}_{N-1}(I) \cap H_0^r(I),$$

which are used as the trial and test function spaces, respectively. To make the method more efficient, we also use some collocation methods to compute the nonlinear term. The semi-discrete LPG and collocation method for (1.1) is to find $u_N(t) \in \mathcal{V}_N$ $(t \geq 0)$ such that for any $v \in \mathcal{W}_{N-1}$,

$$\text{(2.2)} \qquad \begin{cases} (\partial_t u_N(t), v) + (\partial_x P_N F(u_N(t)), v) - (\partial_x^{r+1} u_N(t), \partial_x^r v) = 0, & 0 < t \leq T, \\ (u_N(0), v) = (P_N^{(0)} U_0, v), \end{cases}$$

where the spectral approximation operator P_N can be one of the following:

1. $\mathcal{P}_N : L^2(I) \to \mathbb{P}_N(I)$, the Legendre Galerkin projection operator such that

$$(\mathcal{P}_N u, v) = (u, v), \qquad \forall\, v \in \mathbb{P}_N(I);$$

2. $\mathcal{I}_N : C(\bar{I}) \to \mathbb{P}_N(I)$, the polynomial interpolation operator at Legendre–Gauss–Lobatto points $\{x_j\}_{j=0}^N$ ($x_0 = -1$, $x_N = 1$, and $\{x_j\}_{j=1}^{N-1}$ are the zeros of $P_{N-1}^{(1,1)}(x)$) such that

$$\mathcal{I}_N u(x_j) = u(x_j), \quad j = 0, \cdots, N;$$

3. $\mathcal{I}'_N : C^1(\bar{I}) \to \mathbb{P}_{N-2r+2}(I)$, the polynomial interpolation operator at

$\{x'_j\}_{j=0}^{N-2r+1}$ ($x'_0 = -1$, $x'_{N-2r+1} = 1$, and $\{x'_j\}_{j=1}^{N-2r}$ are the zeros of $P_{N-2r}^{(r+1,r)}(x)$) such that

$$\mathcal{I}'_N u(x'_j) = u(x'_j), \quad j = 0, \cdots, N-2r+1; \quad \partial_x \mathcal{I}'_N u(1) = \partial_x u(1).$$

The third method takes more care of the nonsymmetry of the problem, which is based on the generalized Legendre–Guass quadrature rule introduced in [14] and similar to that adopted in [17]. We let $P_N^{(0)} = P_N$ in the first two cases and, in Case 3, let $P_N^{(0)} U_0 \in \mathcal{V}_N$ such that

$$P_N^{(0)} U_0(x'_j) = U_0(x'_j), \quad j = 1, \cdots, N-2r.$$

Remark 2.1 *We can also choose $\mathcal{I}_N^C : C(\bar{I}) \to \mathbb{P}_N(I)$, the polynomial interpolation operator at Chebyshev-Gauss points, as P_N and $P_N^{(0)}$ in* (2.2)*. This combination of the LPG and Chebyshev collocation method allows for the use of the fast Legendre transformation* [2] *and turns out being very efficient, which will be analyzed in another paper.*

Let τ be the step size in time space and $t_k = k\tau$ ($k = 0, 1, 2, \cdots, n_T$; $T = n_T \tau$). For simplicity, we denote $u^k(x) := u(x, t_k)$ by u^k usually and

$$u_{\hat{t}}^k = \frac{1}{2\tau}(u^{k+1} - u^{k-1}), \qquad \hat{u}^k = \frac{1}{2}(u^{k+1} + u^{k-1}).$$

The fully discrete LPG and collocation method for (1.1) is to find $u_N^k \in \mathcal{V}_N$ such that for any $v \in \mathcal{W}_{N-1}$,

$$\text{(2.3)} \qquad \begin{cases} (u_{N\hat{t}}^k, v) + (\partial_x P_N F(u_N^k), v) - (\partial_x^{r+1} \hat{u}_N^k, \partial_x^r v) = 0, \quad 1 \le k \le n_T - 1, \\ (u_N^1, v) = (P_N^{(0)}[U_0 + \tau \partial_t U(0)], v), \\ (u_N^0, v) = (P_N^{(0)} U_0, v). \end{cases}$$

Here the time-discretization is a leapfrog–Crank–Nicolson scheme, which is of the second-order accuracy in time space. The method is more efficient and of better stability since it is a semi-implicit time advancing scheme, i.e., implicit for the linear term of the higher order derivative but explicit for the nonlinear term. The solution of above scheme can be obtained by solving a system with a diagonal strip matrix at each time level t_k. We refer to [20] for the details in the case of $r = 1$.

3 Preliminaries

Throughout this paper C will denote a positive generic constant. We assume that N is sufficiently large ($N > 4r + 1$). In this section, we analyze approximation properties of some projection operators and give some basic lemmas, which will be needed in the error estimates.

We first recall a basic result of Jacobi approximation [17]. Let $\mathcal{P}_N^{\alpha,\beta} : L^2_{\omega_{\alpha,\beta}}(I) \to \mathbb{P}_N(I)$ be the orthogonal projection operator with the weight $\omega_{\alpha,\beta}(x)$ and $\mathcal{P}_N := \mathcal{P}_N^{00}$ for simplicity.

Lemma 3.1 ([17]) *If $\alpha, \beta > -1$ and $v \in H^\sigma(I)$,*

$$\begin{aligned} \|\partial_x^s(v - \mathcal{P}_N^{\alpha,\beta} v)\|_{\omega_{\alpha+s,\beta+s}} &\le CN^{s-\sigma}\|\partial_x^\sigma(v - \mathcal{P}_N^{\alpha,\beta} v)\|_{\omega_{\alpha+\sigma,\beta+\sigma}} \\ &\le CN^{s-\sigma}\|\partial_x^\sigma v\|_{\omega_{\alpha+\sigma,\beta+\sigma}}, \qquad 0 \le s \le \sigma. \end{aligned} \tag{3.1}$$

We introduce a projection operator (cf. Section 6 of [4]) concerning the higher order derivative term in (1.1). We define

$$P_N^{r+1} u = \bar{\partial}_x^{-r-1} \mathcal{P}_{N-r-1} \partial_x^{r+1} u, \qquad \forall\, u \in H^{r+1}(I),$$

where

$$\bar{\partial}_x^{-1} f(x) = -\int_x^1 f(y)\, dy, \qquad \bar{\partial}_x^{-m} f(x) = (\bar{\partial}_x^{-1})^m f(x),$$

Then, for any $v \in \mathbb{P}_{N-1}(I)$,

$$(\partial_x^{r+1} P_N^{r+1} u, \partial_x^r v) = (\mathcal{P}_{N-r-1} \partial_x^{r+1} u, \partial_x^r v) = (\partial_x^{r+1} u, \partial_x^r v). \tag{3.2}$$

We show that $P_N^{r+1} u \in \mathcal{V}_N$ if $u \in H_0^{r,r+1}(I)$. It is easy to see that

$$(\partial_x^l P_N^{r+1} u)(1) = (\bar{\partial}_x^{-r-1+l} \mathcal{P}_{N-r-1} \partial_x^{r+1} u)(1) = 0, \qquad 0 \le l \le r. \tag{3.3}$$

Integrating by parts, we have for $0 \le l \le r-1$ and $N \ge 2r+1$ that

$$\begin{aligned} (\partial_x^l P_N^{r+1} u)(-1) &= -\int_{-1}^1 \bar{\partial}_x^{-r+l} \mathcal{P}_{N-r-1} \partial_x^{r+1} u(x)\, dx \\ &= \frac{-1}{(r-l)!} \int_{-1}^1 [\partial_x^{r-l} (x+1)^{r-l}][\bar{\partial}_x^{-r+l} \mathcal{P}_{N-r-1} \partial_x^{r+1} u](x)\, dx \\ &= \frac{(-1)^{r-l+1}}{(r-l)!} \int_{-1}^1 (x+1)^{r-l} \mathcal{P}_{N-r-1} \partial_x^{r+1} u(x)\, dx \\ &= \frac{(-1)^{r-l+1}}{(r-l)!} \int_{-1}^1 (x+1)^{r-l} \partial_x^{r+1} u(x)\, dx \\ &= -\int_{-1}^1 \partial_x^{l+1} u(x)\, dx = -(\partial_x^l u)(x)|_{-1}^1 = 0, \qquad 0 \le l \le r-1. \end{aligned} \tag{3.4}$$

Also, for $u \in H_0^{r,r+1}(I)$, the operator P_N^{r+1} keeps the value $\partial_x^r u(-1)$:

$$(\partial_x^r P_N^{r+1} u)(-1) = -\int_{-1}^{1} \mathcal{P}_{N-r-1}\partial_x^{r+1} u(x)\,dx = -(\partial_x^r u)(x)|_{-1}^{1} = \partial_x^r u(-1). \tag{3.5}$$

Furthermore, if $u \in H_0^{r,r+1}(I)$, for any $v \in \mathbb{P}_{N+l-2r-2}(I)$ and $0 \le l \le r$,

$$\begin{aligned}(\partial_x^l (u - P_N^{r+1}u), v) &= (\partial_x^l (u - P_N^{r+1}u), \partial_x^{r+1-l}\bar{\partial}_x^{-r-1+l} v)\\ &= (-1)^{r+1-l}(\partial_x^{r+1}(u - P_N^{r+1}u), \bar{\partial}_x^{-r-1+l} v)\\ &= (-1)^{r+1-l}((I - \mathcal{P}_{N-r-1})\partial_x^{r+1}u, \bar{\partial}_x^{-r-1+l} v) = 0.\end{aligned} \tag{3.6}$$

We have the following approximation result for $P_N^{r+1}u$.

Lemma 3.2 *If $u \in H_0^{r,r+1}(I) \cap H^\sigma(I)$ and $\sigma \ge r+1$,*

$$\begin{aligned}\|\partial_x^l (u - P_N^{r+1}u)\|_{\omega_{l-r-1,l-r-1}} &\le CN^{l-\sigma}\|\partial_x^\sigma (u - P_N^{r+1}u)\|_{\omega_{\sigma-r-1,\sigma-r-1}},\\ &\le CN^{l-\sigma}\|\partial_x^\sigma u\|_{\omega_{\sigma-r-1,\sigma-r-1}}, \qquad 0 \le l \le r+1.\end{aligned} \tag{3.7}$$

Proof. Let $g = \partial_x^l (u - P_N^{r+1}u)$ $(0 \le l \le r+1)$. From (3.3), (3.4), and (3.5), $g \in H_0^{r+1-l}(I)$. Therefore, $g\omega_{l-r-1,l-r-1} \in L^2(I)$ by the Hardy's inequality (see [23], p. 145). We have from (3.1) that

$$\begin{aligned}\|\partial_x^l (u - P_N^{r+1}u)\|^2_{\omega_{l-r-1,l-r-1}} &= (\partial_x^l (u - P_N^{r+1}u), \partial_x^{r+1-l}\bar{\partial}_x^{-r-1+l}[g\omega_{l-r-1,l-r-1}])\\ &= |((I - \mathcal{P}_{N-r-1})\partial_x^{r+1}u, (I - \mathcal{P}_{N-r-1})\bar{\partial}_x^{-r-1+l}[g\omega_{l-r-1,l-r-1}])|\\ &\le \|(I - \mathcal{P}_{N-r-1})\partial_x^{r+1}u\|\ \|(I - \mathcal{P}_{N-r-1})\bar{\partial}_x^{-r-1+l}[g\omega_{l-r-1,l-r-1}]\|\\ &\le CN^{r+1-\sigma}\|\partial_x^{\sigma-r-1}(I - \mathcal{P}_{N-r-1})\partial_x^{r+1}u\|_{\omega_{\sigma-r-1,\sigma-r-1}}N^{l-r-1}\|g\omega_{l-r-1,l-r-1}\|_{\omega_{r+1-l,r+1-l}}\\ &\le CN^{l-\sigma}\|\partial_x^\sigma (I - P_N^{r+1})u\|_{\omega_{\sigma-r-1,\sigma-r-1}}\|g\|_{\omega_{l-r-1,l-r-1}},\end{aligned}$$

which gives (3.7). □

The following lemma gives some inverse properties related to the weight.

Lemma 3.3 *If $\alpha > -1$ and $\beta > -1$,*

$$\|v\|_{\omega_{\alpha,\beta}} \le CM\|v\|_{\omega_{\alpha+1,\beta+1}}, \qquad \forall\, v \in \mathbb{P}_M(I). \tag{3.8}$$

If $\alpha > \min\{-1, -2r_2\}$ and $\beta > \min\{-1, -2r_1\}$ $(r_1, r_2 \ge 0)$,

$$\|\partial_x v\|_{\omega_{\alpha+1,\beta+1}} \le CM\|v\|_{\omega_{\alpha,\beta}}, \qquad \forall\, v \in \mathbb{P}_M(I) \cap H_0^{r_1,r_2}(I). \tag{3.9}$$

Proof. The first result (3.8) can be shown by the Gauss-Jacobi quadrature formula of $M+2$ points and the asymptotic properties of the zeros of Jacobi polynomials (see (8.9.1) of [26]).

For (3.9), we write $v=\tilde{v}\omega_{r_2,r_1}$. By (3.8) and

$$\begin{aligned}\|\partial_x P_n^{(\alpha+2r_2,\beta+2r_1)}\|^2_{\omega_{\alpha+1+2r_2,\beta+1+2r_1}}\\ = n(n+1+\alpha+2r_2+\beta+2r_1)\|P_n^{(\alpha+2r_2,\beta+2r_1)}\|^2_{\omega_{\alpha+2r_2,\beta+2r_1}},\end{aligned} \tag{3.10}$$

we get

$$\begin{aligned}\|\partial_x v\|^2_{\omega_{\alpha+1,\beta+1}} \le C({r_1}^2\|\tilde{v}\|^2_{\omega_{\alpha+2r_2,\beta+2r_1-1}} + {r_2}^2\|\tilde{v}\|^2_{\omega_{\alpha+2r_2-1,\beta+2r_1}})\\ +\|\partial_x\tilde{v}\|^2_{\omega_{\alpha+1+2r_2,\beta+1+2r_1}} \le CM^2\|\tilde{v}\|^2_{\omega_{\alpha+2r_2,\beta+2r_1}},\end{aligned}$$

where we have used the orthogonality properties of $\{\partial_x P_n^{(\alpha+2r_2,\beta+2r_1)}\}$ and $\{P_n^{(\alpha+2r_2,\beta+2r_1)}\}$. $\square$

We define an interpolation operator $\mathcal{I}_N^r : C^r(\bar{I}) \to \mathbb{P}_N(I)$ based on the generalized Legendre–Guass quadrature rule using the values of $\partial_x^l u(\pm 1)$ $(0\le l\le r-1)$ and $\partial_x^r u(1)$ such that

$$\begin{cases}\mathcal{I}_N^r u(x_j')=u(x_j'), \quad j=1,\cdots,N-2r,\\ \partial_x^l\mathcal{I}_N^r u(\pm1)=\partial_x^l u(\pm1), \quad 0\le l\le r-1, \quad \partial_x^r\mathcal{I}_N^r u(1)=\partial_x^r u(1).\end{cases}$$

We need approximation results of the two interpolation operators $\mathcal{I}_N$ and $\mathcal{I}_N^r$. Let $\{x_j\}_{j=0}^N$ $(\{x_j'\}_{j=0}^{N-2r+1})$ be defined in Section 2 and $\{\omega_j\}_{j=0}^N$ $(\{\omega_j'\}_{j=1}^{N-2r}$, $\{{\omega'}_j^-\}_{j=0}^{r-1}$, $\{{\omega'}_j^+\}_{j=0}^{r})$ be the corresponding weights of the quadrature formula. We define

$$(u,v)_N=\sum_{j=0}^N \omega_j u(x_j)v(x_j), \qquad \|v\|_N=(v,v)_N^{1/2},$$

$$(u,v)_{N,',\omega}=\sum_{j=1}^{N-2}\omega_j' u(x_j')v(x_j')\omega(x_j'), \qquad \|v\|_{N,',\omega}=(v,v)_{N,',\omega}^{1/2},$$

where the subscript ω will be dropped whenever $\omega(x)\equiv 1$.

Lemma 3.4 *If* $u\in H^\sigma(I)$, *then*

$$\|u-\mathcal{I}_N u\|_{\omega_{-1,-1}} + N^{-1}\|\partial_x(u-\mathcal{I}_N u)\| \le CN^{-\sigma}\|\partial_x^\sigma u\|_{\omega_{\sigma-1,\sigma-1}}, \qquad \sigma\ge 1, \tag{3.11}$$

$$\|u-\mathcal{I}_N^r u\|_{\omega_{-r-1,-r}} + N^{-1}\|\partial_x(u-\mathcal{I}_N^r u)\|_{\omega_{-r,-r+1}} \le CN^{-\sigma}\|\partial_x^\sigma u\|_{\omega_{\sigma-r-1,\sigma-r-1}}, \quad \sigma\ge r+1. \tag{3.12}$$

Proof. The first result is similar to ones given in (13.25) and (13.26) of [4]. To show the second result, we follow the line in [3] and first prove that

$$\|\mathcal{I}_N^r v\|_{\omega_{-r-1,-r}} \le C(\|v\|_{\omega_{-r-1,-r}} + N^{-1}\|\partial_x v\|_{\omega_{-r,-r+1}}), \qquad \forall\, v \in H_0^{r,r+1}(I). \tag{3.13}$$

We notice that

$$\omega'_j = \frac{\omega^G_{N-2r,j}(r+1,r)}{(1-x'_j)^{r+1}(1+x'_j)^r} \tag{3.14}$$

and the Cotes-Christoffel number (see (15.3.10) of [26])

$$\omega^G_{N-2r,j}(r+1,r) = \left(\sin\frac{\theta_j}{2}\right)^{2r+3}\left(\cos\frac{\theta_j}{2}\right)^{2r+1} O(N^{-1}), \qquad x'_j = \cos\theta_j.$$

We see from (3.14) that

$$0 < \omega'_j \le CN^{-1}\sin\theta_j, \qquad j = 1,\cdots,N-2r. \tag{3.15}$$

With the notation

$$[f]_d := \sum_{j=1}^{N-2r} \omega'_j f(x'_j) + \sum_{j=0}^{r-1} {\omega'}_j^- \partial_x^j f(-1) + \sum_{j=0}^{r} {\omega'}_j^+ \partial_x^j f(1),$$

we have that [14]

$$[f]_d = \int_{-1}^1 f(x)dx, \qquad \forall\, f \in \mathbb{P}_{2N-2r}(I). \tag{3.16}$$

Let $\hat{v}(\theta) = v(\cos\theta)$ and $\hat{\omega}_{\alpha,\beta}(\theta) = \omega_{\alpha,\beta}(\cos\theta)$. We get from (3.16) that

$$\begin{aligned} \|\mathcal{I}_N^r v\|^2_{\omega_{-r-1,-r}} &= \|\mathcal{I}_N^r v\|^2_{N,',\omega_{-r-1,-r}} \\ &\le CN^{-1}\sum_{j=1}^{N-2r} \hat{v}^2(\theta_j)\hat{\omega}_{-r-1,-r}(\theta_j)\sin\theta_j, \qquad \forall\, v \in H_0^{r,r+1}(I). \end{aligned} \tag{3.17}$$

We have $\theta_j \in I_j \subset (0,\pi)$ with I_j of size CN^{-1}. Thus,

$$\begin{aligned} \|\mathcal{I}_N^r v\|^2_{\omega_{-r-1,-r}} &\le CN^{-1}\sum_{j=1}^{N-2r} \sup_{\theta\in I_j} |\hat{v}^2(\theta)\hat{\omega}_{-r-1,-r}(\theta)\sin\theta| \\ &\le C\sum_{j=1}^{N-2r}\left(\|\hat{v}(\hat{\omega}_{-r-1,-r}\sin\theta)^{1/2}\|^2_{L^2(I_j)} + N^{-2}\|[\hat{v}(\hat{\omega}_{-r-1,-r}\sin\theta)^{1/2}]_\theta\|^2_{L^2(I_j)}\right) \\ &\le C\int_0^\pi \big(\hat{v}^2(\theta)\hat{\omega}_{-r-1,-r}(\theta) + N^{-2}\hat{v}^2(\theta)\hat{\omega}_{-r-2,-r-1}(\theta) \\ &\quad + N^{-2}(\hat{v}_\theta)^2(\theta)\hat{\omega}_{-r-1,-r}(\theta)\big)\sin\theta\, d\theta \\ &= C\left(\|v\|^2_{\omega_{-r-1,-r}} + N^{-2}\|v\|^2_{\omega_{-r-2,-r-1}} + N^{-2}\|\partial_x v\|^2_{\omega_{-r,-r+1}}\right), \end{aligned} \tag{3.18}$$

which, by the Hardy's inequality, gives (3.13).

Next, for any $u \in H^\sigma(I)$ $(\sigma \ge r+1)$, there exists $q(x) \in \mathbb{P}_{2r}(I)$ such that $\tilde{u} := u - q \in H_0^{r,r+1}(I)$. Let $u^* = P_N^{r+1}\tilde{u} + q$. Replacing v by $u^* - u = (P_N^{r+1} - I)\tilde{u} \in H_0^{r+1}(I)$ in (3.13) and noting $\mathcal{I}_N^r u^* = u^*$, we get from (3.7)

$$\begin{aligned}
\|u^* - \mathcal{I}_N^r u\|_{\omega_{-r-1,-r}} &\le C(\|u^* - u\|_{\omega_{-r-1,-r}} + N^{-1}\|\partial_x(u^* - u)\|_{\omega_{-r,-r+1}}). \\
&= C(\|P_N^{r+1}\tilde{u} - \tilde{u}\|_{\omega_{-r-1,-r}} + N^{-1}\|\partial_x(P_N^{r+1}\tilde{u} - \tilde{u})\|_{\omega_{-r,-r+1}}) \\
&\le CN^{-\sigma}\|\partial_x^\sigma(I - P_N^{r+1})\tilde{u}\|_{\omega_{\sigma-r-1,\sigma-r-1}} \\
&\le CN^{-\sigma}\|\partial_x^\sigma(I - P_N^{r+1})u\|_{\omega_{\sigma-r-1,\sigma-r-1}},
\end{aligned} \tag{3.19}$$

which gives the first part of (3.12).

The second part of the result (3.12) is obtained by applying (3.9) to $u^* - \mathcal{I}_N^r u$ with $(\alpha, \beta) = (-r-1, -r)$ and using (3.19) . □

The following lemma plays an important role in the convergence analysis.

Lemma 3.5 *Let $u \in H_0^{r,r+1}(I)$ and $v(x) := u(x)/(1-x)$. We have*

$$-(\partial_x^{r+1}u, \partial_x^r v) = \left(r + \frac{1}{2}\right)|v|_r^2 + |\partial_x^r v(-1)|^2, \tag{3.20}$$

$$|v|_l \le |v|_r, \qquad 0 \le l \le r. \tag{3.21}$$

Proof. By the Hardy's inequality, we can check that $v \in H_0^r(I)$ and $(1-x)(\partial_x^r v)^2|_{x=1} = 0$, which is trivial for $u \in \mathcal{V}_N$ as we needed in this paper. We have

$$\begin{aligned}
-(\partial_x^{r+1}u, \partial_x^r v) &= (\partial_x^{r+1}[(x-1)v], \partial_x^r v) = ((x-1)\partial_x^{r+1}v, \partial_x^r v) + (r+1)\|\partial_x^r v\|^2 \\
&= \frac{1}{2}[(x-1)(\partial_x^r v)^2]|_{-1}^1 - \frac{1}{2}(\partial_x^r v, \partial_x^r v) + (r+1)\|\partial_x^r v\|^2,
\end{aligned} \tag{3.22}$$

which gives (3.20). Since $v \in H_0^r(I)$, (3.21) can be get easily by induction. □

Lemma 3.6 *If $v \in \mathcal{V}_N$, then*

$$\|v\|_{\omega_{-r,-r+1}} \le \|v\|_{N,',\omega_{-r,-r+1}} \le C\|v\|_{\omega_{-r,-r+1}}. \tag{3.23}$$

Proof. We follow the line in the proof of (13.18) in [4]. Any $v \in \mathbb{P}_N(I) \cap H_0^{r-1,r}(I)$ can be written as

$$v(x) = \omega_{r,r-1}(x) \sum_{n=0}^{N-2r+1} \hat{v}_n P_n^{(r,r-1)}(x).$$

By the orthogonal property of $\{P_n^{(r,r-1)}(x)\}$

$$\|v\|_{\omega_{-r,-r+1}}^2 = \sum_{n=0}^{N-2r+1} |\hat{v}_n|^2 \|P_n^{(r,r-1)}\|_{\omega_{r,r-1}}^2$$

and by (3.16)

$$[v^2\omega_{-r,-r+1}]_d = \sum_{n=0}^{N-2r} |\hat{v}_n|^2 \|P_n^{(r,r-1)}\|^2_{\omega_{r,r-1}} + |\hat{v}_{N-2r+1}|^2 [(P_{N-2r+1}^{(r,r-1)})^2\omega_{r,r-1}]_d.$$

Since $\partial_x P_{N-2r+1}^{(r,r-1)}(x) = \dfrac{N+1}{2} P_{N-2r}^{(r+1,r)}(x)$ and

$$(P_{N-2r+1}^{(r,r-1)})^2 + \frac{1}{(N-2r+1)^2}(1-x^2)(\partial_x P_{N-2r+1}^{(r,r-1)}) \in \mathbb{P}_{2N-4r+1}(I),$$

we obtain from (3.10)

$$\begin{aligned}
[(P_{N-2r+1}^{(r,r-1)})^2\omega_{r,r-1}]_d &= \|P_{N-2r+1}^{(r,r-1)}\|^2_{\omega_{r,r-1}} + \frac{1}{(N-2r+1)^2}\|\partial_x P_{N-2r+1}^{(r,r-1)}\|^2_{\omega_{r+1,r}} \\
&= \left(1 + \frac{N+1}{N-2r+1}\right) \|P_{N-2r+1}^{(r,r-1)}\|^2_{\omega_{r,r-1}}.
\end{aligned}$$

Therefore, for any $v \in \mathbb{P}_N(I) \cap H_0^{r-1,r}(I)$,

$$\|v\|^2_{\omega_{-r,-r+1}} \le [v^2\omega_{-r,-r+1}]_d \le C\|v\|^2_{\omega_{-r,-r+1}}, \tag{3.24}$$

which yields (3.23) for $v \in \mathcal{V}_N$. ◻

Lemma 3.7 *Let $\alpha > -1,\ \beta > -1$. We have*

$$\|v\|_{N,',\omega_{\alpha,\beta}} \le C\left(\|v\|_{\omega_{\alpha,\beta}} + N^{-1}\|v_x\|_{\omega_{\alpha+1,\beta+1}}\right), \qquad \forall\, v \in H^1_{\omega_{\alpha+1,\beta+1}}(I), \tag{3.25}$$

$$\|v\|_{N,',\omega_{\alpha,\beta}} \le C(1 + MN^{-1})\|v\|_{\omega_{\alpha,\beta}}, \qquad \forall\, v \in \mathbb{P}_M(I), \tag{3.26}$$

$$\|v\|_\infty \le C_1 M\|v\|, \qquad \forall\, v \in \mathbb{P}_M(I). \tag{3.27}$$

Proof. By the same approach as in the proof of (3.17) and (3.18), according to the asymptotic properties of θ_j, we can choose $I_j \subset (\hat{\delta}, \pi - \hat{\delta})$ with $\hat{\delta} = C/N > 0$. Then

$$\begin{aligned}
\|v\|^2_{N,',\omega_{\alpha,\beta}} &\le CN^{-1}\sum_{j=1}^{N-2r} \hat{v}^2(\theta_j)\hat{\omega}_{\alpha,\beta}(\theta_j)\sin\theta_j \\
&\le C\left(\|v\|^2_{\omega_{\alpha,\beta}} + N^{-2}\|v\|^2_{L^2_{\omega_{\alpha-1,\beta-1}}(\tilde{I})} + N^{-2}\|v_x\|^2_{\omega_{\alpha+1,\beta+1}}\right),
\end{aligned} \tag{3.28}$$

where $\tilde{I} = (-1+\delta, 1-\delta)$ with $\delta = C/N^2 > 0$. So (3.25) follows. Then by (3.9) we have (3.26). The result (3.27) is well known [4]. ◻

Lemma 3.8 ([20]) *Assume that*

$E(t), \rho(t)$ are non-negative functions continuous on $[0,T]$, $\rho(t)$ is increasing, and ε, C are positive constants,

for any $t \in (0,T]$, if $\max\limits_{0\le s\le t} E(s) \le \varepsilon$, $E(t) \le \rho(t) + C\displaystyle\int_0^t E(s)\,ds$,

$E(0) \le \rho(0)$ and $\rho(T)\mathrm{e}^{CT} \le \varepsilon$. Then for any $0 \le t \le T$,

$$E(t) \le \rho(t)\mathrm{e}^{Ct}.$$

The following lemma is a discrete form of Lemma 3.8.

Lemma 3.9 (cf. Lemma 4.16 of [13]). *Assume that*

E^k, ρ^k $(k = 0, 1, \cdots, n_T)$ are non-negative set functions, ρ^k is increasing, and ε, C are positive constants,

for any $1 \le n \le n_T$, if $\max\limits_{1\le k\le n} E^k \le \varepsilon$, $E^n \le \rho^n + C\tau\sum\limits_{k=0}^{n-1} E^k$,

for any $1 \le k \le n_T$, $E^k - E^{k-1} \le \dfrac{\varepsilon}{2}$,

$E^0 \le \rho^0$ and $\rho^{n_T}\mathrm{e}^{CT} \le \dfrac{\varepsilon}{2}$. Then for any $0 \le n \le n_T$,

$$E^n \le \rho^{n\tau}\mathrm{e}^{Cn\tau}.$$

Remark 3.1 *The conditions* (ii) *and* (iii) *in Lemma* 3.9 *can be replaced by*

(ii)$'$ *for any $1 \le n \le n_T$, if $\max\limits_{1\le k\le n-1} E^k \le \varepsilon$, $E^n \le \rho^n + C\tau\sum\limits_{k=0}^{n-1} E^k$.*

This form is used for the numerical analysis of explicit (or linear implicit) schemes.

4 The convergence of the semi-discrete scheme

In this section we prove the convergence of optimal rate for the semi-discrete scheme of the LPG and collocation methods. We assume that the solution of (1.1) $U \in C(0,T; H^\sigma(I))$ $(\sigma \ge r+1)$ and the positive generic constant C may be dependent on the norm of U in this space. For simplicity, let $\omega(x) = (1-x)^{-1}$ hereafter. Also, we denote by $C(A, B)$ a positive generic constant dependent on A, B, etc.

Let $u^*(t) = P_N^{r+1}U(t)$, which is used as a comparison function, and $e_N(t) = u_N(t) - u^*(t)$. By (2.1), (2.2), and (3.2), we have for any $v \in \mathcal{W}_{N-1}$ that

$$(4.1)\quad \begin{cases} (\partial_t e_N(t), v) + (\partial_x P_N \tilde{F}_N(t), v) - (\partial_x^{r+1} e_N(t), \partial_x^r v) = (f(t), v), & t \in (0,T], \\ (e_N(0), v) = (P_N^{(0)} U_0 - u^*(0), v), \end{cases}$$

where

$$\tilde{F}_N(t) = F(u_N(t)) - F(u^*(t)),$$
$$f(t) = \partial_x[F(U(t)) - P_N F(u^*(t))] + [\partial_t U(t) - \partial_t u^*(t)] := f_1(t) + f_2(t).$$

Let $\eta_N(t) = e_N(t)\omega \in \mathcal{W}_{N-1}$. Taking $v = 2\eta_N(t)$ in (4.1), we get from (3.20) that

(4.2)
$$\frac{d}{dt}\|e_N(t)\|_\omega^2 + (2r+1)|\eta_N(t)|_r^2 + 2|\partial_x^r \eta_N(-1,t)|^2 \le 2|(f(t) - \partial_x P_N \tilde{F}_N(t), \eta_N(t))|.$$

For given $t \in (0, T]$, we assume that (see the condition (ii) of Lemma 3.8)

(4.3)
$$\max_{0\le s\le t} \|e_N(s)\| \le N^{-1}, \qquad C_F := \max_{|z|\le 2\|U\|_{C(0,T;H^{r+1}(I))}+C_1} |\partial_z F(z)|,$$

where C_1 is a constant appearing in (3.27) so that

$$\|e_N(s)\|_{L^\infty(I)} \le C_1 N \|e_N(s)\| \le C_1, \quad 0 \le s \le t.$$

Consider $0 \le s \le t$. We deal with the nonlinear term and the initial error in three different cases as follows:

Case 1. $P_N = \mathcal{P}_N$. Noting that $\|u^*(s)\|_{L^\infty(I)} \le 2|P_N^{r+1}U|_{r+1} \le 2|U(s)|_{r+1}$, we have from (4.3) that,

$$|(\partial_x \mathcal{P}_N \tilde{F}_N(s), \eta_N(s))| = |(F(u_N(s)) - F(u^*(s)), \partial_x \eta_N(s))|$$
$$\le C_F \|e_N(s)\| \|\partial_x \eta_N(s)\| \le C\|e_N(s)\|^2 + \frac{1}{6}|\eta_N(s)|_r^2.$$

By (3.7) and (3.21),

$$|(f_1(s), \eta_N(s))| = |(F(U(s)) - F(u^*(s)), \partial_x \eta_N(s))|$$
$$\le C_F \|(I - P_N^{r+1})U(s)\| \; |\eta_N(s)|_1 \le CN^{-2\sigma}\|U(s)\|_\sigma^2 + \frac{1}{6}|\eta_N(s)|_r^2,$$

and

$$\|e_N(0)\|_\omega \le \|(I - P_N^{r+1})U_0\|_\omega \le CN^{-\sigma}\|U_0\|_\sigma.$$

Case 2. $P_N = \mathcal{I}_N$. We use the exactness of the quadrature rule, (4.3), and (9.3.2) of [6] to get

$$|(\partial_x \mathcal{I}_N \tilde{F}_N(s), \eta_N(s))| = |(F(u_N(s)) - F(u^*(s)), \partial_x \eta_N(s))_N|$$
$$\le C_F \|e_N(s)\|_N \|\partial_x \eta_N(s)\|_N \le C\|e_N(s)\|^2 + \frac{1}{6}|\eta_N(s)|_r^2.$$

Similarly, it follows from (3.11), (4.3), and (3.7) that

$$
\begin{aligned}
|(f_1(s),\eta_N(s))| &\le |((I-\mathcal{I}_N)F(U(s)),\partial_x\eta_N(s))| + |(F(U(s))-F(u^*(s)),\partial_x\eta_N(s))_N| \\
&\le C(\|(I-\mathcal{I}_N)F(U(s))\| + C_F\|(\mathcal{I}_N-P_N^{r+1})U(s)\|)|\eta_N(s)|_1 \\
&\le CN^{-2\sigma} + \frac{1}{6}|\eta_N(s)|_r^2,
\end{aligned}
$$

provided that $F(z)\in C^\sigma(\mathbb{R})$. For the initial error, we have from (3.11) and (3.7) that

$$
\|e_N(0)\|_\omega \le \|(\mathcal{I}_N-P_N^{r+1})U_0\|_\omega \le CN^{-\sigma}\|U_0\|_\sigma.
$$

Case 3. $P_N=\mathcal{I}_N'$. Since $[\mathcal{I}_N'\tilde F_N(s)]\partial_x\eta_N(s)\in\mathbb{P}_{2N-2r}(I)\cap H_0^{r,r+1}(I)$, we can use (3.16) and (3.26) to get

$$
\begin{aligned}
|(\partial_x\mathcal{I}_N\tilde F_N(s),\eta_N(s))| &= |(F(u_N(s))-F(u^*(s)),\partial_x\eta_N(s))_{N,'}| \\
&\le C_F\|e_N(s)\|_{N,'}\|\partial_x\eta_N(s)\|_{N,'} \le C\|e_N(s)\|^2 + \frac{1}{6}|\eta_N(s)|_r^2.
\end{aligned}
$$

By (3.16), (3.25), (3.26), (3.12), and (3.7)

$$
\begin{aligned}
|(f_1(s),\eta_N(s))| &\le |((I-\mathcal{I}_{N-2r+2}^r)F(U(s)),\partial_x\eta_N(s))| + |((\mathcal{I}_{N-2r+2}^r-I)F(U(s)),\partial_x\eta_N(s))_{N,'}| \\
&\quad + |(F(U(s))-F(u^*(s)),\partial_x\eta_N(s))_{N,'}| \\
&\le C(\|(I-\mathcal{I}_{N-2r+2}^r)F(U(s))\| + N^{-1}|(I-\mathcal{I}_{N-2r+2}^r)F(U(s))|_1 \\
&\quad + C_F\|(\mathcal{I}_{N-2r+2}^r-P_N^{r+1})U(s)\|)|\eta_N(s)|_1 \\
&\le CN^{-2\sigma} + \frac{1}{6}|\eta_N(s)|_r^2,
\end{aligned}
$$

provided that $F(z)\in C^\sigma(\mathbb{R})$. Noting that $P_N^{(0)}U_0=\mathcal{I}_N^rU_0$ in this case,

$$
\|e_N(0)\|_\omega \le \|(\mathcal{I}_N^r-P_N^{r+1})U_0\|_\omega \le CN^{-\sigma}\|U_0\|_\sigma.
$$

Also, it is easy to see from (3.6) and (3.7) that

$$
\begin{aligned}
|(f_2(s),\eta_N(s))| &= |((I-P_N^{r+1})\partial_tU(s),(I-\mathcal{P}_{N-2r-2})\eta_N(s))| \\
&\le CN^{-\sigma}\|\partial_tU(s)\|_{\max\{r+1,\sigma-r\}}|\eta_N(s)|_r \\
&\le CN^{-2\sigma}\|\partial_tU(s)\|_{\max\{r+1,\sigma-r\}}^2 + \frac{1}{6}|\eta_N(s)|_r^2.
\end{aligned}
$$

Putting these estimates into (4.2) and denoting

$$
\begin{aligned}
E(t) &= \|e_N(t)\|_\omega^2 + 2\int_0^t \left(r|\eta_N(s)|_r^2 + |\partial_x^r\eta_N(-1,s)|^2\right)\,ds, \\
\rho(t) &= C\left(\int_0^t \|\partial_tU(s)\|_{\max\{r+1,\sigma-r\}}^2\,ds + \max_{0\le s\le t}\|U(s)\|_\sigma^2\right)N^{-2\sigma},
\end{aligned}
$$

we obtain

$$E(t) \le \rho(t) + C\int_0^t E(s)\,ds, \qquad 0 < t \le T.$$

On the use of Lemma 3.8, we get

$$E(t) \le \rho(t)\mathrm{e}^{Ct}, \qquad 0 < t \le T.$$

Thus, we arrive at the following convergence result via the triangle inequality and (3.7).

Theorem 4.1 *Assume that* $F(z) \in C^1(\mathbb{R})$ *(or* $F(z) \in C^\sigma(\mathbb{R})$ *if* $P_N = \mathcal{I}_N, \mathcal{I}'_N$*),* $\sigma \ge r+1$*, and*

$$U \in C(0,T; H_0^{r,r+1}(I) \cap H^\sigma(I)) \cap H^1(0,T; H_0^{r,r+1}(I) \cap H^{\max\{r+1,\sigma-r\}}(I)).$$

Then for $0 \le t \le T$*,*

$$\|u_N(t) - U(t)\| \le 2\|u_N(t) - U(t)\|_\omega \le CN^{-\sigma}.$$

Remark 4.1 *By the inverse property* (3.9) *we can get an* $H^1_{\omega_{0,1}}$*-estimate from* (3.7) *such that*

$$\begin{aligned}\|\partial_x[u_N(t) - U(t)]\|_{\omega_{0,1}} &\le \|\partial_x e_N(t)\|_{\omega_{0,1}} + \|\partial_x[(P_N^{r+1} - I)U(t)]\|_{\omega_{0,1}} \\ &\le C(N\|e_N(t)\|_\omega + N^{1-\sigma}) \le CN^{1-\sigma}.\end{aligned}$$

5 The convergence of the fully discrete scheme

This section is devoted to the convergence analysis for the fully discrete scheme of the LPG and collocation method (2.3).

Let $u_*^k = P_N^{r+1}U^k$ and $e_N^k = u_N^k - u_*^k$. By (2.1) and (2.3), we have for any $v \in \mathcal{W}_{N-1}$ that

$$\begin{cases} (e^k_{N\hat{t}}, v) + (\partial_x P_N \tilde{F}^k_N, v) - (\partial_x^{r+1}\hat{e}^k_N, \partial_x^r v) = (f^k, v), \quad 1 \le k \le n_T - 1, \\ (e^1_N, v) = (P_N^{(0)}[U_0 + \tau\partial_t U(0)] - u^1_*, v), \\ (e^0_N, v) = (P_N^{(0)}U_0 - u^0_*, v), \end{cases} \tag{5.1}$$

where with the notation $v^k_{t\bar{t}} := \dfrac{1}{\tau^2}(v^{k+1} - 2v^k + v^{k-1})$

$$\tilde{F}^k_N = F(u^k_N) - F(u^k_*),$$

$$f^k = \partial_x[F(U^k) - P_N F(u^k_*)] + \frac{\tau^2}{2}\partial_x F(U^k)_{t\bar{t}} + (\partial_t \hat{U}^k - u^k_{*\hat{t}}) := f^k_1 + f^k_2 + f^k_3.$$

Let $\eta_N^k = e_N^k \omega \in \mathcal{W}_{N-1}$. Taking $v = 2\hat{\eta}_N^k$ in (2.3), we get as in (4.2) that

$$(\|e_N^k\|_\omega^2)_{\hat{t}} + (2r+1)|\hat{\eta}_N^k|_r^2 + 2|\partial_x^r \hat{\eta}_N^k(-1)|^2 \le 2|(f^k - \partial_x P_N \tilde{F}_N^k, \hat{\eta}_N^k)|. \tag{5.2}$$

For given $0 < n \le n_T$, we assume that $\max\limits_{1\le k\le n-1} \|e_N^k\| \le N^{-1}$ (see the condition (ii)$'$ of Lemma 3.9 in Remark 3.1). Then by (3.27),

$$\|e_N^k\|_{L^\infty(I)} \le C_1 N \|e_N^k\| \le C_1, \quad 1 \le k \le n-1.$$

Consider $1 \le k \le n-1$. As in *Case* 1–*Case* 3 of Section 4, we have

$$|(\partial_x P_N \tilde{F}_N^k, \hat{\eta}_N^k)| \le C\|e_N^k\|^2 + \frac{1}{8}|\hat{\eta}_N^k|_r^2,$$
$$|(f_1^k, \hat{\eta}_N^k)| \le CN^{-2\sigma} + \frac{1}{8}|\hat{\eta}_N^k|_r^2.$$

Assume that $F(z) \in C^2(\mathbb{R})$. Since

$$|(f_2^k, \hat{\eta}_N^k)| \le \tau^2 \|F(U^k)_{t\bar{t}}\| |\hat{\eta}_N^k|_1 \le C\tau^4 \|F(U^k)_{t\bar{t}}\|^2 + \frac{1}{8}|\hat{\eta}_N^k|_r^2,$$

we need to bound

$$\begin{aligned}\tau \sum_{k=1}^{n-1} \|F(U^k)_{t\bar{t}}\|^2 &\le C\|\partial_t^2 F(U)\|_{L^2(0,T;L^2(I))}^2 \le CC_F'(\|U\|_{H^2(0,T;L^2(I))}^2 + \|\partial_t U\|_{L^4(I\times(0,T))}^4) \\ &\le CC_F'(\|U\|_{H^2(0,T;L^2(I))}^2 + \|\partial_t U\|_{H^1(I\times(0,T))}^4),\end{aligned}$$

where

$$C_F' = \max_{|z|\le \|U\|_{C(\bar{I}\times[0,T])}} \{|\partial_z F(z)|^2,\ |\partial_z^2 F(z)|^2\}.$$

We introduce the notation

$$\|u\|_{H^{-r}(I)} = \sup_{v\in H_0^r(I), v\neq 0} \frac{|(u,v)|}{|v|_r}.$$

The following estimate can be obtained in the same way as in [19] such that

$$|(f_3^k, \hat{\eta}_N^k)| \le C\left(\|\partial_t \hat{U}^k - U_{\hat{t}}^k\|_{H^{-r}(I)}^2 + N^{-2\sigma}\|U_{\hat{t}}^k\|_{\max\{r+1,\sigma-r\}}^2\right) + \frac{1}{8}|\hat{\eta}_N^k|_r^2\,,$$

and

$$\tau \sum_{k=1}^{n-1} \|\partial_t \hat{U}^k - U_{\hat{t}}^k\|_{H^{-r}(I)}^2 \le C\tau^4 \|\partial_t^3 U\|_{L^2(0,T;H^{-r}(I))}^2,$$
$$\tau \sum_{k=1}^{n-1} \|U_{\hat{t}}^k\|_{\max\{r+1,\sigma-r\}}^2 \le C\|\partial_t U\|_{L^2(0,T;H^{\max\{r+1,\sigma-r\}}(I))}^2.$$

For the initial errors, $\|e_N^0\|_\omega$ is bounded as in Section 4. For $\|e_N^1\|_\omega$, if $P_N^{(0)} = \mathcal{P}_N$, we have from the Taylor's formula and (3.7) that

$$\|e_N^1\|_\omega \le \tau^2 \|\partial_t^2 U\|_{C(0,\tau;L_\omega^2(I))} + \|(I - P_N^{r+1})U^1\|_\omega \le C(\tau^2 + N^{-\sigma}),$$

and if $P_N^{(0)} = \mathcal{I}_N$ or $P_N^{(0)} = \mathcal{I}_N^r$, we have from (3.11), (3.12), and (3.7) that

$$\begin{aligned}\|e_N^1\|_\omega &\le \|(P_N - I)U_0\|_\omega + \tau\|(P_N - I)\partial_t U(0)\|_\omega + \tau^2\|\partial_t^2 U\|_{C(0,\tau;L_\omega^2(I))} + \|(I - P_N^{r+1})U^1\|_\omega \\ &\le C(\tau^2 + \tau N^{-\sigma/2} + N^{-\sigma}) \le C(\tau^2 + N^{-\sigma}),\end{aligned}$$

provided that $\partial_t U(0) \in H^{\sigma/2}(I)$.

Substituting the above estimates into (5.2) and denoting

$$\begin{aligned}E^n &= \|e_N^n\|_\omega^2 + 2\tau \sum_{k=1}^{n-1} \left(r|\eta_N^k|_r^2 + |\partial_x^r \eta_N^k(-1)|^2 \right), \\ \rho^n &= C(\tau^4 + N^{-2\sigma}),\end{aligned}$$

we obtain

$$E^n \le \rho^n + C\tau \sum_{k=1}^{n-1} E^k, \qquad 0 < n \le n_T.$$

Let $\tau\sqrt{N} \le c_0$ being sufficiently small to meet the condition (iv) of Lemma 3.9. Then by Lemma 3.9 with Remark 3.1, we get

$$E^n \le \rho^n \mathrm{e}^{Cn\tau}, \qquad 0 < n \le n_T.$$

Theorem 5.1 *Assume that $F(z) \in C^2(\mathbb{R})$, $\tau\sqrt{N} \le c_0$ being sufficiently small, $\sigma \ge r+1$, and*

$$U \in C(0,T; H_0^{r,r+1}(I) \cap H^\sigma(I)) \cap H^1(0,T; H_0^{r,r+1}(I) \cap H^{\max\{r+1,\sigma-r\}}(I)) \cap H^3(0,T; L_\omega^2(I)).$$

In addition, if $P_N = \mathcal{I}_N$ or $P_N = \mathcal{I}_N'$, assume that $F(z) \in C^\sigma(\mathbb{R})$ and $\partial_t U(0) \in H^{\sigma/2}(I)$. Then for $0 < n \le n_T$,

$$\|u_N^n - U^n\| \le 2\|u_N^n - U^n\|_\omega \le C(\tau^2 + N^{-\sigma}).$$

6 Conclusion

In this paper, we have presented a class of Legendre Petrov–Galerkin and collocation methods for the generalized Korteweg–de Vries equations with non-periodic boundary conditions. This non-symmetric approach is more suitable to the $(2r+1)th$-order differential equations and leads to optimal error estimates. Combining the Legendre Petrov–Galerkin method with a collocation treatment of the nonlinear term, the approximation scheme can be solved efficiently.

To demonstrate the efficiency of the Legendre Petrov–Galerkin method for the problem (1.1), we only used the second-order Crank–Nicolson scheme in time discretization. High-order time integration techniques [9, 11] can be designed for the semi-discrete scheme (2.2). More detailed discussion on high-order time integration methods for semi-discrete approximations of general time-dependent partial differential equations can be found in a review paper [16] by Levy and Tadmor.

References

[1] K. Abe and O. Inoue, *Fourier expansion solution of the KdV equation*, J. Comput. Phys., 34 (1980), pp. 202–210.

[2] B. R. Alpert and V. Rokhlin, *A fast algorithm for the evaluation of Legendre expansions*, SIAM J. Sci. Stat. Comput., 12 (1991), pp. 158–179.

[3] C. Bernardi and Y. Maday, *Polynomial interpolation results in Sobolev space*, J. Comput. & Appl. Math., 43 (1992), pp. 53–82.

[4] C. Bernardi and Y. Maday, *Spectral methods*, in Handbook of Numerical Analysis, Vol. V, Techniques of Scientific Computing (Part 2), P. G. Ciarlet and J. L. Lions, eds., North-Holland, Amsterdam, 1997, pp. 209–486.

[5] N. Bressan and D. Pavoni, *Truncation versus mapping in the spectral approximation to the Korteweg-de Vries equation*, Comput. Methods. Appl. Mech. Engrg., 80 (1990), pp. 443–450.

[6] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral Methods in Fluid Dynamics*, Springer-Verlag, Berlin, 1988.

[7] G. F. Carey and Y. Shen, *Approximations of the KdV equation by least squares finite elements*, Comput. Methods Appl. Mech. Engrg., 93 (1991), pp. 1–11.

[8] T. F. Chan and T. Kerkhoven, *Fourier methods with extended stability intervals for the Korteweg-de Vries equation*, SIAM J. Numer. Anal., 22 (1985), pp. 441–454.

[9] E. Coutsias, T. Hagstrom, J. Hesthaven, and D. Torres, *Integration preconditioners for differential operators in spectral tau-methods*, in Proceedings of the Third International Conference on Spectral and High Order Methods, A. V. Ilin and L. R. Scott, eds., Department of Mathematics, University of Houston, Houston, 1996, TX, pp. 21–38.

[10] K. Djidjeli, W. G. Price, E. H. Twizell, and Y. Wang, *Numerical methods for the solution of the third- and fifth-order dispersive Korteweg-de Vries equations*, J. Comput. Appl. Math., 58 (1995), pp. 307–336.

[11] B. Fornberg and T. Driscoll, *A fast spectral algorithm for nonlinear wave equations with linear dispersion*, J. Comput. Phys., 155 (1999), pp. 456–467.

[12] B. Fornberg and G. B. Whitham, *A numerical and theoretical study of certain nonlinear phenomena*, Phil. Trans. Roy. Soc., 289 (1978), pp. 373–404.

[13] B.-Y. Guo, *Difference Methods for Partial Differential Equations*, Science Press, Beijing, 1988.

[14] W. Z. Huang and D. M. Sloan, *The pseudospectral method for third-order differential equations*, SIAM J. Numer. Anal., 29 (1992), pp. 1626–1647.

[15] P.-Y. Kuo, *Error estimations of the spectral method for solving K. d. V. -Burgers equation*, Talk in Madrid University, March 1981, also see Acta Mathematica Sinica, 28 (1985), pp. 1–15.

[16] D. Levy and E. Tadmor, *From semi-discrete to fully-discrete: stability of Runge-Kutta schemes by the energy method*, SIAM Rev., 40 (1998), pp. 40–73.

[17] J. Li, H.-P. Ma, and W.-W. Sun, *Error analysis for solving the Korteweg-de Vries equation by a legendre pseudo-spectral method*, Numer. Methods Partial Differential Equations, 16 (2000), pp. 513–534.

[18] H.-P. Ma and B.-Y. Guo, *The Fourier pseudospectral method with a restrain operator for the Korteweg-de Vries equation*, J. Comput. Phys., 65 (1986), pp. 120–137.

[19] H.-P. Ma and W.-W. Sun, *A Legendre Petrov-Galerkin and Chebyshev collocation method for third-order differential equations*, SIAM J. Numer. Anal., 38 (2000), pp. 1425–1438.

[20] ——, *Optimal error estimates of the Legendre Petrov–Galerkin method for the Korteweg–de Vries equation*, SIAM J. Numer. Anal., 39 (2001), pp. 1380–1394.

[21] Y. Maday and A. Quarteroni, *Error analysis for spectral approximation of the Korteweg-de Vries equation*, M^2AN, 22 (1988), pp. 499–529.

[22] W. J. Merryfield and B. Shizgal, *Properties of collocation third-derivative operators*, J. Comput. Phys., 105 (1993), pp. 182–185.

[23] D. S. Mitrinović, J. E. Pečarić, and A. M. Fink, *Inequalities Involving Functions and Their Integrals and Derivatives*, Kluwer Academic Publishers, Dordrecht, 1991.

[24] D. Pavoni, *Single and multidomain Chebyshev collocation methods for the Korteweg-de Vries equation*, Calcolo, 25 (1988), pp. 311–346.

[25] J. Shen, *Efficient spectral-Galerkin method I. Direct solvers for second- and fourth-order equations using Legendre polynomials*, SIAM J. Sci. Comput., 15 (1994), pp. 1489–1505.

[26] G. Szegö, *Orthogonal Polynomials*, American Mathematics Society, Providence, RI, 1975.

HAUSDORFF MEASURES OF HOMOGENEOUS CANTOR SETS

Chengqin Qu
Department of Mathematics, Zhongshan University, Guangzhou 510275, P.R. China
E-mail: chengqinqu@163.net
Hui Rao
Department of Mathematics, Wuhan University, Wuhan 430072, P.R.China
E-mail: raohui@nlsc.whu.edu.cn
Weiyi Su
Department of Mathematics, Nanjing University, Nanjing 210093, P.R.China
E-mail:suqiu@netra.nju.edu.cn

Suppose $I = [0, 1]$, let $\{n_k\}_{k\geq 1}$ be a sequence of positive integers, and $\{c_k\}_{k\geq 1}$ be real number sequence satisfying $n_k \geq 2,\ 0 < n_k c_k \leq 1 (k \geq 1)$. For any $k \geq 1$, let $D_k = \{(i_1, \cdots, i_k) :\ 1 \leq i_j \leq n_j, 1 \leq j \leq k\},\ D = \bigcup_{k\geq 0} D_k$, where $D_0 = \emptyset$. If $\sigma = (\sigma_1, \cdots, \sigma_k) \in D_k$, $\tau = (\tau_1, \cdots, \tau_m) \in D_m$, let $\sigma * \tau = (\sigma_1, \cdots, \sigma_k, \tau_1, \cdots, \tau_m)$. Let $\mathcal{F} = \{I_\sigma :\ \sigma \in D\}$ be the collection of the closed sub-intervals of I which satisfy

i) $I_\emptyset = I$;

ii) For any $k \geq 1$ and $\sigma \in D_{k-1}$, $I_{\sigma * i}$ $(1 \leq i \leq n_k)$ are sub-intervals of I_σ. $I_{\sigma*1}, \cdots, I_{\sigma*n_k}$ are arranged from the left to the right, $I_{\sigma*1}$ and I_σ have the same left endpoint, $I_{\sigma*n_k}$ and I_σ have the same right endpoint, and the lengths of the gaps between any two consective sub-intervals are equal. We denote the length of one of the gaps by y_k.

iii) For any $k \geq 1$ and $\sigma \in D_{k-1}$, $1 \leq j \leq n_k$, we have

$$\frac{|I_{\sigma*j}|}{|I_\sigma|} = c_k,$$

where $|A|$ denotes the diameter of A.

Let $E_k = \bigcup_{\sigma\in D_k} I_\sigma$, $E = \bigcap_{k\geq 0} E_k$, we call E the homogeneous Cantor set ([5,6]) determined by $\{n_k\}_{k\geq 1},\ \{c_k\}_{k\geq 1}$ and call $\mathcal{F}_k = \{I_\sigma :\ \sigma \in D_k\}$ the k-order basic intervals of E.

For the homogeneous Cantor set E, its Hausdorff dimension was given ([5])

$$\dim_H(E) = \liminf_{k\to\infty} \frac{\log n_1 \cdots n_k}{-\log c_1 \cdots c_k}.$$

In this paper we determine the exact Hausdorff measures of a class of homogeneous Cantor sets, i.e.

Theorem 1. *Let E be the homogeneous Cantor set determined by $\{n_k\}_{k\geq 1}$, $\{c_k\}_{k\geq 1}$, if $y_{k+1} \leq y_k$ for all $k \geq 1$, then*

$$\mathcal{H}^s(E) = \liminf_{k\to\infty} \prod_{j=1}^{k} n_j c_j^s,$$

where s is the Hausdorff dimension of E.

In order to prove Theorem 1 we need the following lemmas.

Let G_k be a set consisting of all possible union of elements in $\mathcal{F}_k$, define

$$G = \bigcup_{k=0}^{\infty} G_k, \tag{1}$$

and

$$\mathcal{H}_G^\alpha(E) = \liminf_{\delta\to 0}\{\sum |U_i|^\alpha : \ E \subset \cup U_i, \ |U_i| < \delta \text{ and } U_i \in G\}. \tag{2}$$

Lemma 1 ([2]). *Let $\mathcal{H}^\alpha(E)$ be the α-dimensional Hausdorff measure of E, then $\mathcal{H}^\alpha(E) = \mathcal{H}_G^\alpha(E)$.*

For any $\sigma = (\sigma_1, \cdots, \sigma_m) \in D_m$, when $0 < k \leq m$, we denote $\sigma|k = (\sigma_1, \cdots, \sigma_k)$. Let x_k be the length of k-order basic interval, y_k the length of the gap between any two consecutive sub-intervals $I_{\sigma * i}$ and $I_{\sigma*(i+1)}$, where $\sigma \in D_{k-1}$, $1 \leq i \leq n_k - 1$. For any $\sigma, \tau \in D_k$, let $a(\sigma)$ be the left endpoint of I_σ, $b(\tau)$ the right endpoint of I_τ. Take $B = \liminf_{k\to\infty} \prod_{j=1}^{k} n_j c_j^s$. Let μ be the probability measure supported by E, such that for any $A \in \mathcal{F}_k$, we have

$$\mu(A) = (n_1 \cdots n_k)^{-1}.$$

Lemma 2. *Let E be the homogeneous Cantor set satisfying the condition of Theorem 1, and $0 < B < \infty$. Then for any $\epsilon > 0$ there exists $k_0 \in \mathbb{N}$, such that for any $\sigma, \tau \in D_k (k \geq k_0)$ with $a(\sigma) < b(\tau)$ and $\sigma|k_0 = \tau|k_0$ holds*

$$\mu([a(\sigma), b(\tau)]) \leq (B - \epsilon)^{-1} (b(\tau) - a(\sigma))^s. \tag{3}$$

Proof. Since $\liminf_{k\to\infty} (n_1 \cdots n_k)(c_1 \cdots c_k)^s = B$, then for any $\epsilon > 0$ there exists $k_0 > 0$ such that for $k \geq k_0$ we have

$$(n_1 \cdots n_k)^{-1} \leq (B - \epsilon)^{-1} (c_1 \cdots c_k)^s. \tag{4}$$

If $\sigma = \tau \in D_k (k \geq k_0)$, then $b(\tau) - a(\sigma) = c_1 \cdots c_k$, and it follows immediately from (4) that

$$\mu([a(\sigma), b(\tau)]) \leq (B - \epsilon)^{-1} (b(\tau) - a(\sigma))^s.$$

So we need only consider the case that $\sigma, \tau \in D_k (k > k_0)$, $\sigma \neq \tau$. In this case we will prove (3) by induction.

(1) For $n = k_0 + 1$, suppose that $[a(\sigma), b(\tau)]$ contains i $(k_0 + 1)$-order basic intervals. Since $\sigma|k_0 = \tau|k_0$ and $\sigma \neq \tau$, we have $2 \leq i \leq n_{k_0+1}$. On the other hand

$$(b(\tau) - a(\sigma))^s = \left(\frac{n_{k_0+1} - i}{n_{k_0+1} - 1} c_{k_0+1} + \frac{i - 1}{n_{k_0+1} - 1} \right)^s (c_1 \cdots c_{k_0})^s. \tag{5}$$

Using the convexity of x^s we have

$$(b(\tau) - a(\sigma))^s \geq \frac{n_{k_0+1} - i}{n_{k_0+1} - 1} (c_1 \cdots c_{k_0+1})^s + \frac{i - 1}{n_{k_0+1} - 1} (c_1 \cdots c_{k_0})^s. \tag{6}$$

From (4) and (6) we obtain (3).

(2) Now suppose that (3) holds for $n = k (> k_0)$, we will deal with the case for $n = k + 1$ in the following two cases

1° $\sigma, \tau \in D_{k+1}$, $\sigma|k = \tau|k$. In this case, similar to the proof of (1) we have (3).

2° $\sigma, \tau \in D_{k+1}$, $\sigma|k \neq \tau|k$.

i) If $a(\sigma) = a(\sigma|k)$ and $b(\tau) \neq b(\tau|k)$, then $I_{\tau|k}$ lies to the right of $I_\sigma|k$ and $i = \tau_{k+1} < n_{k+1}$. Let $\hat{\tau} \in D_k$, and $I_{\hat{\tau}}$ is the k-order basic interval immediately to the left of $I_{\tau|k}$. Since $y_1 \geq y_2 \geq \cdots \geq y_k$, so the gap between $I_{\hat{\tau}}$ and $I_{\tau|k}$ is of size at least y_k. Put

$$\lambda = (b(\tau|k) - b(\tau)) / (b(\tau|k) - b(\hat{\tau})), \ 0 < \lambda < 1. \tag{7}$$

Then

$$b(\tau) - a(\sigma) = b(\tau) - a(\sigma|k) = \lambda (b(\hat{\tau}) - a(\sigma|k)) + (1 - \lambda)(b(\tau|k) - a(\sigma|k)). \tag{8}$$

Since I_σ, I_τ lie in the k_0-order basic interval $I_{\tau|k_0}$, we have that $I_{\sigma|k}$, $I_{\tau|k}$ and $I_{\hat{\tau}}$ also lie in $I_{\tau|k_0}$. Using the convexity of x^s and (8), we obtain

$$(b(\tau) - a(\sigma))^s \geq \lambda (b(\hat{\tau}) - a(\sigma|k))^s + (1 - \lambda)(b(\tau|k) - a(\sigma|k))^s. \tag{9}$$

By inductive assumption and (9), we have

$$(b(\tau) - a(\sigma))^s \geq (B - \epsilon)(\lambda \mu([a(\sigma|k), b(\hat{\tau})]) + (1 - \lambda) \mu([a(\tau|k), b(\tau|k)])). \tag{10}$$

The condition $y_{k+1} \leq y_k$ implies

$$x_k + y_k \geq x_k + y_{k+1} = n_{k+1}(x_{k+1} + y_{k+1}). \tag{11}$$

From (7) and (11) we have

$$\lambda \leq (n_{k+1}-i)(x_{k+1}+y_{k+1})/(x_k+y_k) \leq 1-\frac{i}{n_{k+1}}. \tag{12}$$

From (10) and (12) we obtain (3).

Using the same method, we also have (3) in the case that $b(\tau) = b(\tau|k)$ and $a(\sigma) \neq a(\sigma|k)$.

ii) If $a(\sigma) \neq a(\sigma|k)$ and $b(\tau) \neq b(\tau|k)$, we have $1 < \sigma_{k+1}$ and $\tau_{k+1} < n_{k+1}$.

a) If $\sigma_{k+1} > \tau_{k+1}$, let $\sigma' \in D_k$, and $I_{\sigma'}$ is the first k-order basic interval to the right $I_{\sigma|k}$. Let $\tau' \in D_{k+1}$, $\tau'|k = \tau|k$, and $\tau'_{k+1} = \tau_{k+1} + n_{k+1} - \sigma_{k+1} + 1$, in this case, $J = [a(\sigma'), b(\tau')]$ and

$$b(\tau) - a(\sigma) \geq b(\tau') - a(\sigma'),$$

$$\mu([a(\sigma), b(\tau)]) = \mu([a(\sigma'), b(\tau')]).$$

By i) we have (3).

b) If $\sigma_{k+1} \leq \tau_{k+1}$, let $\tau' \in D_{k+1}$, $\tau'|k = \tau|k$ and $\tau'_{k+1} = \tau_{k+1} - \sigma_{k+1} + 1$, in this case $J = [a(\sigma|k), b(\tau')]$, and

$$b(\tau) - a(\sigma) = b(\tau') - a(\sigma|k),$$

$$\mu([a(\sigma), b(\tau)]) = \mu([a(\sigma|k), b(\tau')]).$$

By i) we also have (3). This completes the proof of Lemma 2.

Proof of Theorem 1. By [5] , if $B = 0$ or ∞ we have $\mathcal{H}^s(E) = 0$ or ∞ respectively. Therefore it suffices to prove Theorem 1 in the case for $0 < B < +\infty$.

Let $\{U_i\}_i \subset G$ is a δ-covering of E, where $\delta \leq \min\{x_{k_0}, y_{k_0}\}$. For each U_i there exists $\sigma \in D_{k_0}$ such that $U_i \subset I_\sigma$. Suppose U_i is the union of some $k(i)$-order basic intervals, i.e. $U_i = I_{\sigma^{(i)}} \cup \cdots \cup I_{\tau^{(i)}}$, where $= I_{\sigma^{(i)}}$ and U_i have the same left endpoint, $I_{\tau^{(i)}}$ and U_i have the same right endpoint. Then

$$|U_i| = b(\tau^{(i)}) - a(\sigma^{(i)}).$$

By Lemma 2, we have

$$\begin{aligned} 1 = \mu(E) \quad & \leq \textstyle\sum_i \mu(U_i) \leq \sum_i \mu([a(\sigma^{(i)}), b(\tau^{(i)})]) \\ & \leq (B-\epsilon)^{-1} \textstyle\sum (b(\tau^{(i)}) - a(\sigma^{(i)}))^s = (B-\epsilon)^{-1} \sum_i |U_i|^s, \end{aligned}$$

by Lemma 1 we have

$$\mathcal{H}^s(E) = \mathcal{H}^s_G(E) \geq B - \epsilon.$$

Since ϵ is arbitrary we obtain $\mathcal{H}^s(E) \geq B$. On the other hand, we obviously have $\mathcal{H}^s(E) \leq B$, thus $\mathcal{H}^s(E) = B$, which complete the proof of Theorem 1.

References

[1] Falconer K.J. *The geometry of fractal sets.* Cambridge University Press, 1985.

[2] Marion D.J. *Measures de Hausdorff et theorie de Perron-Frobinius des matrice non-negatives.* Ann. Inst. Fourier, Grenoble, 1985, 35(4): 99-125.

[3] Zhou Z.L. and Wu M. *The Hausdorff measure of Siepinski carpet.* Science in China(Series A) , 1990, 42(7):673-680.

[4] Mauldin R.D. and Williams S.C. *Scaling Hausdorff measures.* Mathematika, 1989, 36: 325-333.

[5] Feng D.J. Rao H. and Wu J. *The net measure properties of one dimensional homogeneous Cantor set and its applications. Progr. Nat. Sci., 1996, 6(6): 673-678.*

[6] *Feng D.J. Wen Z.Y. Wu J. Dimension of the homogeneous Moran sets. Science in China (Series A), 1997, 40: 475-482.*

Best approximation of a compact convex set by a ball in an arbitrary norm*

S.I.Dudov and I.V.Zlatorunskaya

The problem of the best approximation of a compact convex set by a ball with respect to an arbitrary norm in the Hausdorff metric corresponding to that norm is considered. This problem is reduced to a convex programming problem, which can be studied by means of convex analysis.

Necessary and sufficient conditions for a solution of this problem are obtained and several properties of its solution are described. It is proved, in particular, that the center of at least one ball of the best approximation lies in the approximated compact set. Conditions ensuring that the centers of all balls of the best approximation lie in this compact set and conditions for uniqueness of the solution are obtained. In addition, several variational properties of the solution are studied and it is shown that the problem may be reduced to a linear programming problem in the case when the approximated set and the ball of the given norm are polytopes.

1 Introduction

Estimation and approximation problems of relatively complex sets by sets of a simple structure have broad applications in natural sciences and in mathematics itself, presenting a branch of nonsmooth analysis.

As simple sets that locally approximate a given set,one usually takes cones of feasible (or tangent, or contingent) directions[1,2,3,4] or bundles of higher order curves[5,6]. Problems of nonlocal set approximation involve different problems of the set estimation, such as problems of outer and inner estimation. Simple sets are usually taken

*The work was supported by the Russian Foundation for Basic Researches (grant no. 00–15–96123).

[1]B.N.Pshenichnyi, Convex analysis and extremal problems, Nauka, Moscow 1980. (Russian)

[2]V.F. Dem'yanov and L.V. Vasil'ev, Nondifferentiable optimization, Nauka, Moscow 1981; English transl., Optimization Software, New York 1985.

[3]V.F. Dem'yanov and A.M.Rubinov, Constractive nonsmooth analysis, Frankfurt a/M, Verlag Peter Lang, 1995.

[4]F.H. Clarke, Optimization and nonsmooth analysis, Wiley, New York 1983.

[5]A.Ya.Dubovitskii and A.A.Milyutin, "Extremal problems in the presence of constraints", Comp. Maths and Math. Phys., 5, (1965), 395–453. (Russian).

[6]U.Ledzewicz and H.Shaettler,"High–Order Approximations and generalized necessary conditions for optimality", SIAM J. on Control and Optimization, 37:1, (1998), 33–53.

in the form of ellipsoid or polytope. The known publications are mostly connected just with these cases.

Alongside with the ellipsoid and the polytope, the ball with respect to an arbitrary norm can be also considered as the simplest set, both in the geometric sense and in the number of required parameters.

The problem of outer estimation of a given compact set by a ball in arbitrary norm was considered by B.Pshenichnyi[1]. It consists in constructing the ball of the smallest radius that contains the estimated compact set. The corresponding inner estimation problem was also considered[7].

In this paper we consider another estimation problem of a given convex compact set by a ball of an arbitrary norm, consisting in the following. Let D be a given convex compact set in the finite–dimensional real space $\mathbb{R}^p$ an estimate of which we are seeking, let $n(x)$ be a function on $\mathbb{R}^p$ satisfying the axioms of norm, let $\rho(A,B) = \sup\limits_{x\in A}\inf\limits_{y\in B} n(x-y)$ be the deviation of the set A from the set B with respect to the norm $n(\cdot)$, let $h(A,B) = \max\{\rho(A,B), \rho(B,A)\}$ be the Hausdorff distance between A and B corresponding to the norm $n(\cdot)$, and let $Bn(x,r) = \{y \in \mathbb{R}^p : n(x-y) \le r\}$ be the ball in the norm $n(\cdot)$ with center at the point x and the radius r. Then the problem of the best approximation of the convex compact set D by a ball with respect to $n(\cdot)$ in the Hausdorff metric corresponding to this norm can be expressed as follows:

$$h(D, Bn(x,r)) \to \min_{x\in\mathbb{R}^p,\, r\ge 0}. \tag{1.1}$$

The problem (1.1) was considered by M.Nikol'skii and D.Silin[8] in the case when $n(\cdot)$ is the Euclidean norm. In their paper they established the existence and uniqueness of the solution, obtained a necessary condition and described some properties of the solution. The authors pointed out that all their results can be easily extended to the case of a more general 'ellipsoidal' norm because of the simple connection between the solutions of the problems for these norms.

The aim of the present paper is to study the problem (1.1) for an arbitrary norm $n(\cdot)$. This problem can be regarded as a problem of approximation theory. Indeed, a given convex compact set, as an element of the space of all convex compact sets $Kv(\mathbb{R}^p)$, can be approximated by elements of a subset of $Kv(\mathbb{R}^p)$, namely, by balls of the given norm in the above metric. On another hand, (1.1) can be called the uniform estimation problem, in comparison with the outer and inner estimation problems.

It was discovered in the work of M.Nikol'skii and D.Silin[8] that the problem (1.1) is connected with another estimation problem. Namely, in the case of Euclidean norm it is proved that the ball of best approximation is unique, its center x_0 lies in

---

[7] S.I.Dudov, "An inner bound of a convex set by a norm body", Comp. Maths Math. Phys., 36:5, (1996), 683–688.

[8] M.S.Nikol'skii and D.B.Silin, "Best approximation of a convex compact set by elements of addial", Proc. Steklov Inst. Math., 211, (1995), 306–321.

D, and is the unique solution of the problem

$$R(x) - \rho_\Omega(x) \to \min_{x \in D}, \tag{1.2}$$

where $R(x) = \max_{y \in D} n(x-y)$, $\rho_\Omega(x) = \min_{y \in \Omega} n(x-y)$, $\Omega = \overline{\mathbb{R}^p \backslash D}$. The radius of this ball is $r_0 = (R(x_0) + \rho_\Omega(x_0))/2$; moreover

$$\min_{x \in \mathbb{R}^p,\, r \geq 0} h(D, Bn(x,r)) = \frac{R(x_0) - \rho_\Omega(x_0)}{2}.$$

Therefore, we can say in that case that problems (1.1) and (1.2) are equivalent.

From the geometrical point of view, the latter problem is the problem of constructing a spherical layer of the 'least thickness' that contains the boundary of the convex compact set D. Interconnection between these problem is natural. Indeed, the smallest thickness of a spherical layer containing the boundary of D can be taken as a measure of approximation of this set by a ball of the given norm. The thinner this layer the better compact set D can be approximated by a ball.

As was pointed out by M.Nikol'skii and D.Silin[8], problem (1.1) and some close problems were considered[9,10,11,12,13,14,15,16] only in the case of the Euclidean norm. In particular, for $p=2$ T.Bonnesen[12] and for $p=3$ N.Kriticos[15] obtained necessary and sufficient conditions and proved the uniqueness of solution of problem (1.2). Corresponding results for arbitrary p were obtained by I.Barany[1]6, where the means of convex analysis were essentially exploited. It must be emphasized that $R(x)$ is a convex function on $\mathbb{R}^p$ and $\rho_\Omega(x)$ is a concave function on D. Hence (1.2) is a convex problem.

However, as was established by the authors of this paper, problems (1.1) and (1.2) are not necessarily equivalent for any norm. Examples show that some balls of the best approximation may have centers outside D. Moreover, a solution of problem (1.1) is not necessarily unique. Still, we show in section 3 for any norm $n(\cdot)$, (1.1) is equivalent to a convex problem that is slightly different from (1.2). Namely, we establish that (1.1) is equivalent to the following problem

$$R(x) + P(x) \to \min_{x \in \mathbb{R}^p}, \tag{1.3}$$

[9]M. D'Ocagne, "Sur certaine figures minimales", Bull. Soc. Math. France 12 (1884), 168–177.

[10]H. Lebesgue, "Sur quelques questions de minimum, relatives aux courbes orbiformes, et sur leurs rapports avec le calcul des variations", J. Math. Pures Appl. 4 (1921), 67–96.

[11]T. Bonnesen and W. Fenchel, Theorie der konvexen Korper, Springer-Verlag, Berlin 1934.

[12]T. Bonnesen, Uber das isoperimetrische Defizit ebener Figuren, Math. Ann., 91(1924), 252–268

[13]St. Vincze, "Uber den Minimalkreisring einer Eiline", Acta Sci. Math. (Szeged) 11:3 (1947), 133–138.

[14]I. Vincze, "Uber Kreisringe, die eine Eiline einschlissen", Studia. Sci. Math. Hungar. 9:1/2 (1974), 155–159.

[15]N. Kritikos, "Uber konvexe Flachen und einschliessende Kugeln", Math. Ann. 96 (1927),583–586.

[16]I.Barany, "On the minimal ring containing the boundary of convex body", Acta Sci. Math. (Szeged) 52:1/2 (1988), 93–100.

where $P(x) = \rho_D(x) - \rho_\Omega(x)$. The main results are obtained by investigation of properties of the function $\Phi(x) = R(x) + P(x)$.

The paper has the following structure. In section 2 we establish some properties of the function $R(x)$ and of the distance function $\rho_A(x)$, in particular, in the case, when the norm $n(\cdot)$ is a strictly quasiconvex function. In addition, we prove the convexity of the function $P(x)$ on $\mathbb{R}^p$ and derive formula for its subdifferential. In section 4 we obtain a necessary and sufficient condition for solution of problem (1.1), which is a basic tool for the further study. Also we prove here that the center of at least one ball of the best approximation lies in D. In section 5 we present conditions ensuring that the entire set of centers of the balls of the best approximation lies in D. Conditions for the uniqueness of solution are given in section 6. In section 7 several variational properties of the solution are studied. Finally, in section 8 it is shown that the problem may be reduced to a linear programming problem in the case when the approximated set and the balls of the given norm are polytopes.

2 Auxiliary functions and their properties

One would expect that the properties of the norm involved affect the properties of the solutions of (1.1). They can also be crucial for one's choice of analytic tools. For instance there exists a simple formula expressing the distance between sets in the Hausdorff metric generated by the Euclidean norm in terms of the support functions of these sets [17]. This was evidently decisive for the systematic use of support functions techniques in the work of M.Nicol'skii and D.Silin[8]. As regards this paper, the following axiliary functions will be important in what follows:
$R(x) = \max_{y \in D} n(x - y)$, the radius of the smallest ball with center at x containing the set D;
$\rho_A(x) = \min_{y \in A} n(x - y)$, the distance from x to a closed set A;
$P(x) = \rho_D(x) - \rho_\Omega(x)$, where $\Omega = \overline{\mathbb{R}^p \backslash D}$.

In this section we discuss some properties of the norm and these functions; we will use the following notation: $\overline{A}, intA, coA, \partial A$ are the closure, interior, convex hull, and boundary of a set A, respectively;
$\langle v, w \rangle$ is the inner product of elements v and w;

$$A + B = \{a + b : a \in A, b \in B\}; \qquad A - B = \{a - b : a \in A, b \in B\};$$

$$K(A) = \{v \in \mathbb{R}^p : \exists \alpha \geq 0,\, a \in A,\, v = \alpha a\};$$
$$K^+ = \{w \in \mathbb{R}^p : \langle v, w \rangle \geq 0 \quad \forall v \in K\};$$

$K(x, A)$ is the cone of feasible directions of the set A at the point x, that is, $K(x, A) = \overline{\gamma}(x, A)$, where $\gamma(x, A) = \{g \in \mathbb{R}^p : \exists \alpha_g \geq 0 : x + \alpha g \in A,\, \forall \alpha \in (0, \alpha_g)\}$;

$$Q^R(x, D) = \{y \in D : n(x - y) = R(x)\};$$

[17] K.Leichtveiss, Konvexe Mengen, Springer-Verlag, Berlin 1985.

$$Q^{\rho}(x, A) = \{z \in A : n(x - z) = \rho_A(x)\};$$

$W(\psi, A) = \sup\limits_{v \in A} \langle v, \psi \rangle$ is the support function of the set A;

$[x_1, x_2] = co\,\{x_1, x_2\}$ is the closed interval connecting x_1 and x_2;

$\|x\| = (\sum\limits_{i=1}^{p} (x^{(i)})^2)^{1/2}$ is the Euclidean norm of the vector $x \in \mathbb{R}^p$.

1. The norm $n(x)$ is a finite convex function on $\mathbb{R}^p$, therefore it is a continuous function that is differentiable at each point $x \in \mathbb{R}^p$ in each direction $g \in \mathbb{R}^p$, and its subdifferential $\partial n(\cdot)$, regarded as a set-valued map, is upper semicontinuous; in addition,

$$n'(x, g) \equiv \lim_{\alpha \downarrow 0} \alpha^{-1}[n(x + \alpha g) - n(x)] = \max_{v \in \partial n(x)} \langle v, g \rangle.$$

The following formula for the subdifferential of the norm is well-known[1]:

$$\begin{aligned} &\partial n(0_p) = \{v \in \mathbb{R}^p : n^*(v) \leq 1\}, \\ &\partial n(x) = \{v \in \mathbb{R}^p : n^*(v) = 1,\, n(x) = \langle v, x \rangle\}, \quad x \neq 0_p. \end{aligned} \tag{2.1}$$

Here $n^*(\cdot) = \max\limits_{n(v) \leq 1} \langle v, \cdot \rangle$ is the polar norm, $0_p = (0, \ldots, 0) \in \mathbb{R}^p$.

Definition 2.1. We call $n(\cdot)$ a smooth norm if it is continuously differentiable at all points $x \neq 0_p$.

According to the concept of a strictly quasiconvex functions we give here the following definition.

Definition 2.2. We say that a norm $n(\cdot)$ is strictly quasiconvex if

$$n(\alpha x_1 + (1 - \alpha)x_2) < \max\{n(x_1), n(x_2)\} \quad \text{for } x_1 \neq x_2,\ \alpha \in (0, 1). \tag{2.2}$$

Examples of a strictly quasiconvex norm are Euclidean norm, ellipsoidal norms (in which balls are ellipsoids), and each calibration function[18] generated by a strictly convex[1]7 or a strongly convex[19] compact set symmetric with respect to origin. An example of a norm not satisfying Definition 2.2 is the Chebyshev norm $n(x) = \max\limits_{i=1,\ldots,p} |x^{(i)}|$, where $x = (x^{(1)}, \ldots, x^{(p)}) \in \mathbb{R}^p$.

Let us give some simple properties of a strictly quasiconvex norm.

Lemma 2.1. *If $n(\cdot)$ is a strictly quasiconvex norm and elements $x_1 \neq 0_p$ and $x_2 \neq 0_p$ have property $x_2 \neq \lambda x_1$ for each $\lambda > 0$, then*

$$n(\alpha x_1 + (1 - \alpha)x_2) < \alpha n(x_1) + (1 - \alpha)n(x_2), \quad \alpha \in (0, 1). \tag{2.3}$$

Proof. If $n(x_1) = n(x_2)$, then (2.3) is an immediate consequence of (2.2). Assume that $n(x_1) \neq n(x_2)$. Consider $x_3 = \dfrac{n(x_2)}{n(x_1)} x_1$. In view of our assumptions,

[18]R.Rockafellar, Convex analysis, Princeton Univ. Press, Princeton, NJ 1970.

[19]E.S.Polovinkin, "Strongly convex analysis", Mat. sb. 187:2 (1996), 102-130; English transl. in Sb.Math. 187 (1996).

$x_3 \neq x_2$; at the same time $n(x_3) = n(x_2)$. Hence, in accordance with Definition 2.2 we have

$$n(\beta x_3 + (1-\beta)x_2) < n(x_2) \quad \text{for all } \beta \in (0,1). \tag{2.4}$$

We consider an arbitrary $\alpha \in (0,1)$ and associate with it the quantities

$$\beta_0 = \frac{\alpha n(x_1)}{\alpha n(x_1) + (1-\alpha)n(x_2)}, \qquad \lambda_0 = \frac{\alpha n(x_1) + (1-\alpha)n(x_2)}{n(x_2)}.$$

It is easy to see that $\beta_0 \in (0,1)$ and $\alpha x_1 + (1-\alpha)x_2 = \lambda_0(\beta_0 x_3 + (1-\beta)x_2)$. Now, using inequality (2.4), we can deduce (2.3):

$$n(\alpha x_1 + (1-\alpha)x_2) = \lambda_0 n(\beta_0 x_3 + (1-\beta_0)x_2) < \lambda_0 n(x_2) = \alpha n(x_1) + (1-\alpha)n(x_2).$$

□

LEMMA 2.2. *If $n(\cdot)$ is a strictly quasiconvex norm, then the equality*

$$n(x_2) = n(x_1) + n(x_1 - x_2) \tag{2.5}$$

holds for $x_1 \neq 0_p$ if and only if there exists $\lambda \geq 1$ such that $x_2 = \lambda x_1$.

PROOF. The sufficiency is obvious. We now establish the necessity. Assume that equality (2.5) holds and $x_1 \neq 0_p$. If the assertion of Lemma 2.2 is false, then using Lemma 2.1 we obtain inequality (2.3). From it and (2.5) we deduce the inequality

$$n(x_2) - n(\alpha x_1 + (1-\alpha)x_2) > \alpha n(x_1 - x_2), \qquad \alpha \in (0,1),$$

which contradicts the triangle inequality for the norm $n(\cdot)$. □

LEMMA 2.3. *If $n(\cdot)$ is a smooth norm, then the polar norm $n^*(\cdot)$ is a strictly quasiconvex norm.*

PROOF. Since $n(x)$ is the convex positively homogeneous function, it is sufficiently to prove (2.2) for any $\alpha \in (0,1)$ and $x_1 \neq x_2$ such that $n^*(x_1) = n^*(x_2) = 1$.

Assume the contrary, that is, there exist the points $x_1 \neq x_2$, such that $n^*(x_1) = n^*(x_2) = 1$, and $\alpha_0 \in (0,1)$, for which $n^*(\alpha_0 x_1 + (1-\alpha_0)x_2) = 1$. Then by convexity of the function $n^*(\cdot)$ it follows

$$n^*(\alpha x_1 + (1-\alpha)x_2) = 1 \quad \text{for all } \alpha \in [0,1]. \tag{2.6}$$

It means that the line segment $[x_1, x_2]$ belongs to the boundary of the ball $Bn^*(0_p, 1) = \{y \in \mathbb{R}^p : n^*(y) \leq 1\}$. Let us build the support hyperplane to the ball $Bn^*(0_p, 1)$

$$\pi = \{x \in \mathbb{R}^p : \langle v_0, x \rangle = \lambda\},$$

which contains the line segment $[x_1, x_2]$. Without loss of generality we can assume $\lambda > 0$. Then, in accordance with construction, we have

$$\langle v_0, \alpha x_1 + (1-\alpha)x_2 \rangle = \lambda, \quad \alpha \in (0,1), \tag{2.7}$$

$$\langle v_0, y\rangle \le \lambda \quad \text{for all } y \in Bn^*(0_p, 1). \tag{2.8}$$

It is known[18] that $n(v) = \max\limits_{n^*(x)\le 1} \langle v, x\rangle$ for all $v \in \mathbb{R}^p$. Therefore using (2.6) and (2.7) we obtain

$$n(v_0) = \max_{n^*(x)\le 1} \langle v_0, x\rangle \ge \langle v_0, \alpha x_1 + (1-\alpha)x_2\rangle = \lambda.$$

On the other hand, by (2.8) it follows $n(v_0) \le \lambda$. Thus we conclude

$$n(v_0) = \lambda. \tag{2.9}$$

Now from (2.6), (2.7) and (2.9), in accordance the formula (2.1), it follows $\alpha x_1 + (1-\alpha)x_2 \in \partial n(v_0)$ for all $\alpha \in [0,1]$. This contradicts the smoothness of norm $n(\cdot)$, since the subdifferential of a smooth norm at any point $v \ne 0_p$ consists of the only element.

2. In this subsection we deal with some properties of the function $R(x)$.

Lemma 2.4. *The function $R(x)$ is globally Lipschitz on $\mathbb{R}^p$, and for all $x, y \in \mathbb{R}^p$ the following inequality holds:*

$$|R(x) - R(y)| \le n(x - y). \tag{2.10}$$

Proof. Using the triangle inequality, we obtain

$$|R(x) - R(y)| = |\max_{z\in D} n(x-z) - \max_{y\in D} n(x-y)| \le \max_{z\in D} |n(x-z) - n(y-x)| \le n(x-y).$$

The global Lipschitz property of the function $R(x)$ follows from (2.10) in view of the equivalence of norms in $\mathbb{R}^p$.

On the basis of subdifferential calculus for convex functions[1,2] we can easily establish the following result.

Lemma 2.5. *The function $R(x)$ is convex on $\mathbb{R}^p$ and its subdifferential can be expressed by the formula*

$$\underline{\partial R}(x) = \operatorname{co}\{\partial n(x-z) :\, z \in Q^R(x, D)\}. \tag{2.11}$$

The following property of the function $R(x)$ is very important for our purposes.

Lemma 2.6. *If x_1 and x_2 are points such that*

$$R(x_2) = R(x_1) + n(x_1 - x_2), \tag{2.12}$$

then for all $\alpha \in [0,1]$ the following equality holds:

$$R(\alpha x_1 + (1-\alpha)x_2) = \alpha R(x_1) + (1-\alpha)R(x_2). \tag{2.13}$$

If on the other hand $n(\cdot)$ is a strictly quasiconvex norm and points x_1 and x_2 satisfy the inequalities

$$R(x_1) \le R(x_2) < R(x_1) + n(x_1 - x_2), \tag{2.14}$$

then the strict inequality

$$R(\alpha x_1 + (1-\alpha)x_2) < \alpha R(x_1) + (1-\alpha)R(x_2), \quad \alpha \in (0,1). \tag{2.15}$$

holds.

PROOF. (a) Assume that (2.12) holds, but there exists $\alpha_0 \in (0,1)$ such that

$$R(\alpha_0 x_1 + (1-\alpha_0)x_2) < \alpha_0 R(x_1) + (1-\alpha_0)R(x_2) = R(x_2) - \alpha_0 n(x_1 - x_2).$$

Then, using Lemma 2.4 we deduce the contradicting inequality

$$R(x_2) - R(\alpha_0 x_1 + (1-\alpha_0)x_2) \le \alpha_0 n(x_1 - x_2),$$

which proves the first part of the statement.

(b) Now let $n(\cdot)$ be a strictly quasiconvex norm and assume that (2.14) holds. If we additionally assume that there exists at least one $\alpha_0 \in (0,1)$ failing (2.15), then it easily follows from the convexity of $R(x)$ that (2.13) holds for all $\alpha \in [0,1]$, that is,

$$R(x_1 + (1-\alpha)(x_2 - x_1)) - R(x_1) = (1-\alpha)(R(x_2) - R(x_1)). \tag{2.16}$$

The function $R(x)$ is a finite convex function on $\mathbb{R}^p$, therefore dividing both parts of equality (2.16) by $\beta = (1-\alpha)$ and passing to the limit as $\beta \downarrow 0$ we obtain

$$R'(x_1, x_2 - x_1) \equiv \lim_{\beta \downarrow 0} \beta^{-1}[R(x_1 + \beta(x_2 - x_1)) - R(x_1)] = R(x_2) - R(x_1). \tag{2.17}$$

The upper continuity of the set-valued map $\partial n(\cdot) : \mathbb{R}^p \to 2^{\mathbb{R}^p}$, the boundedness of $\partial n(x)$ (see (2.1)) and the closedness of the set $Q^R(x, D)$ demonstrate that $\{\partial n(x - z)/\, z \in Q^R(x, D)\}$ is a compact set. Now, applying Lemma 2.5 we see that there exists $z_0 \in Q^R(x, D)$ such that

$$R'(x_1, x_2 - x_1) = \max_{v \in \underline{\partial R}(x_1)} \langle v, x_2 - x_1 \rangle = \max_{v \in \{\partial n(x_1 - z)/z \in Q^R(x_1, D)\}} \langle v, x_2 - x_1 \rangle =$$

$$= \max_{v \in \partial n(x_1 - z_0)} \langle v, x_2 - x_1 \rangle = n'(x_1 - z_0, x_2 - x_1). \tag{2.18}$$

Since $z_0 \in D$ and $R(x_1) = n(x_1 - z_0)$, it follows by right-hand inequality in (2.14) that

$$n(x_2 - z_0) \le R(x_2) < n(x_1 - z_0) + n(x_1 - x_2).$$

Next, using in turn Lemmae 2.2 and 2.1; we obtain

$$n(\alpha x_1 + (1-\alpha)x_2 - z_0) < \alpha n(x_1 - z_0) + (1-\alpha)n(x_2 - z_0), \quad \alpha \in (0,1).$$

Thus, the convex function $f(x) = n(x - z_0)$ is not an affine function on $[x_1, x_2]$ and therefore

$$n(x_2 - z_0) = f(x_2) > f(x_1) + f'(x_1, x_2 - x_1) = n(x_1 - z_0) + n'(x_1 - z_0, x_2 - x_1).$$

Hence, using (2.18) and (2.17) we obtain

$$R(x_2) \geq n(x_2 - z_0) > R(x_1) + n'(x_1 - z_0, x_2 - x_1) =$$

$$= R(x_1) + R'(x_1, x_2 - x_1) = R(x_2),$$

which is a contradiction. □

3.Let us examine now some properties of the distance function $\rho_A(\cdot)$.

Similarly as in the case of Euclidean norm[1,4] we obtain the following statement

LEMMA 2.7.*For the arbitrary set $A \in \mathbb{R}^p$ the function $\rho_A(x)$ is globally Lipschitz on $\mathbb{R}^p$, and for all $x, y \in \mathbb{R}^p$ the following inequality holds:*

$$|\rho_A(x) - \rho_A(y)| \leq n(x - y), \tag{2.19}$$

Under the condition that $D \subset \mathbb{R}^p$ is an arbitrary convex closed set the following two facts are proved[20].

LEMMA 2.8. *The function $\rho_D(x)$ is convex on $\mathbb{R}^p$ and its subdifferential can be expressed by the formula*

$$\underline{\partial \rho}_D(x) = \partial n(x - z) \cap -K^+(z, D), \quad \text{for all } z \in Q^\rho(x, D). \tag{2.20}$$

LEMMA 2.9. *The function $\rho_\Omega(x)$ is concave on D, and its superdifferential can be expressed by the formula*

$$\overline{\partial \rho_\Omega}(x) = \operatorname{co}\{\partial n(x - z) \cap K^+(z, D) : \, z \in Q^\rho(x, \Omega)\}, \quad x \in \operatorname{int} D. \tag{2.21}$$

At the boundary points of D the function $\rho_\Omega(x)$ is differentiable in each direction $g \in \mathbb{R}^p$ and

$$\rho'_\Omega(x, g) \equiv \lim_{\alpha \downarrow 0} \alpha^{-1}[\rho_\Omega(x + \alpha g) - \rho_\Omega(x)] = \max\{0, \min_{\substack{w \in K^+(x,D), \\ n^*(w)=1}} \langle w, g\rangle\}. \tag{2.22}$$

REMARK 2.1. It is easy to show[2] that $K^+(z, D) \subset K(\partial n(x - z))$ for all $z \in Q^\rho(x, D)$ if $x \in D$. Hence we also have the formula

$$\overline{\partial \rho_\Omega}(x) = \operatorname{co}\{w \in K^+(z, D) : \, z \in Q^\rho(x, \Omega), n^*(w) = 1\}, \quad x \in \operatorname{int} D. \tag{2.23}$$

Later we shall use both formulae for $\overline{\partial \rho_\Omega}(x)$.

LEMMA 2.10. *If D is a strictly convex set and points $x_1, x_2 \in D$ satisfy the inequalities*

$$\rho_\Omega(x_1) \leq \rho_\Omega(x_2) < \rho_\Omega(x_1) + n(x_1 - x_2), \tag{2.24}$$

then the strict inequality

$$\rho_\Omega(\alpha x_1 + (1 - \alpha)x_2) > \alpha \rho_\Omega(x_1) + (1 - \alpha)\rho_\Omega(x_2), \quad \alpha \in (0, 1). \tag{2.25}$$

[20]S.I. Dudov, "Subdifferentiability and superdifferentiability of distance functions", Math. Zametki 61:4 (1997), 530-542; English transl. in Math. Notes 61:4 (1997).

holds.

PROOF. The case $x_i \in \partial D$, $i = 1, 2$ is obvious. Let $x_1 \in intD$ or $x_2 \in intD$, then $\alpha\rho_\Omega(x_1) + (1-\alpha)\rho_\Omega(x_2) > 0$, $\alpha \in (0,1)$. For arbitrary point z satisfying

$$n(x_1 + (1-\alpha)x_2 - z) \le \alpha\rho_\Omega(x_1) + (1-\alpha)\rho_\Omega(x_2)), \tag{2.26}$$

that is, $z \in Bn(\alpha x_1 + (1-\alpha)x_2, \alpha\rho_\Omega(x_1) + (1-\alpha)\rho_\Omega(x_2))$, consider points

$$z_i = \frac{x_i + [z - (\alpha x_1 + (1-\alpha)x_2)]\rho_\Omega(x_i)}{\alpha\rho_\Omega(x_1) + (1-\alpha)\rho_\Omega(x_2)}, \quad i = 1, 2. \tag{2.27}$$

Using (2.26) it is easy to verify

$$z_i \in Bn(x_i, \rho_\Omega(x_i)), \quad i = 1, 2, \tag{2.28}$$

and also $z = \alpha z_1 + (1-\alpha)z_2$. Let us show that

$$z_1 \ne z_2. \tag{2.29}$$

Indeed, assuming the contrary, from (2.27) we can deduce

$$x_2 - x_1 = \frac{[z - (\alpha x_1 + (1-\alpha)x_2)](\rho_\Omega(x_2) - \rho_\Omega(x_1))}{\alpha\rho_\Omega(x_1) + (1-\alpha)\rho_\Omega(x_2)}.$$

Hence applying (2.26) we have $n(x_2 - x_1) \le \rho_\Omega(x_2) - \rho_\Omega(x_1)$. It contradicts (2.24).

Since $Bn(x_i, \rho_\Omega(x_i)) \subset D$, $i = 1, 2$ and D is strictly convex set, then (2.28)-(2.29) imply $z = \alpha z_1 + (1-\alpha)z_2 \in intD$. In view of the arbitrary choice of $z \in Bn(\alpha x_1 + (1-\alpha)x_2, \alpha\rho_\Omega(x_1) + (1-\alpha)\rho_\Omega(x_2))$ we proved

$$Bn(\alpha x_1 + (1-\alpha)x_2, \alpha\rho_\Omega(x_1) + (1-\alpha)\rho_\Omega(x_2)) \subset intD.$$

It implies inequality (2.25). $\square$

4. In this subsection let us consider function $P(x) = \rho_D(x) - \rho_\Omega(x)$, assuming that the set D is convex and closed.

The following statement is completely obvious.

LEMMA 2.11. *Let $f_1(t)$ and $f_2(t)$ be finite convex functions on the closed intervals $[a, b]$ and $[b, c]$, respectively, where $a < b < c$, and let $f_1(b) = f_2(b)$. If there exists $\alpha \in (0,1)$ such that the function*

$$f(t) = \begin{cases} f_1(t), & t \in [a, b], \\ f_2(t), & t \in [b, c] \end{cases}$$

satisfies the inequality

$$f(\alpha a + (1-\alpha)c) > \alpha f(a) + (1-\alpha)f(c),$$

and if

$$b = \beta a + (1-\beta)c, \quad \beta \in (0,1),$$

then

$$f(b) > \beta f(a) + (1-\beta) f(c).$$

The main properties of function $P(x)$ are mapped by following statement.

LEMMA 2.12. *The function $P(x)$ is finite and convex on $\mathbb{R}^p$, and its subdifferential can be expressed as follows:*

$$\underline{\partial P}(x) = \begin{cases} \partial n(x-z) \cap -K^+(z,D), & \forall z \in Q^\rho(x,D), \quad x \notin D, \\ \operatorname{co}\{v \in -K^+(z,D) : n^*(v) = 1, & z \in Q^\rho(x,\Omega)\}, \quad x \in D. \end{cases} \tag{2.30}$$

PROOF. Assume that $P(x)$ is not convex on $\mathbb{R}^p$, that is, there exist points x_1, x_2 and quantity $\alpha_0 \in (0,1)$ such that

$$P(\alpha_0 x_1 + (1-\alpha_0)x_2) > \alpha_0 P(x_1) + (1-\alpha_0) P(x_2). \tag{2.31}$$

We see from Lemmae 2.8 and 2.9 and also from the definition of $P(x)$ that only cases when $x_1 \in intD$ and $x_2 \notin D$ or $x_1, x_2 \notin D$, but $[x_1, x_2] \cap intD \neq \emptyset$ are non-trivial.

(a) Let $x_1 \in \operatorname{int} D, x_2 \notin D$. Let x_0 be the boundary point of D lying on the line segment $[x_1, x_2]$. The function $P(x)$ is convex on $[x_1, x_0]$ and on $[x_0, x_2]$ by Lemmae 2.9 and 2.8 respectively. Hence, by Lemma 2.11, we see from inequality (2.31) for the point $x_0 = \beta x_1 + (1-\beta)x_2$, where $\beta \in (0,1)$, that:

$$P(x_0) > \beta P(x_1) + (1-\beta) P(x_2). \tag{2.32}$$

We claim that the following inequality for $g = \dfrac{x_2 - x_1}{\|x_2 - x_1\|}$ is a consequence of (2.32):

$$P'(x_0, g) < -P'(x_0, -g). \tag{2.33}$$

Indeed, from (2.32) we immediately obtain

$$(1-\beta)(P(x_2) - P(x_0)) < \beta(P(x_0) - P(x_1)). \tag{2.34}$$

Since $\dfrac{\beta}{1-\beta} = \dfrac{\|x_2 - x_0\|}{\|x_1 - x_0\|}$, from (2.34) it follows that

$$\frac{P(x_2) - P(x_0)}{\|x_2 - x_0\|} < \frac{P(x_0) - P(x_1)}{\|x_1 - x_0\|}. \tag{2.35}$$

The function $P(x)$ is convex on $[x_0, x_2]$ and $[x_1, x_0]$, therefore

$$P'(x_0, g) \le \frac{P(x_2) - P(x_0)}{\|x_2 - x_0\|}, \quad P'(x_0, -g) \le \frac{P(x_1) - P(x_0)}{\|x_1 - x_0\|}. \tag{2.36}$$

A combination of (2.35) and (2.36) yields (2.33).

We now claim that, in fact,

$$P'(x_0, g) \ge -P'(x_0, -g)). \tag{2.37}$$

Actually, since x_0 is a boundary point of a convex set D, it is a well-known result of convex analysis[2,18] that there exists a support hyperplane π containing x_0 such that D lies in one of the two half-spaces into which it partitions the space $\mathbb{R}^p$. It now follows from the definition of $P(x)$, the properties of the support hyperplane π, and our choice of x_1 and x_2 that for sufficiently small $\alpha \geq 0$ we have

$$\rho_\pi(x_0+\alpha g) = \rho_\pi(x_0-\alpha g), \quad P(x_0+\alpha g) = \rho_D(x_0+\alpha g) \geq \rho_\pi(x_0+\alpha g),$$
$$P(x_0-\alpha g) = -\rho_\Omega(x_0-\alpha g) \geq -\rho_\pi(x_0-\alpha g), \quad \rho_\pi(x_0) = P(x_0) = 0.$$

Hence it easily follows that

$$\begin{gathered} P'(x_0,g) \geq \rho'_\pi(x_0,g), \quad P'(x_0,-g) \geq -\rho'_\pi(x_0,-g), \\ \rho'_\pi(x_0,g) = \rho'_\pi(x_0,-g). \end{gathered} \tag{2.38}$$

From (2.38) we can deduce (2.37). The contradiction between (2.37) and (2.33) means that our assumption (2.31) fails in the case under consideration.

(b) Assume now that $x_1, x_2 \notin D$, but $[x_1,x_2] \cap \operatorname{int} D \neq \emptyset$. The function $P(x)$ is continuous. Hence it follows from its definition that there exists $x_0 \in [x_1,x_2] \cap \operatorname{int} D$ such that

$$P(x_0) = \min_{x\in[x_1,x_2]} P(x) < 0 < \min\{P(x_1), P(x_2)\}. \tag{2.39}$$

As shown in (a), the function $P(x)$ is convex on the closed intervals $[x_1,x_0]$ and $[x_0,x_2]$. By assumption (2.31), using Lemma 2.11 for the point $x_0 = \beta x_1 + (1--\beta)x_2$ where $\beta \in (0,1)$, we obtain the inequality $P(x_0) > \beta P(x_1) + (1-\beta)P(x_2)$, which contradicts (2.39).

(c) It remains to prove formula (2.30). It is an immediate consequence of Lemmae 2.8 and 2.9 for $x \notin D$ or $x \in \operatorname{int} D$. Hence it remains to prove it in the case when x is a boundary point of D. Using Lemmae 2.8 and 2.9, and formula (2.1) we obtain

$$\begin{aligned} P'(x,g) &= \rho'_D(x,g) - \rho'_\Omega(x,g) = \\ &= \max_{v\in\partial n(0)\cap -K^+(x,D)} \langle v,g\rangle - \max\{0, \min_{\substack{w\in K^+(x,D),\\ n^*(w)=1}} \langle w,g\rangle\} = \\ &= \max\{0, \max_{\substack{v\in -K^+(x,D),\\ n^*(v)=1}} \langle v,g\rangle\} + \min\{0, \max_{\substack{v\in -K^+(x,D),\\ n^*(v)=1}} \langle v,g\rangle\} = \\ &= \max_{\substack{v\in -K^+(x,D),\\ n^*(v)=1}} \langle v,g\rangle = \max_{\operatorname{co}\{v\in -K^+(x,D): n^*(v)=1\}} \langle v,g\rangle, \quad \forall\, g \in \mathbb{R}^p. \end{aligned} \tag{2.40}$$

On the other hand, for the finite convex function $P(x)$ we have the following expression for its directional derivative in terms of the subdifferential $\underline{\partial P}(x)$, which is a convex compact set:

$$P'(x,g) = \max_{v\in\underline{\partial} P(x)} \langle v,g\rangle, \quad \forall g \in \mathbb{R}^p. \tag{2.41}$$

From (2.40) and (2.41) we see that $\underline{\partial P}(x) = \mathrm{co}\,\{v \in -K^+(x,D) : n^*(v) = 1\}$, which corresponds to formula (2.30) in the case $x \in \partial D$. $\square$

3 Reduction to a convex problem

In this section we prove that the problems (1.1) and (1.3) are equivalent.

The deviation of the convex compact set A from the convex set B in an arbitrary norm $n(\cdot)$ can be expressed by formula[1]

$$\rho(A,B) = \sup_{n^*(\psi)\le 1} \{W(\psi,A) - W(\psi,B)\}.$$

Hence the corresponding Hausdorff distance between the convex compact sets A and B we can write in the form

$$h(A,B) = \sup_{n^*(\psi)\le 1} |W(\psi,A) - W(\psi,B)|. \tag{3.1}$$

As above we suppose that $D \subset \mathbb{R}^p$ is the convex compact set and $\Omega = \overline{\mathbb{R}^p \setminus D}$.

LEMMA 3.1. *The following equality holds*

$$\min_{n^*(\psi)=1} \max_{v\in D}\langle v-x,\psi\rangle = \begin{cases} -\rho_D(x), & \text{if} \quad x \notin D, \\ \\ \rho_\Omega(x), & \text{if} \quad x \in D. \end{cases} \tag{3.2}$$

PROOF. (a) Let $x \notin D$. Consider the positively homogeneous function $\varphi(\psi) = \max_{v\in D}\langle v-x,\psi\rangle$. Since $x \notin D$ it is easy to see that $\varphi(\psi)$ may receive the negative value. Therefore the following equality

$$\min_{n*(\psi)=1} \max_{v\in D}\langle v-x,\psi\rangle = \min_{n*(\psi)\le 1} \max_{v\in D}\langle v-x,\psi\rangle \tag{3.3}$$

is valid.

Using well-known minimax theorem[3] we obtain

$$\min_{n*(\psi)\le 1} \max_{v\in D}\langle v-x,\psi\rangle = \max_{v\in D} \min_{n*(\psi)\le 1}\langle v-x,\psi\rangle = -\min_{v\in D} \max_{n*(\psi)\le 1}\langle x-v,\psi\rangle =$$

$$= -\min_{v\in D} n(x-v) = -\rho_D(x). \tag{3.4}$$

In the considered case (3.2) follows from (3.3)-(3.4).

(b) Let now $x \in D$. Since $Bn(x,\rho_\Omega(x)) \subset D$ we have

$$\min_{n*(\psi)=1} \max_{v\in D}\langle v-x,\psi\rangle \ge \min_{n*(\psi)=1} \max_{v\in Bn(x,\rho_\Omega(x))}\langle v-x,\psi\rangle =$$

$$\min_{n*(\psi)=1} \max_{v\in x+\rho_\Omega(x)Bn(0_p,1)}\langle v-x,\psi\rangle = \rho_\Omega(x)\cdot \min_{n*(\psi)=1} \max_{n(v)\le 1}\langle v,\psi\rangle =$$

$$= \rho_\Omega(x) \cdot \min_{n*(\psi)=1} n^*(\psi) = \rho_\Omega(x). \tag{3.5}$$

On the other hand take any point $v^* \in Q^\rho(x, \Omega)$. It is the boundary point of the convex set D. Therefore by the supporting hyperplane theorem there exists element $\psi^* \neq 0_p$ such that

$$\langle v, \psi^* \rangle \leq \langle v^*, \psi^* \rangle \quad \text{for all } v \in D.$$

Hence we obtain

$$\max_{v \in D} \langle v - x, \psi^* \rangle \leq \langle v^* - x, \psi^* \rangle. \tag{3.6}$$

Without loss of generality we assume that $n^*(\psi) = 1$. It is easy to see, that

$$\min_{n^*(\psi)=1} \max_{v \in D} \langle v - x, \psi \rangle \leq \max_{v \in D} \langle v - x, \psi^* \rangle, \tag{3.7}$$

$$\langle v^* - x, \psi^* \rangle \leq \max_{n^*(\psi) \leq 1} \langle v^* - x, \psi \rangle = n(x - v^*) = \rho_\Omega(x). \tag{3.8}$$

From (3.6)–(3.8) we have inequality

$$\min_{n^*(\psi)=1} \max_{v \in D} \langle v - x, \psi \rangle \leq \rho_\Omega(x),$$

which together with (3.5) gives us (3.2) for the case in question. □

Now we will use (3.1) and Lemma 3.1 to obtain one interesting formula.

LEMMA 3.2. *For any $x \in \mathbb{R}^p$ and $r \geq 0$ the following formula holds*

$$h(D, Bn(x, r)) = \max\{R(x) - r, r + P(x)\}. \tag{3.9}$$

PROOF. In accordance with (3.1) we obtain

$$h(D, Bn(x, r)) = \sup_{n^*(\psi) \leq 1} |W(\psi, D) - W(\psi, Bn(x, r))| =$$

$$= \max_{n^*(\psi) \leq 1} |\max_{v \in D} \langle v, \psi \rangle - \max_{v \in x + r\, Bn(0_p, 1)} \langle v, \psi \rangle| =$$

$$\max_{n^*(\psi) \leq 1} |\max_{v \in D} \langle v - x, \psi \rangle - r \max_{n(v) \leq 1} \langle v, \psi \rangle| = \max_{n^*(\psi) \leq 1} |\max_{v \in D} \langle v - x, \psi \rangle - r\, n^*(\psi)| =$$

$$= \max\{\max_{n^*(\psi) \leq 1} f(\psi), - \min_{n^*(\psi) \leq 1} f(\psi)\}, \tag{3.10}$$

where $f(\psi) = \max_{v \in D} \langle v - x, \psi \rangle - r\, n^*(\psi)$.

Since $f(\psi)$ is the positively homogeneous function, then

$$\begin{gathered} \max_{n^*(\psi) \leq 1} f(\psi) = \max\{0, \max_{n^*(\psi)=1} f(\psi)\}, \\ \min_{n^*(\psi) \leq 1} f(\psi) = \min\{0, \min_{n^*(\psi)=1} f(\psi)\}. \end{gathered} \tag{3.11}$$

It is easy to see that

$$\max_{n^*(\psi)=1} f(\psi) = \max_{v\in D} \max_{n^*(\psi)=1} \langle v-x, \psi\rangle - r =$$

$$\max_{v\in D} n(v-x) - r = R(x) - r, \tag{3.12}$$

and by the Lemma 3.1,

$$\min_{n^*(\psi)=1} f(\psi) = \begin{cases} -(r+\rho_D(x)) & for \quad x \notin D, \\ \\ -(r-\rho_\Omega(x)) & for \quad x \in D. \end{cases} \tag{3.13}$$

Substituting (3.12) and (3.13) to (3.11), we have

$$\begin{aligned} &\max_{n^*(\psi)\le 1} f(\psi) = \max\{0, R(x)-r\}, \\ \\ &\min_{n^*(\psi)\le 1} f(\psi) = \min\{0, -(r+P(x))\}. \end{aligned} \tag{3.14}$$

Now the formula (3.9) follows from (3.10) and (3.14). □

It is obvious, that $R(x) \ge P(x)$, therefore we obtain

$$\min_{r\ge 0} \max\{R(x)-r, r+P(x)\} = \max\{R(x)-r_0, r_0+P(x)\} =$$

$$= \frac{R(x)+P(x)}{2}, \tag{3.15}$$

where $r_0 = \dfrac{R(x)-P(x)}{2}$.

Directly from formula (3.15) and Lemma 3.2 the following statement, which plays important role, follows.

Theorem 3.1. *The problem (1.1) is equivalent to the problem (1.3). In addition, if the pair (x_0, r_0) is a solution of (1.1), then the point x_0 is a solution of (1.3) and $r_0 = (R(x_0) - P(x_0))/2$. Conversely, if x_0 is a solution of (1.3), then the pair (x_0, r_0), where $r_0 = (R(x_0) - P(x_0))/2$, is a solution of (1.1). Moreover,*

$$\min_{x\in\mathbb{R}^p,\, r\ge 0} h(D, Bn(x_0, r)) = \frac{R(x_0)+P(x_0)}{2}.$$

4 Necessary and sufficient condition of solution

Obviously the problem (1.3) is equivalent to the problem

$$\Phi(x) \equiv R(x) + P(x) \to \min_{x\in D(z)},$$

where $D(z) = \{x \in \mathbb{R}^p : \Phi(x) \le \Phi(z)\}$ for any $z \in \mathbb{R}^p$. It is easy to see that the set $D(z)$ is bounded and closed and function $\Phi(x)$ is continuous. From this fact we have the existence of solution of problem (1.3), and, therefore, of problem (1.1). The formulae of subdifferentials of $R(x)$ and $P(x)$ make it possible to obtain necessary and sufficient condition for solution of the problem (1.1).

THEOREM 4.1. *A pair (x_0, r_0) solves the problem (1.1) if and only if*

$$0_p \in \underline{\partial \Phi}(x_0) \equiv \underline{\partial R}(x_0) + \underline{\partial P}(x_0),$$

where $\underline{\partial R}(x_0)$ and $\underline{\partial P}(x_0)$ are defined by formulae (2.11) and (2.30), respectively, and $r_0 = (R(x_0) - P(x_0))/\,2$.

PROOF. As it follows from Lemmae 2.5 and 2.12, the function $\Phi(x)$ is finite and convex on $\mathbb{R}^p$. Its subdifferential $\underline{\partial \Phi}(x)$ is the algebraic sum of $\underline{\partial R}(x)$ and $\underline{\partial P}(x)$ by the Moreau-Rockafellar theorem[18]. From convex analysis[1,2,18], we know that a necessary and sufficient condition of the solution of (1.3) is the inclusion $0_p \in \underline{\partial \Phi}(x_0)$. Now it remains to use theorem 3.1. □

REMARK 4.1. It follows from Theorem 3.1 that

$$X(D) = \{y \in \mathbb{R}^p : \, \Phi(y) = \min_{x \in \mathbb{R}^p} \Phi(x)\}$$

is the set of centers of the best approximation balls for the convex compact set D and, in accordance with Theorem 4.1,

$$x_0 \in X(D) \Leftrightarrow 0_p \in \underline{\partial R}(x_0) + \underline{\partial P}(x_0). \tag{4.1}$$

COROLLARY 4.1. *Let D be a convex compact set symmetric with respect to a point x_0. Then $x_0 \in X(D)$.*

PROOF. It is easy to see that $x_0 \in D$ and the sets $Q^R(x_0, D)$ and $Q^\rho(x_0, \Omega)$ possess central symmetry relative to x_0. One sees from formulae (2.1), (2.11), and (2.30) that this yields the inclusions

$$0_p \in \underline{\partial R}(x_0), \quad 0_p \in \underline{\partial P}(x_0).$$

Hence (4.1) also holds. □

REMARK 4.2. Thus if convex compact set D possesses central symmetry and the solution of the problem (1.1) is unique, then $X(D) = \{x_0\}$, where point x_0 is the center of symmetry. Example 4.1 below shows that even if D possesses central symmetry, the set $X(D)$ is not necessarily a singleton.

EXAMPLE 4.1. Let $p = 3$, $x = (x^{(1)}, x^{(2)}, x^{(3)}) \in \mathbb{R}^3$,

$$n(x) = \max\{|x^{(1)}|, |x^{(2)}|, |x^{(3)}|\},$$

$$D = \mathrm{co}\,\{(0,1,1/2), (0,1,-1/2), (0,-1,1/2), (0,-1,-1/2)\}.$$

In that case $\mathrm{int}\, D = \emptyset$ and D is a set symmetric with respect to $x_0 = (0,0,0)$ and $r_0 = (R(x_0) - \rho_\Omega(x_0))/2 = R(x_0)/2$=1/2. Hence it follows by Theorem 3.1 and

Corollary 4.1 that $Bn(0_3, 1/2)$ is the best approximation ball for D and $h_0(D) = h(D, Bn(0_3, 1/2)) = 1/2$. However, it is easy to see that for $x \in \mathrm{co}\,\{(0,0,1/2),(0,0,-1/2)\}$ we have $h(D, Bn(x, 1/2)) = 1/2$, that is, the balls $Bn(x, 1/2)$ are also best approximation balls. Hence

$$\mathrm{co}\,\{(0,0,1/2),(0,0,-1/2)\} \subset X(D).$$

Corollary 4.2. *If $D = [x_1, x_2]$ – is a line segment connecting x_1 and x_2; then the best approximation ball in the problem (1.1) is unique. Moreover, it has its center at the point $x_0 = (x_1 + x_2)/2$ and its radius is $r_0 = n(x_1 - x_2)/4$.*

Proof. Clearly, x_0 is a minimum point of the function $R(x)$ in $\mathbb{R}^p$ and it is the only such point on the interval $[x_1, x_2]$. Since $P(x) = 0$ for $x \in [x_1, x_2]$, x_0 is also the unique minimum point of $\Phi(x)$ on this closed interval. Moreover, it has the property $\Phi(x_0) = \min\limits_{x \in \mathbb{R}^p} \Phi(x)$ by Corollary 4.1. It remains to observe that for $x \notin D$ we have $P(x) = \rho_D(x) > 0$, so that

$$\Phi(x) = R(x) + P(x) > R(x_0) = \Phi(x_0) = \min_{x \in \mathbb{R}^p} \Phi(x).$$

Using Theorem 1.1 we can calculate the radius of the best approximation ball:

$$r_0 = \frac{R(x_0) - P(x_0)}{2} = \frac{R(x_0)}{2} = \frac{n(x_1 - x_2)}{4}. \ \square$$

The following fact is important.

Theorem 4.2. *The following relation*

$$X(D) \cap D \neq \emptyset, \tag{4.2}$$

is valid. Moreover if $x_0 \in X(D)$ but $x_0 \notin D$, then

$$co\{x_0, Q^\rho(x_0, D)\} \subset X(D). \tag{4.4}$$

Proof. Assume that $x_0 \in X(D)$ but $x_0 \notin D$. We consider an arbitrary point $z \in Q^\rho(x_0, D)$. Then $P(x_0) = \rho_D(x_0) = n(x_0 - z)$, $P(z) = 0$, and therefore

$$R(x_0) + n(x_0 - z) = R(x_0) + P(x_0) = \Phi(x_0) \leq \Phi(z) = R(z). \tag{4.4}$$

On the other hand, by Lemma 2.4 we obtain

$$R(z) - R(x_0) \leq n(x_0 - z). \tag{4.5}$$

From (4.4) and (4.5) we see that $R(z) - R(x_0) = n(x_0 - z)$, and therefore we also have $\Phi(z) = \Phi(x_0)$, that is, $z \in X(D)$. Since $\Phi(x)$ is a convex function, the entire segment $[x_0, z] \subset X(D)$ lies in $X(D)$. We have thus proved the inclusion (4.3). Relation (4.2) appears, actually, as its consequence. $\square$

REMARK 4.3. Let $Pr_B A = \{b \in B : n(a-b) = \min_{y \in B} n(a-y) \text{ for some } a \in A\}$ be the projection of a set A onto a set B. It is easy to see that the assertion of Theorem 4.2 is equivalent to the relation

$$\text{co}\,\{X(D), Pr_D X(D)\} = X(D).$$

Relation (4.2) means that at least one best approximation ball has its center in D.

We now discuss the case when the center of a best approximation ball lies at the boundary of D.

COROLLARY 4.3. *If $x_0 \in X(D) \cap \partial D$, then*

$$\underline{\partial R}(x_0) \cap K^+(x_0, D) \neq \emptyset. \tag{4.6}$$

PROOF. Since $x_0 \in X(D)$, the inclusion (4.1) follows from Theorems 3.1 and 4.1. For $x_0 \in \partial D$ we have $Q^\rho(x_0, \Omega) = \{x_0\}$, therefore using formula (2.30) we can write the inclusion (4.1) as follows:

$$0_p \in \underline{\partial R}(x_0) + \text{co}\,\{v \in -K^+(x_0, D) : \, n^*(v) = 1\}.$$

Hence we immediately obtain (4.6). □

REMARK 4.4. From convex analysis[1,2] we know that (4.6) is equivalent to

$$R(x_0) = \min_{x \in D} R(x).$$

COROLLARY 4.4. *Let $int\, D = \emptyset$ and $x_0 \in D$. Then $x_0 \in X(D)$ if and only if (4.6) holds.*

PROOF. The necessity follows from Corollary 4.3. Assume that (4.6) holds. Since $int\, D = \emptyset$, it follows that $P(x) = 0$ for all $x \in D$. Hence, using Theorem 4.2 and taking into account Remark 4.4 we obtain

$$\Phi(x_0) = R(x_0) = \min_{x \in D} R(x) = \min_{x \in D} \Phi(x) = \min_{x \in \mathbb{R}^p} \Phi(x).$$

This shows that $x_0 \in X(D)$. □

REMARK 4.5. Simple examples show that if $int\, D \neq \emptyset$, then relation (4.6) is not sufficient for validity of $x_0 \in X(D) \cap \partial D$.

COROLLARY 4.5. *Let $int\, D \neq \emptyset$, and let $n(\cdot) = \|\cdot\|$ be the Euclidean norm. Then a point x_0 is the center of a best approximation ball if and only if $x_0 \in \text{int}\, D$ and*

$$0_p \in \text{co}\{\frac{x_0 - y}{\|x_0 - y\|} : \, y \in Q^R(x_0, D)\} - \text{co}\{\frac{x_0 - z}{\|x_0 - z\|} : \, z \in Q^\rho(x_0, \Omega)\}. \tag{4.7}$$

PROOF. For the case of Euclidean norm the solution of problem (1.1) is unique and also[8]

$$X(D) = \{x_0\} \in intD.$$

The smoothness of Euclidean norm involves $\partial n(x) = \{n'(x)\}$ for all $x \neq 0_p$, where gradient $n'(x) = x\|x\|^{-1}$. Therefore taking into account $P(x_0) = -\rho_\Omega(x_0)$ and formulae (2.11)-(2.21) we have

$$\begin{aligned} \underline{\partial R}(x_0) &= co\{\frac{x_0 - y}{\|x_0 - y\|} : y \in Q^R(x_0, D)\}, \\ \underline{\partial P}(x_0) &= -co\{\frac{x_0 - z}{\|x_0 - z\|} : z \in Q^\rho(x_0, \Omega)\}. \end{aligned} \tag{4.8}$$

Now it is easy to see that inclusion (4.1) is equivalent to (4.7) in this case. □

Remark 4.6. As it was noted in section 1 problem (1.1) is equivalent to the problem (1.2) for Euclidean norm. It is not difficult to verify that the necessary and sufficient condition, obtained by I.Barany[16], is equivalent to (4.7).

Remark 4.8. Let $p = 2$, let D be a triangle, and let $n(\cdot)$ be the Euclidean norm. Using relation (4.7) it is easy to prove that the point x_0, that is the center of the best approximation ball (disc), is the point of intersection of the perpendicular bisector of the largest side of the triangle and the bisector of the smaller angle adjoining this side. The radius of this disc is $r_0 = (R(x_0) + \rho_\Omega(x_0))/2$.

Corollary 4.6. *If $intD = \emptyset$ and $n(\cdot) = \|\cdot\|$ is the Euclidean norm, then*

$$x_0 \in X(D) \Leftrightarrow x_0 \in coQ^R(x_0, D). \tag{4.9}$$

Proof. It is not difficult to prove that for the Euclidean norm

$$\min_{x \in D} R(x) = \min_{x \in \mathbb{R}^p} R(x).$$

Hence in accordance with Corollary 4.4 and Remark 4.4 the point $x_0 \in X(D)$ if and only if

$$R(x_0) = \min_{x \in \mathbb{R}^p} R(x). \tag{4.10}$$

As we know[1,2,18], (4.10) is equivalent to inclusion

$$0_p \in \underline{\partial R}(x_0).$$

Now using formula (4.8) it is easy to obtain (4.9). □

5 Conditions providing an estimated set contains all centers of the best approximation balls

Relation (4.2) means that at least one best approximation ball has its center in D. In this section we will deduce some conditions ensuring the inclusion $X(D) \subset D$.

Let $Sn(x, r) = \{y \in \mathbb{R}^p : n(x - y) = r\}$ be the sphere with respect to the norm $n(\cdot)$ with center at x and of radius r.

Lemma 4.1. *If $n(\cdot)$ is a strictly quasiconvex norm, then the spheres $Sn(x_1, r_1)$ and $Sn(x_2, r_1 + n(x_1 - x_2))$ have the unique common point $y_0 = x_1 + r_1(n(x_1 - x_2))^{-1}(x_1 - x_2)$.*

Proof. If $y \in Sn(x_1, r_1) \cap Sn(x_2, r_1 + n(x_1 - x_2))$, then

$$n(x_1 - y) = r_1, \quad n(x_2 - y) = r_1 + n(x_1 - x_2), \tag{5.1}$$

that is, $n(x_2 - y) = n(x_1 - y) + n(x_1 - x_2)$. By Lemma 2.2 this means that $x_2 - y = \lambda(x_1 - y)$, where $\lambda > 1$. Expressing y and substituting this expression into (5.1) we obtain the uniqueness of the common point and the required formula for it. □

Theorem 5.1. *Assume that at least one of the following conditions is fulfilled:*

1) $n(\cdot)$ is a smooth norm,

2) $n(\cdot)$ is a strictly quasiconvex norm,

3) $p = 2$,

4) the set D possess central symmetry.

Then

$$X(D) \subset D. \tag{5.2}$$

Proof. Assume the contrary: there exists at least one point $x_0 \in X(D)$ such that $x_0 \notin D$. Then we see from (4.1) that

$$0_p \in \partial\Phi(x_0) = \underline{\partial R}(x_0) + \underline{\partial \rho_D}(x_0). \tag{5.3}$$

(a) Consider the case when $n(\cdot)$ is a smooth norm. By Lemma 2.8 we obtain

$$\underline{\partial \rho_D}(x_0) = n'(x_0 - z), \quad \forall z \in Q^{\rho}(x_0, D). \tag{5.4}$$

We claim that there exists $z^* \in Q^R(x_0, D)$ such that

$$n'(x_0 - z) = -n'(x_0 - z^*), \quad \forall z \in Q^{\rho}(x_0, D). \tag{5.5}$$

The inclusion (5.3) means, in view of (2.11) and (5.4), that there exists a positive integer m, elements $\{z_i\}_{i=\overline{1,m}} \subset Q^R(x_0, D)$, and quantities $\alpha_i > 0,\ i = \overline{1,m}$, such that $\sum\limits_{i=1}^{m} \alpha_i = 1$ and

$$n'(x_0 - z) = -\sum_{i=1}^{m} \alpha_i n'(x_0 - z_i), \quad z \in Q^{\rho}(x_0, D). \tag{5.6}$$

Since $\partial n(x_0 - z_i) = \{n'(x_0 - z_i)\}$, it follows by (2.1) that

$$n^*(n'(x_0 - z_i)) = 1. \tag{5.7}$$

Hence, taking into account the equality

$$n(x) = \max_{n^*(v) \le 1} \langle v, x \rangle, \tag{5.8}$$

we obtain

$$\langle n'(x_0 - z_i), x_0 - z\rangle \leq n(x_0 - z). \tag{5.9}$$

In addition, in accordance with (2.1) we have

$$\langle n'(x_0 - z), x_0 - z\rangle = n(x_0 - z). \tag{5.10}$$

Thus, we see from (5.6), (5.9), (5.10) that

$$\langle -n'(x_0 - z_i), x_0 - z\rangle = n(x_0 - z), \quad i = \overline{1, m}, \quad z \in Q^\rho(x_0, D). \tag{5.11}$$

Comparing (2.1) with (5.7) and (5.11) we conclude that

$$-n'(x_0 - z_i) \subset \partial n(x_0 - z) = n'(x_0 - z),$$

that is, $-n'(x_0 - z_i) = n'(x_0 - z)$, $i = \overline{1, m}$, $z \in Q^\rho(x_0, D)$. We have thus established (5.5).

We now fix a point $z_0 \in Q^\rho(x_0, D)$. The ball $Bn(x_0, \rho_D(x_0))$ touches the convex set D at the point z_0. Hence there exists a hyperplane through this point that separates $Bn(x_0, \rho_D(x_0))$ from D. In the case under consideration the ball has a smooth boundary and $n'(x_0 - z_0)$ is the normal to it at z_0, therefore this hyperplane has the equation

$$\pi = \{x \in R^p : \langle x, n'(x_0 - z_0)\rangle = \langle z_0, n'(x_0 - z_0)\rangle\}.$$

Moreover,

$$\langle x, n'(x_0 - z_0)\rangle \leq \langle y, n'(x_0 - z_0)\rangle, \quad \forall x \in D, \quad y \in Bn(x_0, \rho_D(x_0)). \tag{5.12}$$

Since $Q^R(x_0, D) \subset D$ and $x_0 \in Bn(x_0, \rho_D(x_0))$, it follows by (5.12) that

$$\langle z - x_0, n'(x_0 - z_0)\rangle \leq 0, \quad \text{for each } z \in Q^R(x_0, D).$$

Now, considering here point $z^* \in Q^R(x_0, D)$ satisfying (5.5) and using (2.1) we arrive at a contradiction:

$$\langle z^* - x_0, -n'(x_0 - z^*)\rangle = n(x_0 - z^*) \leq 0.$$

(b) Let $n(\cdot)$ be a strictly quasiconvex norm: and let $z_0 \in Q^\rho(x_0, D)$ be an arbitrary point. By Theorem 4.2, $[x_0, z_0] \subset X(D)$, therefore $\Phi(x_0) = \Phi(z_0)$. Since $\Phi(x_0) = R(x_0) + \rho_D(x_0) = R(x_0) + n(x_0 - z_0)$ and $\Phi(z_0) = R(z_0)$, it follows that $R(z_0) = R(x_0) + n(x_0 - z_0)$, which means that $Bn(x_0, R(x_0)) \subset$ $\subset Bn(z_0, R(z_0))$, and the boundary spheres $Sn(x_0, R(x_0))$ and $Sn(z_0, R(z_0))$ of these balls are by Lemma 5.1 tangent at the unique point $y_0 = x_0 + R(x_0)(n(x_0 -$ $-z_0))^{-1}(x_0 - z_0)$. By the definition of the function $R(\cdot)$ the set D has common points with $Sn(x_0, R(x_0))$ and $Sn(z_0, R(z_0))$. Hence we conclude that $y_0 \in D$.

The convex compact sets D and $Bn(x_0, \rho_D(x_0))$ have no common interior points. Hence, by the separating hyperplane theorem there exists $g_0 \in \mathbb{R}^p$, $g_0 \neq 0_p$, such that

$$\langle x, g_0\rangle \leq \langle z_0, g_0\rangle \leq \langle y, g_0\rangle, \quad \forall x \in Bn(x_0, \rho_D(x_0)),\ y \in D. \tag{5.13}$$

Since $x_0 \in \operatorname{int} Bn(x_0, \rho_D(x_0))$, it follows that $\langle x_0, g_0\rangle < \langle z_0, g_0\rangle$. Hence, bearing in mind that $y_0 \in D$, we obtain

$$\langle y_0, g_0\rangle = \langle x_0, g_0\rangle + \frac{R(x_0)}{n(x_0 - z_0)}\langle x_0 - z_0, g_0\rangle < \langle z_0, g_0\rangle.$$

This contradicts the right-hand side of (5.13).

(c) Let the dimension of the space be $p = 2$. By (5.3) we conclude that there exists an element $v_0 \in \underline{\partial R}(x_0) \cap -\underline{\partial \rho}_D(x_0)$. The inclusion $v_0 \in -\underline{\partial \rho}_D(x_0)$ means in accordance with (2.20) and (2.1) that for each $z_0 \in Q^\rho(x_0, D)$ we have the relations

$$n^*(v_0) = 1, \tag{5.14}$$

$$v_0 \in K^+(z_0, D), \tag{5.15}$$

$$\langle v_0, z_0 - x_0\rangle = n(x_0 - z_0). \tag{5.16}$$

We now claim that if $p = 2$, then the inclusion $v_0 \in \underline{\partial R}(x_0)$ and relation (5.14) yield the existence of elements $z_1, z_2 \in Q^R(x_0, D)$, appropriate points $v_1 \in \partial n(x_0 - z_1)$, $v_2 \in \partial n(x_0 - z_2)$ and $\alpha \in [0, 1]$ such that

$$v_0 = \alpha v_1 + (1 - \alpha)v_2. \tag{5.17}$$

For it follows from (2.11) by Caratheodory's theorem[2,18] that there exist (not necessarily distinct) elements $\{z_i\}_{i=1,2,3} \subset Q^R(x_0, D)$, appropriate (distinct) elements $v_i \in \partial n(x_0 - z_i), i = 1, 2, 3$, and quantities $\alpha_i \geq 0,\ i = \overline{1, 3},\ \sum\limits_{i=1}^{3} \alpha_i = 1$, such that $v_0 = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3$. Moreover, in view of (2.1) we have $n^*(v_i) = 1$, $i = 1, 2, 3$. If the elements $\{v_i\}_{i=1,2,3}$ do not all lie on the same line and the quantities α_i are all strictly positive, then $v_0 \in \operatorname{int}\operatorname{co}\{v_1, v_2, v_3\}$, and therefore $n^*(v_0) < 1$ which contradicts (5.14). Hence we can assume without loss of generality that $\alpha_3 = 0$, which means that the element v_0 can be represented in the form (5.17).

It follows by (5.15) that

$$\langle v_0, x - z_0\rangle \geq 0, \quad \text{for all } x \in D$$

and therefore $\langle v_0, z_i - x_0\rangle + \langle v_0, x_0 - z_0\rangle \geq 0, \quad i = 1, 2$. In view of (5.16), the last inequalities are equivalent to

$$\langle v_0, z_i - x_0\rangle \geq n(x_0 - z_0), \quad i = 1, 2. \tag{5.18}$$

For $v_i \in \partial n(x_0 - z_i)$, in accordance with (2.1), we have

$$\langle v_i, z_i - x_0\rangle = -n(z_i - x_0), \quad n^*(v_i) = 1,\ i = 1, 2. \tag{5.19}$$

Now, substituting (5.17) in (5.18) and using (5.19) we obtain

$$-\alpha n(z_1 - x_0) + (1-\alpha)\langle v_2, z_1 - x_0\rangle \geq n(x_0 - z_0), \tag{5.20}$$

$$\alpha\langle v_1, z_2 - x_0\rangle - (1-\alpha) n(z_2 - x_0) \geq n(x_0 - z_0). \tag{5.21}$$

However, since $n^*(v_i) = 1,\ i = 1, 2$, it follows by (5.8) that

$$\langle v_2, z_1 - x_0\rangle \leq n(z_1 - x_0), \quad \langle v_2, z_2 - x_0\rangle \leq n(z_2 - x_0).$$

Hence for $\alpha \in [1/2, 1]$ inequality (5.20) leads to a contradiction. For $\alpha \in [0, 1/2]$ contradiction results from (5.21).

d) Let D be a convex compact set symmetric with respect to a point x^*. It means by Corollary 4.1 that $x^* \in X(D)$. Since $\Phi(x)$ is the convex function we have $[x^*, x_0] \subset X(D)$, that is,

$$\Phi(x) = R(x) + P(x) = const \quad \text{for all } x \in [x^*, x_0]. \tag{5.22}$$

It is easy to see that point x^*, as center of symmetry, possesses the properties

$$R(x^*) = \min_{x \in \mathbb{R}^p} R(x), \quad P(x^*) = \min_{x \in \mathbb{R}^p} P(x).$$

Therefore the convex functions $R(x)$ and $P(x)$ are nonincreasing functions on the line segment $[x_0, x^*]$ when x tends from x_0 to x^*. Taking into account (5.22) we conclude that the functions $R(x)$ and $P(x)$ are the constant quantities on $[x_0, x^*]$. On the other hand since $x^* \in D$ and $x_0 \notin D$ it follows that $P(x^*) \leq 0$ and $P(x_0) = \rho_D(x_0) > 0$. Thus we obtain contradiction. The proof is over. $\square$

Remark 5.1. The conditions 1) and 2) of Theorem 5.1, which are sufficient for inclusion (5.2) mean the strict convexity and smoothness of balls respectively. Examples show that smoothness and strict convexity of convex compact set D are insufficient for validity of (5.2).

We now present an example when $X(D) \not\subset D$.

Example 5.1. Let $p = 3$, $x = (x^{(1)}, x^{(2)}, x^{(3)}) \in \mathbb{R}^3$, $n(x) = \max\limits_{i=1,2,3} |x^{(i)}|$, $D = \mathrm{co}\{z_1, z_2, z_3,\}$, where $z_1 = (1,1,1)$, $z_2 = (1,-1,-1)$, $z_3 = (-1,-1,1)$. Clearly, $x_0 = (0,0,0) \notin D$ and $Q^\rho(x_0, D) = \{z_0\}$, where $z_0 = (1/3, -1/3, 1/3)$. Using (2.1) it is now easy to obtain

$$\partial n(x_0 - z_0) = \mathrm{co}\,\{(-1,0,0), (0,1,0), (0,0,-1)\},$$

$$\partial n(x_0 - z_1) = \mathrm{co}\,\{(-1,0,0), (0,-1,0), (0,0,-1)\},$$

$$\partial n(x_0 - z_2) = \mathrm{co}\,\{(-1,0,0), (0,1,0), (0,0,1)\},$$

$$\partial n(x_0 - z_3) = \mathrm{co}\,\{(1,0,0),(0,1,0),(0,0,-1)\}. \tag{5.23}$$

We have $\{z_1, z_2, z_3\} \subset Q^R(x_0, D)$, therefore $\mathrm{co}\,\{\partial n(x_0 - z_i) : i = 1,2,3\} \subset \subset \underline{\partial R}(x_0))$ by (2.11). Substituting here (5.23) we obtain

$$\mathrm{co}\,\{(1,0,0),(0,-1,0),(0,0,1)\} \subset \underline{\partial R}(x_0). \tag{5.24}$$

It is easy to calculate that $K^+(z_0, D) = \{x = (\lambda, -\lambda, \lambda) \in \mathbb{R}^3 : \lambda \in \mathbb{R}^1\}$. Hence $-\underline{\partial \rho}_D(x_0) = \partial n(z_0 - x_0) \cap K^+(z_0, D) = \{y_0\}$, where $y_0 = (1/3, -1/3, 1/3)$. Since this point can be represented as

$$y_0 = \frac{1}{3}(1,0,0) + \frac{1}{3}(0,-1,0) + \frac{1}{3}(0,0,1),$$

and with (5.24) taken into account, it follows that $y_0 \in \underline{\partial R}(x_0) \cap -\underline{\partial \rho}_D(x_0)$, that is, $0_3 \in \underline{\partial R}(x_0) + \underline{\partial \rho}_D(x_0)$. Hence, in accordance with (4.1), the point x_0 lies in $X(D)$. However, $x_0 \notin D$, therefore $X(D) \not\subset D$.

6 Condition of the uniqueness of solution

Examples 4.1 and 5.1 simultaneously show that solution of problem (1.1) can be nonunique. Let us give sufficient conditions of the uniqueness of the solution.

Theorem 6.1. *If at least one of the following conditions is valid:*

1) $n(\cdot)$ is a strictly quasiconvex norm,

2) D is a strictly convex set and $n(\cdot)$ is a smooth norm,

3) D is a strictly convex set and it possesses central symmetry,

then the solution of problem (1.1) is unique.

Proof. Assume that a solution is not unique and there exist at least two best approximation balls with centers at $x_1, x_2 \in X(D)$. By Theorem 5.1 we have the inclusion $X(D) \subset D$, and since $[x_1, x_2] \subset X(D)$, it follows that

$$\Phi(x) = R(x) - \rho_\Omega(x) = \min_{x \in \mathbb{R}^p} \Phi(x) = \Phi(x_1), \quad \forall x \in [x_1, x_2]. \tag{6.1}$$

Hence the convex function $R(x)$ and concave function $\rho_\Omega(x)$ are affine on $[x_1, x_2]$.

(a) Let $n(\cdot)$ be a strictly quasiconvex norm. Assume for definiteness that $R(x_2) \geq R(x_1)$. Since the function $R(x)$ is affine on $[x_1, x_2]$, it follows by Lemma 2.6 that

$$R(x_2) = R(x_1) + n(x_1 - x_2). \tag{6.2}$$

By (6.1) and (6.2) we also obtain

$$\rho_\Omega(x_2) = \rho_\Omega(x_1) + n(x_1 - x_2). \tag{6.3}$$

It now follows from (6.3) that

$$R(x_1) > \rho_\Omega(x_1). \tag{6.4}$$

Actually, if one assumes that $R(x_1) = \rho_\Omega(x_1)$, then the inclusions

$$Bn(x_1, \rho_\Omega(x_1)) \subset D \subset Bn(x_1, R(x_1))$$

show that $D = Bn(x_1, \rho_\Omega(x_1))$ and therefore $\rho_\Omega(x_1) = \max_{x \in D} \rho_\Omega(x)$. This contradicts (6.3).

We conclude from (6.2) and Lemma 5.1 that boundary spheres of the nested balls $Bn(x_1, R(x_1))$ and $Bn(x_2, R(x_2))$ are tangent at the unique point

$$y_0 = x_1 + \frac{x_1 - x_2}{n(x_1 - x_2)} R(x_1). \tag{6.5}$$

Moreover, in view of the inclusions

$$D \subset Bn(x_1, R(x_1)) \subset Bn(x_2, R(x_2))$$

and since the ball $Bn(x_2, R(x_2))$ must have a common point with D, it follows that $y_0 \in D$. Next, $x_2 \in \operatorname{int} D$ by (6.3), therefore

$$\alpha x_2 + (1-\alpha) y_0 \in \operatorname{int} D, \quad \text{for all } \alpha \in (0,1). \tag{6.6}$$

On the other hand we see from (6.3) and Lemma 5.1 that the boundary spheres of the nested balls $Bn(x_1, \rho_\Omega(x_1))$ and $Bn(x_2, \rho_\Omega(x_2))$ are tangent at the unique point

$$y_1 = x_1 + \frac{x_1 - x_2}{n(x_1 - x_2)} \rho_\Omega(x_1). \tag{6.7}$$

Moreover, in view of the inclusions

$$Bn(x_1, \rho_\Omega(x_1)) \subset Bn(x_2, \rho_\Omega(x_2)) \subset D$$

and since the ball $Bn(x_1, \rho_\Omega(x_1))$ has a common point with Ω, it follows that

$$y_1 \in \Omega. \tag{6.8}$$

From (6.4), (6.5), and (6.7) we see that $y_1 = \alpha_0 x_2 + (1 - \alpha_0) y_0$ for some $\alpha_0 \in (0,1)$. In view of (6.6), this contradicts (6.8).

(b) Let now D be a strictly convex compact set and $n(\cdot)$ be a smooth norm. In accordance with the definition of a strictly convex set[17] we have $\alpha x_1 + (1-\alpha) x_2 \in$ *int* D for all $\alpha \in (0,1)$. Therefore without loss of generality we shall assume that $x_1, x_2 \in intD$ and $\rho_\Omega(x_1) \le \rho_\Omega(x_2)$. The function $\rho_\Omega(x)$ is affine on $[x_1, x_2]$, therefore application of Lemma 2.10 gives us

$$\rho_\Omega(x_2) = \rho_\Omega(x_1) + n(x_1 - x_2). \tag{6.9}$$

By (6.1) and (6.9) we also obtain

$$R(x_2) = R(x_1) + n(x_1 - x_2). \tag{6.10}$$

Now from affineness of functions $R(x)$ and $\rho_\Omega(x)$ on $[x_1, x_2]$ and equalities (6.9),(6.10) it follows

$$R'(x_1, x_2 - x_1) = \rho'_\Omega(x_1, x_2 - x_1) = n(x_1 - x_2). \tag{6.11}$$

b1) We claim that there exists element $z_0 \in Q^R(x_1, D)$ such that

$$n'(x_1 - z_0) = n'(x_2 - x_1). \tag{6.12}$$

Let us recall that the smoothness of norm $n(\cdot)$ means[2]

$$\partial n(x) = \{n'(x)\}, \quad x \neq 0_p. \tag{6.13}$$

Using relation between $n(\cdot)$ and $n^*(\cdot)$ we can write

$$n(x_2 - x_1) = \max_{n^*(v) \leq 1} \langle v, x_2 - x_1 \rangle. \tag{6.14}$$

From (6.13) and (2.1) we have

$$n(x) = \langle n'(x), x \rangle, \quad n^*(n'(x)) = 1 \quad \text{for all } x \neq 0_p. \tag{6.15}$$

By lemma 2.3 $n^*(\cdot)$ is a strictly quasiconvex norm. Therefore the ball $Bn^*(0_p, 1) = \{v \in \mathbb{R}^p : n^*(x) \leq 1\}$ is a strictly convex set. Then from (6.14),(6.15) we obtain

$$n(x_2 - x_1) < \langle v, x_2 - x_1 \rangle \text{ for } v \in Bn^*(0_p, 1),\ v \neq n'(x_2 - x_1). \tag{6.16}$$

Applying Lemma 2.5 and using the property of subdifferential as differential characteristic of a convex function[1,2,3,18] we have

$$R'(x_1, x_2 - x_1) = \max_{v \in \underline{\partial R}(x_1)} \langle v, x_2 - x_1 \rangle = n(x_2 - x_1), \tag{6.17}$$

$$\underline{\partial R}(x_1) = co\,\{n'(x_1 - z) : z \in Q^R(x_1, D)\}. \tag{6.18}$$

Also by (6.15) and (6.18) it follows that

$$\underline{\partial R}(x_1) \subset Bn^*(0_p, 1). \tag{6.19}$$

Now from relations (6.15)–(6.19) we obtain (6.12).

b2) Let us show that superdifferential $\overline{\partial \rho_\Omega}(x_1)$ is a singleton, moreover

$$\overline{\partial \rho_\Omega}(x_1) = \{n'(x_2 - x_1)\}, \quad n'(x_2 - x_1) = n'(x_1 - z) \text{ for all } z \in Q^\rho(x_1, \Omega). \tag{6.20}$$

Indeed, applying Lemma 2.9 and using the differential property of superdifferential for a concave function, and also (6.11),(6.13) we have

$$\rho'_\Omega(x_1, x_2 - x_1) = \min_{w \in \overline{\partial \rho_\Omega}(x_1)} \langle w, x_2 - x_1 \rangle = n(x_2 - x_1), \tag{6.21}$$

$$\overline{\partial}\rho_\Omega(x_1) = co\,\{n'(x_1 - z) : z \in Q^\rho(x_1, \Omega)\}. \tag{6.22}$$

Since $\overline{\partial}\rho_\Omega(x_1) \subset Bn^*(0_p, 1)$ by (6.15) and (6.22), (6.20) follows from (6.15),(6.16) and (6.21).

b3) According to (6.12) and (6.20) there exists some element $z_0 \in Q^R(x_1, D)$, for which

$$n'(x_1 - z_0) = n'(x_1 - z) \quad \text{for all } z \in Q^\rho(x_1, \Omega). \tag{6.23}$$

Let us show that (6.23) leads to contradiction with our assumption of nonuniqueness of solution.

First we notice that for given normal $w_0 \neq 0_p$ we can construct unique support hyperplane to the convex set D

$$\pi = \{x \in \mathbb{R}^p : \langle w_0, x\rangle = \lambda\},$$

such that

$$\pi^+ = \{x \in \mathbb{R}^p : \langle w_0, x\rangle \leq \lambda\} \subset D. \tag{6.24}$$

The point $z_0 \in Q^R(x_1, D)$ is the boundary point of the convex compact set D and the ball $Bn(x_1, R(x_1))$. Since $n(\cdot)$ is a smooth norm then the support hyperplane of the ball $Bn(x_1, R(x_1))$ and set D at the point z_0 must be unique. Moreover the normal of this support hyperplane is $w_0 = n'(x_1 - z_0)$. On the other hand any $z \in Q^\rho(x_1, \Omega)$ is the boundary point of the convex compact set D and the ball $Bn(x_1, \rho_\Omega(x_1))$. Their support hyperplane at the point z can be constructed in unique way and its normal is $w_0 = n'(x_1 - z)$. In both cases we suppose the inclusion (6.24).

Taking into account (6.23) and the arguments stated above we conclude that there exists hyperplane which is a support hyperplane of the balls $Bn(x_1, \rho_\Omega(x_1))$ and $Bn(x_1, R(x_1))$ simultaneously. It is possible only for case $R(x_1) = \rho_\Omega(x_1)$. This means that the set D is the ball with center at point x_1, which is a unique solution of problem (1.3).

(c) Let D be a strictly convex compact set possessing central symmetry. Without loss of generality we assume that its center of symmetry is the point $x_1 = 0_p$. By Theorems 3.1, 5.1 and Corollary 4.1 it follows that

$$\min_{x\in\mathbb{R}^p,\, r\geq 0} h(D, Bn(x, r)) = h(D, Bn(0_p, r_1)) = h_0, \tag{6.25}$$

$$r_1 = \frac{R(0_p) + \rho_\Omega(0_p)}{2}, \quad h_0 = \frac{R(0_p) - \rho_\Omega(0_p)}{2}. \tag{6.26}$$

Supposing that $Bn(x_2, r_2)$, where $x_2 \neq 0_p$, is also the best approximation ball, we have

$$h(D, Bn(x_2, r_2)) = h_0, \tag{6.27}$$

$$r_2 = \frac{R(x_2) + \rho_\Omega(x_2)}{2}. \tag{6.28}$$

Since the point $x_1 = 0_p$ is the center of symmetry for the set D, then

$$R(0_p) = \min_{x \in \mathbb{R}^p} R(x), \quad \rho_\Omega(0_p) = \max_{x \in D} \rho_\Omega(x).$$

Therefore the convex function $R(x)$ is nonincreasing function and concave function $\rho_\Omega(x)$ is nondecreasing function on $[x_1, x_2]$. Taking into account (6.1) we conclude that the functions $R(x)$ and $P(x)$ are the constant quantities on $[x_1, x_2]$. By (6.26) and (6.28) it follows that

$$r_1 = r_2. \tag{6.29}$$

Let us consider Minkowski function generated by the set D

$$\varphi_D(x) = inf\,\{\alpha > 0 : x \in \alpha D\}.$$

It is easy to see that $\varphi_D(x)$ is a strictly quasiconvex norm. Therefore, as it is proved in (a), the problem

$$h(B\varphi_D(x, \rho), Bn(0_p, r_1)) \to \min_{x \in \mathbb{R}^p,\, \rho \geq 0}, \tag{6.30}$$

where $B\varphi_D(x, \rho) = \{y \in \mathbb{R}^p : \varphi_D(x - y) \leq \rho\}$, have unique solution. Moreover in accordance with Corollary 4.1

$$\min_{x \in \mathbb{R}^p\ \rho \geq 0} h(B\varphi_D(x, \rho), Bn(0_p, r_1)) = \min_{x \in \mathbb{R}^p\ \rho \geq 0} h(B\varphi_D(0_p, \rho), Bn(0_p, r_1)) =$$

$$= \min_{\rho \geq 0} h(\rho D, Bn(0_p, r_1)). \tag{6.31}$$

Using the symmetry of set D with respect to the point $x_1 = 0_p$ and Lemma 3.2 we obtain

$$h(\rho D, Bn(0_p, r_1)) = \max\{\rho R(0_p) - r_1, r_1 - \rho_\Omega(0_p)\}. \tag{6.32}$$

Now by (6.31),(6.32) and (6.25),(6.26) it follows that

$$\min_{x \in \mathbb{R}^p\ \rho \geq 0} h(B\varphi_D(x, \rho), Bn(0_p, r_1)) = h(D, Bn(0_p, r_1)) = h_0,$$

that is, $D = B\varphi_D(0_p, 1)$ is the best approximation ball for set $Bn(0_p, r_1)$ and norm $\varphi_D(\cdot)$ in problem (6.30).

However since

$$h(D, Bn(x_2, r_2)) = h(D - x_2, Bn(0_p, r_2)) =$$

$$= h(B\varphi_D(-x_2, 1), Bn(0_p, r_2)),$$

then by (6.27) and (6.29) we have

$$h(B\varphi_D(-x_2, 1), Bn(0_p, r_1)) = h_0,$$

that is, $B\varphi_D(-x_2, 1)$ is also the best approximation ball in problem (6.30). It contradicts the uniqueness of solution of problem (6.30). The proof is over. $\square$

REMARK 6.1. The strict quasiconvexity of the norm means the strict convexity of balls with respect to this norm. We have shown that the problem (1.3) is uniquely soluble in this case. On the contrary the following example shows that the strict convexity of the set D doesn't guarantee the uniqueness of solution.

EXAMPLE 6.1. Let $p = 2$, $x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^2$, $n(x) = \max\{|x^{(1)}|, |x^{(2)}|\}$; $D = \{x \in \mathbb{R}^2 : (x^{(1)}+1)^2 + (x^{(2)}-1)^2 \le 4, (x^{(1)}-2)^2 + (x^{(2)}+2)^2 \le 10\}$ is a strictly convex compact set, $x_0 = (0, 0)$, $x_1 = ((1 + \sqrt{2} - \sqrt{5})/2, (\sqrt{5} - \sqrt{2} - 1)/2)$.

(a) It is easy to verify that $Q^R(x_0, D) = \{z_1; z_2\}$, where $z_1 = (-1, -1)$, $z_2 = (1, 1)$ and $Q^\rho(x_0, \Omega) = \{z_0\}$, where $z_0 = (2 - \sqrt{5}, \sqrt{5} - 2)$. Bearing in mind that $n^*(x) = |x^{(1)}| + |x^{(2)}|$ and using (2.1) we obtain

$$\partial n(x_0 - z_0) = \mathrm{co}\,\{(1, 0), (0, -1)\}, \quad \partial n(x_0 - z_1) = \mathrm{co}\,\{(1, 0), (0, 1)\},$$

$$\partial n(x_0 - z_2) = \mathrm{co}\,\{(-1, 0), (0, -1)\},$$
$$K^+(z_0, D) = \{x \in \mathbb{R}^2 : x^{(1)} = -x^{(2)}, x^{(1)} \ge 0\}.$$

Next, using formulae (2.11) and (2.21) we obtain

$$\begin{aligned} \underline{\partial R}(x_0) &= \{x \in \mathbb{R}^2 : |x^{(1)}| + |x^{(2)}| \le 1\}, \\ \overline{\partial \rho_\Omega}(x_0) &= \{(1/2, -1/2)\}. \end{aligned} \tag{6.33}$$

Since $x_0 \in \mathrm{int}\, D$, it follows that $\underline{\partial P}(x_0) = -\overline{\partial \rho_\Omega}(x_0)$. Hence we see from (6.33) that $0_2 \in \underline{\partial R}(x_0) + \underline{\partial P}(x_0) \equiv \underline{\partial \Phi}(x_0)$; this means that$x_0 \in X(D)$ and x_0 is the center of some best approximation ball.

(b) We can also verify that $Q^R(x_1, D) = \{z_1; z_2\}$ and $Q^\rho(x_1, \Omega) = \{z_0; z_3\}$, where $z_3 = (\sqrt{2} - 1, 1 - \sqrt{2})$. Moreover,

$$\partial n(x_1 - z_0) = \mathrm{co}\,\{(1, 0), (0, -1)\}, \ \partial n(x_1 - z_3) = \mathrm{co}\,\{(-1, 0), (0, -1)\},$$

$$\partial n(x_1 - z_1) = \{(1, 0)\}, \quad \partial n(x_1 - z_2) = \{(0, -1)\},$$
$$K^+(z_3, D) = \{x \in \mathbb{R}^2 : x^{(1)} = -x^{(2)}, x^{(1)} \le 0\}.$$

Next, using (2.11) and (2.21), we obtain

$$\underline{\partial R}(x_1) = \mathrm{co}\,\{\partial n(x_1 - z_1), \partial n(x_1 - z_2)\} = \mathrm{co}\,\{(1, 0); (0, -1)\},$$

$$\begin{aligned} \overline{\partial \rho_\Omega}(x_1) &= \mathrm{co}\,\{v \in K^+(z_0, D) \cup K^+(z_3, D) : n^*(v) = 1\} = \\ &= \mathrm{co}\,\{(-1/2, 1/2); (1/2, -1/2)\}. \end{aligned}$$

Similarly to (a) we can now conclude that $0_2 \in \underline{\partial \Phi}(x_1)$ and x_1 too is the center of a best approximation ball. Thus, the problem has more than one solution in this case.

7 Variational properties of solution

1. Let $Kv(\mathbb{R}^p)$ be the space of all non-empty convex compact subsets of $\mathbb{R}^p$. It is easy to see that the Hausdorff distance $h(A,B)$ induced by an arbitrary norm $n(\cdot)$ is a metric in $Kv(\mathbb{R}^p)$. Generalizing results of M.Nikol'skii and D.Silin[8] we can discuss some properties of the function

$$h_0(D) = \min_{x\in\mathbb{R}^p,\, r\geq 0} h(D, Bn(x,r)) : Kv(\mathbb{R}^p) \to \mathbb{R}^1$$

and the set-valued map $X(D) : Kv(\mathbb{R}^p) \to 2^{\mathbb{R}^p}$.

THEOREM 7.1. The function $h_0(D)$ is Lipschitz on $Kv(\mathbb{R}^p)$ with Lipschitz constant 1, that is,

$$|h_0(D_1) - h_0(D_2)| \leq h(D_1, D_2), \quad \forall D_1, D_2 \in Kv(\mathbb{R}^p). \tag{7.1}$$

PROOF. Let $Bn(x_1, r_1)$ be a best approximation ball for D_1, and $Bn(x_2, r_2)$ is similar ball for D_2. Then, using the properties of a metric we obtain

$$h_0(D_1) = h(D_1, Bn(x_1, r_1)) \leq h(D_1, Bn(x_2, r_2)) \leq$$

$$\leq h(D_1, D_2) + h(D_2, Bn(x_2, r_2)) = h(D_1, D_2) + h_0(D_2),$$

that is,

$$h_0(D_1) - h_0(D_2) \leq h(D_1, D_2). \tag{7.2}$$

In a similar way

$$h_0(D_2) - h_0(D_1) \leq h(D_1, D_2). \tag{7.3}$$

From (7.2) and (7.3) we deduce (7.1). $\square$

THEOREM 7.2. The set-valued map $X(D) : Kv(\mathbb{R}^p) \to 2^{\mathbb{R}^p}$ is upper semicontinuous on $Kv(\mathbb{R}^p)$.

PROOF. We fix $D_0 \in Kv(\mathbb{R}^p)$; and let M be the set of $D \in Kv(\mathbb{R}^p)$ such that $h(D, D_0) \leq 1$. It is easy to see that there exists a finite number λ such that

$$\|x\| + r < \lambda, \quad \text{for each } x \in X(D), \text{ and for } r = (R(x) - P(x))/2, \quad D \in M. \tag{7.4}$$

Consider an arbitrary sequence $\{D_i\}_{i=1,2,\ldots} \subset Kv(\mathbb{R}^p)$, such that $h(D_i, D_0) \to 0$, $i \to \infty$. By (7.4) we can choose $x_i \in X(D_i)$ and $r_i = (R(x_i) - P(x_i))/2$ such that $x_i \to x^*$ and $r_i \to r^*$ as $i \to \infty$. We claim that $x^* \in X(D_0)$.

Assume that, on the contrary,

$$h(D_0, Bn(x^*, r^*)) - h_0(D_0) = \delta > 0.$$

Then for sufficiently large i we have

$$h(D_i, Bn(x_i, r_i)) - h(D_i, Bn(x_0, r_0)) \geq \delta/2 > 0, \tag{7.5}$$

where $x_0 \in X(D_0)$ and $r_0 = (R(x_0) - P(x_0))/2$, that is, $Bn(x_0, r_0)$ is one of the best approximation balls for D_0. However, (7.5) contradicts the assumption that $Bn(x_i, r_i)$ is a best approximation ball for D_i. □

Immediately from Theoremes 6.1 and 7.1 we obtain

COROLLARY 7.1. *If $n(\cdot)$ is a strictly quasiconvex norm, then the map $X(D)$ is single-valued and continuous on $Kv(\mathbb{R}^p)$.*

REMARK 7.1. In the following example we demonstrate that the set-valued map $X(D) : Kv(\mathbb{R}^p) \to 2^{\mathbb{R}^p}$ is not always lower semicontinuous.

EXAMPLE 7.1. Let $p = 3$, $x = (x^{(1)}, x^{(2)}, x^{(3)}) \in \mathbb{R}^3$,

$$n(x) = \max\{|x^{(1)}|, |x^{(2)}|, |x^{(3)}|\}.$$

It was shown in Example 4.1 that each point $x \in [(0, 0, 1/2), (0, 0, -1/2)]$ is the center of the best approximation ball for set

$$D_0 = \mathrm{co}\,\{(0, 1, 1/2), (0, 1, -1/2), (0, -1, 1/2), (0, -1, -1/2)\}.$$

Let $D_i = \mathrm{co}\,\{D_0, (1/i, 0, 0), (-1/i, 0, 0)\}$. Clearly, $h(D_0, D_i) \to 0$ as $i \to \infty$, and the D_i are compact sets symmetric with respect to the same point $x_0 = (0, 0, 0)$. Hence by Corollary 4.1 $x_0 \in X(D_i)$. Moreover x_0 is the center of a ball of the smallest radius containing D_i and, more importantly, x_0 is the unique maximum point of $\rho_{\Omega_i}(x)$ in $\Omega_i = \overline{\mathbb{R}^3 \setminus D_i}$. Hence $X(D_i) = \{x_0\}$. This shows that elements of $X(D_0)$ distinct from x_0 cannot be limit points of elements of $X(D_i)$. Hence the set-valued map $X(D))$ is not lower semicontinuous on element $D_0 \in Kv(\mathbb{R}^p)$.

2. Consider two classes of transformation : 1) displacements $Tx = x + a$, where $a \in \mathbb{R}^p$ is a fixed element; 2) extensions $Tx = \lambda x$, where λ is a fixed quantity.

It is not difficult to prove that the properties of displacements and extensions for the Euclidean norm[8] are reserved for an arbitrary norms.

THEOREM 7.3. *For the displacements there are fulfilled the relations*

$$r_0(TD, Tx) = r_0(D, x), \quad h_0(TD) = h_0(D), \tag{7.6}$$

and for the extensions, respectively,

$$r_0(TD, Tx) = |\lambda| r_0(D, x), \quad h_0(TD) = |\lambda| h_0(D),$$

where $r_0(D, x)$ is the radius of the best approximation ball for the set D with center $x \in X(D)$. Moreover for both classes of transformations

$$X(TD) = TX(D). \tag{7.7}$$

REMARK 7.2. In the work of M.Nikol'skii and D.Silin[8] it was shown for the case of Euclidean norm that orthogonal linear transformations also possess the properties (7.6) and (7.7). Simple examples show that it can be wrong for non-Euclidean norms.

8 Reduction to a linear programming problem

Everywhere in this section suppose that a convex compact set D is the polytope of the form
$$D = \{y \in \mathbb{R}^p : \langle A_i, y\rangle + a_i \geq 0, \quad i = \overline{1,m}\}, \tag{8.1}$$
where $A_i \in \mathbb{R}^p$, $a_i \in \mathbb{R}^1$, $i = \overline{1,m}$. And let unit ball of the norm also be a polytope formed as
$$Bn(0_p, 1) = \{y \in \mathbb{R}^p : \langle B_j, y\rangle + b_j \geq 0, \quad j = \overline{1,l}\}, \tag{8.2}$$
where $B_j \in \mathbb{R}^p$, $b_j \in \mathbb{R}^1$, $b_j > 0$, $j = \overline{1,l}$. Remark that the example of such norm is $n(x) = \max\limits_{i=\overline{1,p}} |x^{(i)}|$.

As it follows from Theorems 3.1 and 4.2, for obtaining at least one of the solutions of the problem (1.3) we can to find the minimum point of the function $\Phi(x)$ on the set D. That is, taking into account that $P(x) = -\rho_\Omega(x)$ for $x \in D$, it is sufficient to solve the problem
$$R(x) - \rho_\Omega(x) \to \min_{x\in D}. \tag{8.3}$$
Define concretely the forms of the functions $R(x)$ and $\rho_\Omega(x)$ for the considered case.

LEMMA 8.1. *If the polytope D is given in form (8.1) and $x \in D$, then*
$$\rho_\Omega(x) = \min_{i=\overline{1,m}} \{\langle C_i, x\rangle + c_i\}, \tag{8.4}$$
where
$$C_i = \frac{A_i}{n^*(A_i)}, \qquad c_i = \frac{a_i}{n^*(A_i)}. \tag{8.5}$$

PROOF. Clearly, for the point $x \in D$
$$\rho_\Omega(x) = \min_{i=\overline{1,m}} \rho_{\pi_i}(x),$$
where $\pi_i = \{x \in \mathbb{R}^p : \langle A_i, x\rangle + a_i = 0\}$, $i = \overline{1,m}$ are the hyperplanes forming the sides of the polytope D. Hence, using the formula
$$\rho_{\pi_i}(x) = \frac{|\langle A_i, x\rangle + a_i|}{n^*(A_i)},$$
notations (8.5) and taking into account that $|\langle A_i, x\rangle + a_i| = \langle A_i, x\rangle + a_i$ for $x \in D$, we obtain (8.4). $\square$

LEMMA 8.2. *If the unit ball of the norm $n(\cdot)$ is given in form (8.2), then*
$$R(x) = \max_{j=\overline{1,l}} \{\langle T_j, x\rangle + t_j\}, \tag{8.6}$$
where
$$T_j = -\frac{B_j}{b_j}, \qquad t_j = \max_{y\in -D} \langle T_j, y\rangle. \tag{8.7}$$

PROOF. By (8.2) it is not difficult to conclude

$$Bn^*(0_p,1)=\{v\in\mathbb{R}^p: n^*(v)\le 1\}=co\{-\frac{B_j}{b_j}: j=\overline{1,l}\}.$$

Hence, using the formula $n(x)=\max\limits_{n^*(v)\le 1}\langle v,x\rangle$, we have

$$R(x)=\max_{y\in D}\max_{n^*(v)\le 1}\langle v,x-y\rangle=\max_{y\in D}\max_{v\in co\{-B_j/b_j:j=\overline{1,l}\}}\langle v,x-y\rangle. \tag{8.8}$$

Since,

$$\max_{v\in co\{-B_j/b_j:j=\overline{1,l}\}}\langle v,x-y\rangle=\max_{v\in\{-B_j/b_j:j=\overline{1,l}\}}\langle v,x-y\rangle,$$

then, using notations (8.7), by (8.8) it follows that

$$R(x)=\max_{j=\overline{1,l}}\max_{y\in D}\langle T_j,x-y\rangle=\max_{j=\overline{1,l}}\{\langle T_j,x\rangle+\max_{y\in D}\langle T_j,-y\rangle\}=$$

$$=\max_{j=\overline{1,l}}\{\langle T_j,x\rangle+t_j\}. \square$$

Now we can reduce the problem (8.3) to a linear programming problem.

THEOREM 8.1. *If D and $Bn(0_p,1)$ are the polytopes given in forms (8.1) and (8.2) respectively, then the problem (8.3) is equivalent to the linear programming problem*

$$\begin{cases} z=x^{(p+1)}\to\min,\\ x^{(p+1)}-\langle T_j-C_i,x\rangle-t_j+c_i\ge 0, \quad i=\overline{1,m},\ j=\overline{1,l},\\ x\in D,\end{cases} \tag{8.9}$$

where T_j,t_j,C_i,c_i is defined by (8.7) and (8.5). Moreover, if the pair (x^*,z^*) is a solution of (8.9) then the point x^* is a solution of (8.3). Conversely, if x^* is a solution of (8.3), then the pair (x^*,z^*), where $z^*=R(x^*)-\rho_\Omega(x^*)$, is a solution of (8.9).

PROOF. Lemmae 8.1 and 8.2 allow us to rewrite the problem (8.3) in form

$$\max_{i=\overline{1,m},\,j=\overline{1,l}}\{\langle T_j-C_i,x\rangle+t_j-c_i\}\to\min_{x\in D}.$$

The equivalence of this problem and the problem (8.9) is quite clear. $\square$

9 Conclusion

The reduction of the problem (1.1) to the convex problem (1.3) enables us to get its approximate solution by using numerical methods of convex programming. The formulae for the subdifferentials of the functions $R(x)$ and $P(x)$, for example, make it possible to apply subgradient methods[2]. Constructing a numerical solution algorithm the authors used the idea of the Kelley's method[21,2]. This led to the necessity of solving problems of the form (8.9) at each step of the algorithm, which were the result of an outer estimating of a compact set under approximation and the unit ball of chosen norm by polytopes.

Authors wish to express their gratitude to M.S.Nicol'skii for valuable advice.

[21] J.F.Kelley, "The cutting–plane methods for solving convex problems", SIAM J.Appl. Math., 8:4, (1960), 703–712.

Aspects of Numbers Theory in terms of Potential Theory

N. Boboc and Gh. Bucur

1. Preliminaries concerning elementary Potential Theory.

Let $(X, \mathcal{B})$ be a measurable space and V be a kernel on $(X, \mathcal{B})$. If $p\mathcal{B}$ denotes the set of all positive numerical $\mathcal{B}$-measurable functions on X then V appear as map $V : p\mathcal{B} \to p\mathcal{B}$ which is additive, increasing, $V0 = 0$ and σ-continuous in order from below.

The principal notion in the elementary Potential Theory [7] associated with the kernel V is so called *V-supermedian function* (or simply supermedian function) i.e. a function $s \in p\mathcal{B}$ such that $Vs \leq s$. The Potential Theory associated to V means, roughly speaking, the study of the set of all supermedian functions denoted by $\bar{\mathcal{S}}$.

Obviously this set $\bar{\mathcal{S}}$ is a min-stable convex subcone of pB such that for any increasing (resp. decreasing) sequence $(s_n)_n$ in $\bar{\mathcal{S}}$ we have $\sup s_n \in \bar{\mathcal{S}}$ (resp. $\inf s_n \in \bar{\mathcal{S}}$.

Moreover for any $f \in p\mathcal{B}$ the function

$$Rf : \{s \in \bar{\mathcal{S}} / f \leq s\}$$

called, *the reduite of f*, is also supermedian and if there exists $s, t \in \bar{\mathcal{S}}$ with $f + t = s$ then there exists $t_1 \in \bar{\mathcal{S}}$ with $Rf + t_1 = s$. (Mokobodzki theorem)

We associate to V an other kernel, called *Green kernel*, $G = G_V$ given by

$$G = \sum_n V^n$$

where V° is the identity map and $V^{n+1} = V \circ V^n$, for all $n \in \mathbb{N}$. For any $f \in p\mathcal{B}$ the function Gf is supermedian and it is called *potential function*. If Gf is finite then f is uniquely determined by Gf since we have

$$Gf = f + V(Gf).$$

Also if p is a finite potential then any supermedian function s with $s \leq p$ is also a potential. The kernel G satisfies the following property: if s is V-supermedian, $f \in \mathcal{B}$ and $s \geq Gf$ an $\{x \in X \mid f(x) > 0\}$ then $s \geq Gf$ on X.

A function $h \in p\mathcal{B}$ such that $Vh = h$ is called V*-invariant* (or simply invariant). Obviously any invariant function is a supermedian one. Moreover any supermedian function can be decomposed in a summ between a potentil and an invariant function (Watanable theorem) and this decomposition is unique if the given supermedian function is finite.

In the sequel we denote by $\mathcal{S} = \mathcal{S}_V$ the set of all finite V-supermedian functions. In the set $\mathcal{S}$ we consider two order relations: the first is the usual order relation denoted $\leq$, called the *natural order relation* (i.e. $s \leq t \overset{\text{det}}{\Longleftrightarrow} s(x) \leq t(x) \forall x \in X$); the second, called *specific order relation*, denoted $\preceq$, is defined by

$$s \preceq t \overset{\text{det}}{\Longleftrightarrow} \exists u \in \mathcal{S} \text{ with } s + u = t$$

It is proved that the ordered sets $(\mathcal{S}, \leq)$, $(\mathcal{S}, \preceq)$ are ordered convex cones such that any subset of $\mathcal{S}$ possesses an infimum in both order relations. Moreover any specific majorant of a nonempty subset A of $\mathcal{S}$ is a specific majorant of the natural supremum of A in $\mathcal{S}$.

To have a consistent Potential Theory associated to the kernel V we assume that V is transient i.e. there exists $f_0 \in p\mathcal{B}$, $0 < f_0 \leq 1$ such that Gf is finite or equivalently, there exists $s \in \mathcal{S}$ such that

$$Vs(x) < s(x) \quad \forall x \in X.$$

In this case the liniar vector space $\mathcal{S} - \mathcal{S}$ is a vector lattice of real $\mathcal{B}$-measurable functions such that there exists a strictly positive function $f_0 \in p\mathcal{B}$, $f_0 \leq 1$ for which

$$f \in pB, \quad f \leq f_0 \Rightarrow f \in \mathcal{S} - \mathcal{S}.$$

Hence for any functional

$$\varphi : \mathcal{S} \longrightarrow \mathbb{R}_+$$

which is additive, increasing and continuous in order f from below there exists a unique positive σ-finite measure μ on $(X, \mathcal{B})$ such that $\varphi(f) = \int f d\mu$.

Also for any $s \in \mathcal{S}$ there exists an increasing sequence $(p_n)_n$ of potentials in $\mathcal{S}$ such that $\sup_n p_n = s$.

An element $s \in \mathcal{S}$ is called *subtractible* if for any $t \in \mathcal{S}$ we have

$$s \leq t \Longrightarrow s \preceq t.$$

It is easy to see that any $s \in \mathcal{S}$ which is V-invariant is subtractible. Also the set $\mathcal{S}_0$ of all subtractible elements of $\mathcal{S}$ is a convex cone which is solid in $\mathcal{S}$ with respect to the specific order.

If $s_n)_n$ is an increasing sequence in $\mathcal{S}_0$ dominated in $(\mathcal{S}, \leq)$ then $\sup_n s_n \in \mathcal{S}_0$.

An element $s \in \mathcal{S}$ is called *extremal* if for any $s \in \mathcal{S}$ with $s' \preceq s$ there exists $\alpha \geq 0$ with $s' = \alpha s$. It is easy to see that any extremal element of $\mathcal{S}$ is a V-potential or a V-invariant function.

2. Potential theory on special ordered sets

In this section $(X, \leq)$ is an ordered set such that for any $x, y \in X$ with $x \leq y$ the interval $[x, y] := \{z \in X \mid x \leq z \leq y\}$ is finite. In the sequel X will be considered as a measurable space endowed with the σ-algebra $\mathcal{B}$ of all subsets of X. The case X finite was considered in [2]. For other general cases see [3], [4] and [5].

We denote by m the measure on X given by $m(\{x\}) = 1$ for any $x \in X$.

If $x \in X$, a point $x' \in X$ is called *precedent of* x if $x' < x$ (i.e. $x' \leq x$ and $x' \neq x$) and there is no $z \in X$ with $x' < z < x$. We note that if $x \in X$ is not minimal then for any $z \in X$, $z < x$ there exists x' precedent of x with $x \leq x' < x$. The set of all precedents of x is denoted by Π_x. Obviously $\Pi_x = \emptyset$, iff x is minimal.

We denote by V the kernel on X given by

$$Vf(x) = \begin{cases} 0 \text{ if } x \text{ is minimal} \\ \displaystyle\sum_{x' \in \Pi_x} f(x') \text{ if } x \text{ is not minimal} \end{cases}$$

The kernel V is called *the associated kernel on the ordered set* $(X, \leq)$.

If $x \in X$, a point $x' \in X$ is called succesor of x if x is a precedent of x'. The set of all succesors of x is denoted by Π_x^*. Obviously $\Pi_x^* = \emptyset$ iff x is maximal in X and for any $x \in X$ which is not maximal and $z \in X$ with $x < z$ there exists a succesor x' of x with $x < x' \leq z$.

We denote by V^* the kernel on X given by

$$Vf(x) = \begin{cases} 0 \text{ if } x \text{ is maximal} \\ \displaystyle\sum_{x' \in \Pi_x^*} f(x') \text{ if } x \text{ is not maximal} \end{cases}$$

Remark *If we denote by $\leq^*$ the order relation on X given by*

$$x \leq^* y \Longleftrightarrow y \leq x$$

then, for any $x, y \in X$ with $x \leq^* y$, the interval $[x, y]$ in $(X, \leq^*)$ is finite. Also x in minimal in $(X, \leq^*)$ iff x is maximal in $(X, \leq)$ and x' is precedent of x in $(X, \leq^*)$ iff x' is a succesor of x in $(X, \leq)$. Hence the kernel V^* is the associated kernel of the ordered set $(X, \leq^*)$.

In the sequel we denote by G (resp. G^*) the Green kernel associated with the kernels V (resp. V^*) and by $\mathcal{S}$ (resp. $\mathcal{S}^*$) the set of all V (resp. V^*) supermedian functions. If $x \in X$ we denote by G_x (resp. G_x^*) the function $G_x = G1_{\{x\}}$ (resp. $G_x^* = G^*1_{\{x\}}$).

Since by definition for any $x, y \in X$ with $x \leq y$ there exists a finite system $(x_k)_{0 \leq k \leq n}$ with $x_0 = x$, $x_k \in \Pi_{x_{k+1}} \forall 0 \leq k < n$, $x_n = y$ it follows that for any $s \in \mathcal{S}$ we have

$$s(x_{k+1}) \geq Vs(x_{k+1}) \geq s(x_k) \;\forall\; 0 \leq k < n-1$$

and so

$$s(x) = s(x_0) \leq s(x_n) = s(y)$$

i.e. s is an increasing function on the ordered set $(X, \leq)$.

Proposition 2.1. For any two positive function f, g on X we have

$$\int f(V^*)^n g dm = \int gV^n f dm \; \forall n \in \mathbb{N}.$$

Particularly we have

$$G^*_y(x) = G_x(y) \quad \forall x, y \in X$$

Proof. We have

$$\int fV^* g dm = \sum_{x \in X} f(x) V^* g(x) =$$
$$\sum_x \sum_{x' \in \Pi^*_x} f(x) g(x') = \sum_{x \in X} \sum_{x \in \Pi_{x'}} f(x) g(x') =$$
$$= \sum_{x' \in X} \sum_{x \in \Pi_{x'}} f(x) g(x') = \sum_{x' \in X} g(x') V f(x') = \int gV f dm$$

i.e.

$$\int fV^* g dm = \int gV f dm.$$

Suppose now that

$$\int f(V^*)^n g dm = \int gV^n f dm \quad \forall f, g \geq 0.$$

Then

$$\int f(V^*)^{n+1} g dm = \int f(V^*)^n (V^* g) dm = \int V^* g V^n f dm$$
$$= \int gV(V^n f) dm = \int gV^{n+1} f dm.$$

Since

$$G = \sum_n V^n, \quad G^* = \sum_n V^{*n}$$

it follows

$$\int fG^*gdm = \int gGfdm \quad \forall f, g \geq 0$$

and particularly, taking $f = 1_{\{x\}}$, $g = 1_{\{y\}}$ we get

$$G^*_y(x) = \int 1_{\{x\}}G^*(1_{\{y\}})dm = \int 1_y G1_{\{x\}}dm = G_x(y)$$

Proposition 2.2 *For any $x, y \in X$ we have*

$$G_y(x) = 0 \quad \text{if } y \not\leq x,$$

$$G_y(x) = \sum_{x' \in \Pi_x} G_y(x') + 1_{\{y\}}(x).$$

Proof. Suppose that $y \not\leq x$. We show inductively that $V^n 1_y(x) = 0$. Indeed this relation holds, obviously for $n = 0$. Suppose that $V^n_{1_y}(x) = 0$. Since $y \not\leq x$ then $y \not\leq x'$ for all $x' \in \Pi_x$ and so

$$V^{n+1}1_y(x) = \sum_{x' \in \Pi_y} V^n 1_y(x') = 0.$$

From the above considerations we deduce

$$y \not\leq x \Rightarrow G_y(x) = 0.$$

If $y \neq x$ we get

$$G_y = G1_y = V(G1_y) + 1_y = VG_y + 1_y$$

and so

$$G_y(x) = \sum_{x \in \Pi_x} G_y(x') + 1_y(x).$$

Proposition 2.3. *For any $x, y \in X$ we have*

$$G_y(x) = card\mathcal{L}_{y,x}$$

where $\mathcal{L}_{y,x}$ is the set of all maximal totally ordered subsets of $[y, x]$.

Proof. We show that

$$G_y(x) = card\mathcal{L}_{y,x}$$

by induction following $card\,[y,x]$. If $card\,[y,x] = 0$ (i.e. $y \not\leq x$) the relation follows since $\mathcal{L}_{y,x} = \emptyset$ and $G_y(x) = 0$.

Suppose now that the relation holds for $card\,[y,x] = n$ and let x, y such that $card\,[y,x] = n+1$ (where $n \geq 0$). From the preceding Proposition we have

$$y = x \Rightarrow G_y(x') = 0 \quad \forall x' \in \Pi_x$$

and so

$$G_y(x) = \sum_{x' \in \Pi_x} G_y(x') + 1_y(x) = 1_y(x) = 1.$$

On the other hand

$$y = x \Rightarrow card\,\mathcal{L}_{y,x} = 1$$

Suppose that $y \neq x$. In this case from $card\,[y,x] = n+1$ it follows $y < x$ and therefore, using the preceeding Proposition, we have

$$G_y(x) = \sum_{x' \in \Pi_x} G_y(x').$$

On the other hand we have

$$x' \in \Pi_x, y < x \Rightarrow card\,[y,x'] = n$$

and so by hypothesis of induction,

$$G_y(x') = card\,\mathcal{L}_{y,x'}.$$

Hence

$$G_y(x) = \sum_{x' \in \Pi_x} card\,\mathcal{L}_{yx'} = card\,\mathcal{L}_{yx}$$

Proposition 2.4. *For any $x, y \in X$ the following assertion are equivalent:*

1) $x \leq y$

2) $G_x(y) > 0$

3) $G_x^* \leq G_y^*$

4) $G_y \leq G_x$

Proof. 1) $\Leftrightarrow$ 2) follows from Proposition 3)

1) $\Leftrightarrow$ 3). If $z \in X$ we have

$$G_x^*(z) = G_z(x), G_y^*(z) = G_z(y)$$

and therefore, G_z beeing increasing,

$$G_z(x) \leq G_z(y).$$

3) $\Leftrightarrow$ 1) If $G_x^* \leq G_y^*$ then we have

$$1 = G_x^*(x) \leq G_y^*(x) = G_x(y), \quad G_x(y) > 0$$

and so $x \leq y$.

1) $\Leftrightarrow$ 4) follows from 1) $\Leftrightarrow$ 3) and from $x \leq y \Leftrightarrow y \leq^* x$ using the fact that $G_z(u) = G_u^*(z) \; \forall z, u \in X$.

Proposition 2.5. *Suppose that*

i) for any $x \in X$ there exists $x' \in X$ minimal $x' \leq x$;

ii) the set of all minimal elements of X is at most countable;

iii) The kernel V^* is transient.

Then X is at most countable. Particularly the kernel V is also transient.

Proof. Let f_0 be a positive function on X, $0 < f_0 \leq 1$ such that $G^* f_0 < \infty$. For any $n \in \mathbb{N}^*$ we denote

$$A_n := [f_0 \geq \frac{1}{n}].$$

For any $x \in X$ we have

$$\sum_{y \in A_n} G_y^*(x) \leq n G^* f_0(x) < \infty$$

Since

$$G_y(x) = card\, \mathcal{L}_{y,x}$$

it follows that $\cup_{y\in A_n}[x,y]$ is finite. Therefore the set

$$\{y \in A_n \mid y \geq x\}$$

is finite. Since

$$\{y \in X \mid y \geq x\} \subset \cup_n \{y \in A_n \mid y \geq x\}$$

it follows that the set

$$\{y \in X \mid y \geq x\}$$

is at most countable.

By hypothesis we have

$$X = \bigcup_{\substack{p\in X \\ p\ minimal}} \{y \in X \mid y \geq p\}$$

and so X is at most countable.

Proposition 2.6. *We consider the following assertions:*

1. For any $x \in X$ the set of all minorants of x is finite.

2. Any V-invariant function on X is equal zero.

3. There is no strictly decreasing sequence in X.

4. For any $x \in X$ there exists a minimal element x' of X with $x' \leq x$.

5. For any $x \in X$ the set of all minimal elements of x' of X with $x' \leq x$ is finite.

Then we have

1) $\Rightarrow$ 2) $\Leftrightarrow$ 4) and 1) $\Leftrightarrow$ 4) + 5).

Generally 2) $\Rightarrow$ 1) and 4) $\Leftarrow$ 3) are not true.

Proof

1) $\Rightarrow$ 4) + 5) is trivial

4) + 5) $\Rightarrow$ 1) follows from the fact that for each x the set M_x of all minorants of x is given by

$$M_x = \cap_{x'\in A_x}[x', x]$$

where A_x is the set of all minimal x' elements of X with $x' \leq x$.

2) $\Leftarrow$ 3) Suppose that 3) is not true and let $(x_n)_n$ be a strictly decreasing sequence in X. It follows that the sequence $(G_{x_n})_n$ is increasing. Let us denote

$$h = \sup_n G_{x_n}$$

Since

$$VG_{x_n} = G_{x_n} \text{on} X \setminus \{x_n\}$$

$$VG_{x_n} \leq G_{x_n} \text{on} X$$

and

$$VG_{x_{n+1}}(x_n) = G_{x_{n+1}}(x_n) \geq G_{x_n}(x_n)$$

it follows that

$$h \geq Vh = \sup_n VG_{x_n} \geq \sup_n G_{x_n} = h.$$

i.e. h is V-invariant. By hypothesis $h = 0$ and therefore $G_{x_n} = 0$ contradiction.

3) $\Leftrightarrow$ 2) Suppose that there exists non zero V-invariant function h. If we put

$$h_0 = \sup_n nh$$

it follows that h_0 is V-invariant, $h_0 \neq 0$ and

$$h_0(x) > 0 \Leftrightarrow h_0(x) = +\infty.$$

Let $x \in X$ with $h_0(x) = +\infty$. Since

$$+\infty = h_0(x_0) = \sum_{x' \in \Pi_x} h_0(x')$$

then there exists $x' \in \Pi_x$ with $h_0(x') = +\infty$.

By induction we can construct a sequence $(x_n)_n$ in X such that $h_0(x_n) = +\infty$ and $x_{n+1} \in \Pi_{x_n}$. Obviously $(x_n)_n$ is strictly decreasing, contradiction.

1) $\Rightarrow$ 3) is trivial.

2) $\Rightarrow$ 1) is not generally true. We consider the ordered set $X = P \cup \{a\}$ with $a \notin P$ such that P is infinit any two diferent elements of P are incompatible and $a > p$ for

any $p \in P$. Obviously the assertion 1) is not true for this set. If h is a V-invariant function then, from the fact that any $p \in P$ is minimal, we have $h(p) = 0$ and so

$$h(a) = \sum_{p \in P} h(p) = 0.$$

3) $\Rightarrow$ 4) Let $x \in X$ be such that there is no minimal elements x' of X with $x' \le x$. Then there exists a strictly decreasing sequence $(x_n)_n$ in X with $x_0 = x$ which contradicts 3).

4) $\Rightarrow$ 3) is not generally true. We consider the ordered set $X = \{(x, y) \mid x \in Z, y \in \{0, 1\}\}$ where the order relation is the following:

$$(x, y) \le (x', y') \Leftrightarrow \quad x \le x', y = y' = 1 \text{ or } x = x', y = 0, y' = 1.$$

Obviously any element of X has a minorant minimal (i.e. (4) is true) but there exists a strictly decreasing sequence in X (i.e. 3) is not true).

Proposition 2.7 *Let $u = Gf$ be a finite V-potential. Then u will be subtractible iff $f(x) = 0$ for all non minimal element x of X then $u = 0$.*

Proof. Let $x = X$ not minimal. Then $G1_{\{x\}}$ is not subtractible. Indeed suppose that $G1_{\{x\}}$ is subtractible and let $x' \in X$ be such that $x' < x$. We get $G1_{\{x\}} \le G1_{\{x'\}}$ and therefore there exists v, V-supermedian with $G1_{\{x\}} + v = G1_{\{x'\}}$. Since $v = Gg$ with $g \ge 0$ it follows that

$$1_{\{x\}} + g = 1_{\{x'\}}$$

which is a contradiction. Hence if $u = Gf$ is subtractible and $x \in X$ is not minimal then $f(x)G1_{\{x\}}$, being specifically dominated by u, is also subtractible and therefore by the above considerations $f(x) = 0$. Conversely suppose that $f(x) = 0$ for any non minimal element x of X. To show that Gf is subtractible it will be sufficient to suppose that $f = 1_{\{x\}}$ where x is a minimal element of X. Indeed in this case if v is V-supermedian such that $G1_{\{x\}} \le v$ we have

$$\begin{aligned} V(v - G1_{\{x\}})(y) &= Vv(y) - V(G1_{\{x\}})(y) = Vv(y) - G1_{\{x\}}(y) \\ &\le v(y) - G1_{\{x\}}(y) \quad \forall y \in X, y \ne x \end{aligned}$$

and

$$V((v - G1_{\{x\}}))(x) = 0 \leq v(x) - G1_{\{x\}}$$

i.e. $v - G1_{\{x\}}$ is V-supermedian.

Proposition 2.8 *Let $u = Gf$ be a finite V-potential. Then u will be extremal in $\mathcal{S}$ iff there exists $\alpha \geq 0$ and $x \in X$ with $u = \alpha G_x$.*

Proof. Suppose that u is extremal in $\mathcal{S}$. We have

$$u = f(x) \cdot G_x + G(g)$$

where g is the function equal f in any $y \neq x$ and equal zero at x. From $f(x) \cdot G_x \preceq u$ it follows that

$$f(x)G_x = \alpha u$$

for a suitable $\alpha > 0$. We get $u = 1/\alpha f(x) \cdot G_x$, $u = \beta G_x$ where $\beta = \alpha f(x)$.

Conversely suppose that $u = \beta G_x$ with $x \in X$ and $\beta > 0$ and let $s \in \mathcal{S}$ be such that

$$s \preceq u$$

then s is a potential $s = Gf$ such that

$$0 \leq f \leq \beta 1_x.$$

Hence $f = \alpha 1_{\{x\}}$ with $\alpha \geq 0$ and so

$$s = \alpha G_x = \frac{\alpha}{\beta} u$$

A positive real function s on the ordered set $(X, \leq)$ is called *totally increasing* (resp. *totally decreasing*) function if for any system $(x_1, x_2, \ldots, x_n)$ of strictly minorants (resp. strictly majorant) of x such that $x_i \not\leq x_j$ for all $i \neq j$ we have

$$s(x) \geq \sum_{i=1}^{n} s(x_i).$$

Obviously any totally increasing (resp. totally decreasing) function is increasing (resp. decreasing). Also a totally decreasing function with respect to the order relation $\leq$ is nothing also that a totally increasing function with respect to the order relation $\leq^*$.

Proposition 2.9 *A positive real function s on the ordered set $(X, \leq)$ will be totally increasing iff s is V-supermedian.*

Proof. Suppose that s is V-supermedian. For any $x \in X$ let $J = \{x_1, x_2, \dots, x_n\}$ a finite system of strictly minorants of x such that $x_i \not\leq x_j$ for all $i \neq j$. We show that

$$s(x) \geq \sum_{i=1}^{n} s(x_i)$$

by induction following the natural number

$$\mu(J) := \sup_{1 \leq i \leq n} \; card\, [x_i, x].$$

If $\mu(J) = 1$ then each x_i is a precedent of x and in this case we have

$$s(x) \geq \sum_{x' \in \pi_x} s(x') \geq \sum_{i=1}^{n} s(x_2).$$

Suppose that the assertion is true for $q \leq m$ and suppose $\mu(x_1, \dots x_n) = m + 1$. Since x_i is a strictly minorant of x there exists a precedent x' of x with $x_i \leq x'$. We put $A = \{x' \in \Pi_x / \exists\, 1 \leq i \leq n \text{ with } x_i \leq x'\}$.

For any $x' \in A$ let us denote $J_{x'} = \{x_i \in J / x_i \leq x'\}$

Obviously we have $\mu(J_{x'}) \leq m$ and so

$$s(x') \geq \sum_{x_i \in J_{x'}} s(x_i),$$

Hence

$$s(x) \geq \sum_{x' \in \Pi_x} s(x') \geq \sum_{x' \in A} s(x') \geq \sum_{x' \in A} \sum_{x_i \in J_{x'}} s(x_i) \geq \sum_{x_i \in J} s(x_j).$$

Conversely suppose that s is totally increasing on the ordered set $(X, \leq)$. Then for any $x \in X$ any finite subset A of Π_x satisfies the relation

$$x, y \in A, x \neq y \Rightarrow x \not\leq y$$

and so

$$s(x) \geq \sum_{x' \in A} s(x').$$

The set A being arbitrary we deduce

$$s(x) \geq \sum_{x' \in \Pi_x} s(x').$$

Lemma 2.10 *Let $\mathcal{T}$ be a convex cone of positive real functions on ordered set $(X, \leq)$ such that:*

1) $\mathcal{T}$ is min-stable and for any $t, t_1, t_2 \in \mathcal{T}$ with $t \leq t_1 + t_2$ there exists $t', t'' \in \mathcal{T}$ with $t = t' + t''$, $t' \leq t_1$, $t'' \leq t_2$.

2) for any $x, y \in X$ such that $x \leq y$ and any $t \in \mathcal{T}$ we have $t(x) \leq t(y)$.

3) for any $x, y \in X$ such that $x \not\leq y$ there exists $t \in \mathcal{T}$ with $t(x) > 0$ and $t(y) = 0$.

Then $\mathcal{T} \subset \mathcal{S}$.

Proof. In fact we show that any $t \in \mathcal{T}$ is totally increasing function on the ordered set $(X, \leq)$. Let $t \in \mathcal{T}$, $x \in X$ and let $x_1, \ldots, x_n)$ a finite system of strictly minorants of x with $x_i \not\leq x_j$ for all $i \neq j$. For any $i, j \in \{1, 2, \ldots, n\}$ with $i \neq j$ let $t_{ij} \in \mathcal{T}$ be such that $t_{ij}(x_i) = 1$ and $t_{ij}(x_j) = 0$. Let us consider the element of $\mathcal{T}$ given by

$$t_i = \inf\{t_{ij} / i \neq j\}.$$

We have

$$t_i(x_i) = 1, \quad t_i(x_j) = 0 \; \forall j \neq i.$$

We consider the element $s_0 \in \mathcal{T}$, $s \in \mathcal{T}$ given by

$$s_0 = \sum_{i=1}^{n} t(x_i) \cdot t_i, s = \inf(t, s_0).$$

Obviously $s \leq s_0$. From the property 1) of $\mathcal{T}$ there exists $s_1, s_2, \dots, s_n \in \mathcal{T}$ with

$$s = \sum_{i=1}^{n} s_i, \quad s_i \leq t(x_i) \cdot t_i \ \forall 1 \leq i \leq n.$$

It follows $s_i(x_j) = 0 \ \forall j \neq i$ and

$$s_i(x_i) = s(x_i), \quad s_0(x_i) = t(x_i) = s(x_i)$$

and therefore

$$t(x_i) = s_i(x_i) \ \forall 1 \leq i \leq n.$$

From property 2) we get

$$t(x) \geq s(x) = \sum_{i=1}^{n} s_i(x) \geq \sum_{i=1}^{n} s_i(x_i) = \sum_{i=1}^{n} t(x_i)$$

i.e. t is totally increasing function.

Corollary 2.11 *Let W be a kernel on X such that the connex cone $\mathcal{S}_W$ of all real W-supermedian functions satisfies the following properties:*

a) $x \leq y \Rightarrow s(x) \leq s(y) \ \forall s \in \mathcal{S}_W$

b) $x \not\leq y \Rightarrow \ \exists s \in \mathcal{S}_W$ *with* $s(x) > 0$ *and* $s(y) = 0$.

Then $\mathcal{S}_W \subset \mathcal{S}$.

3. Potential theory on the set of natural numbers associated with the divisibility.

In this section X will be the set of all natural numbers $n \geq 2$ which is considered as a measurable space with respect to the σ-algebra of all subsets of X.

On X we distinguish two remarkable order relatins: the first is denoted by $x|y$ and is defined by

$$x|y \overset{\text{det}}{\Leftrightarrow} x \text{ is a divisor of } y;$$

the second is denoted by $x\|y$ and is defined by

$$x\|y \overset{\text{det}}{\Leftrightarrow} x|y \text{ and } (x, \frac{y}{x}) = 1$$

where if $u, v \in X$, (u, v) denotes the greatest common divizor of (u, v). This second order relation was considered in [6].

Obviously for any $x, y \in X$ we have

$$x\|y \Rightarrow x|y.$$

The set of minimal elements in the ordered set $(X, |)$ is the set P of all prime numbers and the set $\widetilde{P}$ of minimal elements in the ordered set $(X, \|)$ is the set of all natural numbers of the form p^k with $p \in P$ and $k \in \mathbb{N}^*$.

If $x \in X$ is of the form

$$x = p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

where $p_i \in P$, $i \neq j \Rightarrow p_i \neq p_j$ and $x_k \in \mathbb{N}^*$ then the set Π_x of the precedents of x in the ordered set $(X, |)$ is given by

$$\Pi_x = \{x/p_i \mid i \in \{1, 2, \dots k\}\}$$

and the set $\widetilde{\P}1i_2$ of the precedents of x in the ordered set $(X, \|)$ is given by

$$\widetilde{\Pi}_x = \{x/p_i^{x_i} \mid i \in \{1, 2, \dots, k\}\}.$$

In the sequel we denote by V, V^* (resp. $\widetilde{V}, \widetilde{V}^*$) the kernels associated with the ordered set $(X, |)$ (resp. $(X, \|)$) and by G, G^* (resp. $\widetilde{G}, \widetilde{G}^*$) the Green kernel associated with V, V^* (resp. $\widetilde{V}, \widetilde{V}^*$).

Also, we denote by $\mathcal{S}$, $\mathcal{S}^*$ (resp. $\widetilde{\mathcal{S}}$, $\widetilde{\mathcal{S}}^*$) the set of all finite V-supermedian, V^*-supermedian (resp. $\widetilde{V}$-supermedian, $\widetilde{V}^*$-supermedian) function.

Proposition 3.1 *For any $x, y \in X$*

$$x = \prod_{i=1}^{k} p_i^{x_i}, \quad y = \prod_{i=1}^{k} p_i^{y_i}, \quad x_i, y_i \in \mathbb{N}$$

such that $y|x$ *(resp.* $y\|x$*) we have*

$$G_y(x) = \frac{\left(\sum_{i=1}^{k}(x_i - y_i)\right)!}{\prod_{i=1}^{k}((x_i - y_i)!)}$$

$$(resp. \widetilde{G}_y(x) = l!)$$

where $l = card\{i \in 1, 2, \ldots, k\} \mid y_i = 0 \; and \; 0 < x_i\}$

Proof. Since $y|x$ it follows that $y_i \leq x_i$ for all $i \in \{1, 2, \ldots, k\}$. On the other hand a maximal chain starting in y and ending in x is a system of the following form

$$(y, yp_{i_1}, yp_{i_1}p_{i_2}, \ldots yp_{i_1}p_{i_2} \ldots p_{i_q})$$

where $q = \sum_{i=1}^{k}(x_i - y_i)$ and where for any $j \in \{1, 2, \ldots k\}$ we have

$$\text{card}\{s \in \{1, 2 \ldots q\} / i_s = j\} = x_j - y_j.$$

The set of all these systems coincides with the set of all partitions of the set $\{1, 2, \ldots q\}$ of the form $\{A_1, A_2 \ldots A_k\}$ where

$$\text{card} A_j = x_j - y_j \quad \forall 1 \leq j \leq k.$$

Hence

$$\text{card}\mathcal{L}(y, x) = \frac{q!}{\prod_{j=1}^{k}[(x_j - y_j)!]} = \frac{\left(\sum_{j=1}^{n}(x_j - y_j)\right)!}{\prod_{j=1}^{k}[(x_j - y_j)!]}.$$

Now, if $y\|x$ we may suppose

$$x = y \cdot p_1^{x_1} \cdot p_2^{x_2} \ldots p_l^{x_l}, \quad x_i \geq 1, p_i \not| y \;\; i = 1, 2, \ldots l$$

A maximal chain with respect to the order relation $\|$ starting in y and ending in x is of the form

$$\left(y, yp_i^{x_{i_1}}, yp_{i_1}^{x_{i_1}} p_{i_2}^{x_{i_2}}, \ldots yp_{i_1}^{x_{i_2}}, \ldots yp_{i_1}^{x_{i_1}} p_{i_2}^{x_{i_2}} \ldots p_{i_l}^{x_{i_l}}\right)$$

and therefore the cardinal of the set of these maximal chains is equal $l!$

Proposition 3.2. *Any element of $\mathcal{S}$ (resp. $\widetilde{\mathcal{S}}$) is a V-potential (resp $\widetilde{V}$-potential).*

Proof. The assertion follows from Proposition 2.5 and from the fact that for any $x \in X$ the set of all minorants of x in $(X, |)$ or $(X, \|)$ is finite.

Proposition 3.3. *Any non zero extremal element of $\mathcal{S}$ (resp. $\widetilde{\mathcal{S}}$) is of the form αG_x (resp. $\alpha \widetilde{G}_x$) with $\alpha > 0$ and $x \in X$. Also any subtractible element of $\mathcal{S}$ (resp. $\widetilde{\mathcal{S}}$) is of the form Gf where $f(x) = 0$ for all $x \neq P$ (resp. $x \notin \widetilde{P}$).*

Proof. The assertion follows from Proposition 2.7 and Proposition 2.8.

Proposition 3.4 *Any subtractible elements of $\mathcal{S}^*$ (resp. $\widetilde{\mathcal{S}^*}$) is V^* (resp. $\widetilde{V}^*$)-invariant*

Proof. The assertion follows from Proposition 2.7.

Proposition 3.5 *For any $x, y \in X$ we have*

$$x|y \Leftrightarrow G_x^* \leq G_y^*, \quad x\|y \Leftrightarrow \widetilde{G}_x^* \leq \widetilde{G}_y^*.$$
$$x|y \Leftrightarrow G_x(y) > 0, \quad x\|y \Leftrightarrow \widetilde{G}_x(y) > 0.$$

Proof. The assertion follows from Proposition 2.4.

Now we develop a compactification precedure which is usual in Potential theory [1].

We denote by L (resp. $\widetilde{L}$) the subset of $\mathcal{S}^*$ (resp. $\widetilde{\mathcal{S}}^*$) defined by

$$L = \{t \in \mathcal{S}^* / \sum_{x \in P} t(x) \leq 1\}$$

$$(\text{resp.} \widetilde{L} = \{t \in \widetilde{\mathcal{S}}^* / \sum_{x \in \widetilde{P}} t(x) \leq 1\}).$$

It is easy to see that L (resp. $\widetilde{L}$) is a convex subset of $\mathcal{S}^*$ (resp. $\widetilde{\mathcal{S}}^*$).

We endowed L (resp. $\widetilde{L}$) with the topology τ generated by the functions

$$t \longrightarrow t(x), x \in X$$

Obviously since X is countable the topology τ is metrisable.

Proposition 3.6 *The metric space (L, τ) (resp. $\widetilde{L}, \tau)$ is compact.*

Proof. Let $(t_n)_n$ be a sequence in L (resp. $\widetilde{L}$). Since X is countable and $t(x) \le 1$, $\forall x \in X$, there exists a subsequence $(t_{k_n})_n$ of $(t_n)_n$ such that $(t_{k_n}(x))_n$ is convergent for any $x \in X$. We consider the function

$$t(x) = \lim_{n\to\infty} t_{k_n}(x).$$

Since t_{k_n} is totally decreasing on $(X, |)$ (resp. $(X, \|)$) it follows that t is also totally decreasing on $(X, |)$ (resp. $(X, \|)$). Hence $t \in \mathcal{S}$ resp. $(t \in \mathcal{S}^*)$. We have

$$\sum_{x\in P} t(x) \le \liminf_{n\to\infty} \sum_{x\in P} t_{k_n}(x) \le 1$$

$$(\text{resp.} \sum_{x\in \widetilde{P}} t(x) \le \liminf_{n\to\infty} \sum_{x\in \widetilde{P}} t_{k_n}(x) \le 1)$$

and so $t \in L$ (resp. $\widetilde{L}$). Hence $(t_{k_n})_n$ converges in (L, τ) (resp. $(\widetilde{L}, \tau)$) to t.

Corollary 3.7 *For any $s \in L$ (resp. $\widetilde{L}$) there exists a probability measure θ on the extrem points of L (resp. $\widetilde{L}$) such that*

$$s = \int_E t d\theta(t)$$

where E is the G_σ-set of all extrem points of L (resp. $\widetilde{L}$).

Proof. We consider the liniar space $\mathcal{L}$ of all bounded real function on X endowed with the topology of simply convergence. Obviously L (resp. $\widetilde{L}$) is a metrisable convex compact subset of $\mathcal{L}$ and therefore by Choquet representation theorem the set of all extrem points of L is a G_δ-set and any $s \in L$ (resp. $\widetilde{L}$) is the baricenter of a probability measure θ on E. i.e.

$$s = \int_E t d\theta(t)$$

Notation. For any $t \in \mathcal{S}^*$ and any $x \in X$ we denote by t_x the function on X given by

$$t_x(y) = t(xy) \quad \forall y \in X.$$

Proposition 3.8 *For any $t \in \mathcal{S}^*$ and any $x \in X$ the function t_x belongs to $\mathcal{S}^*$ and we have*

$$t_x \leq t,$$

$$(V^* t_x)(y) = (V^* t)_x(y).$$

Particularly if t is V^*-invariant then t_x is also V^*-invariant and $t_x \preceq t$.

Proof. We have, for any $y \in X$,

$$\begin{aligned}(V^* t_x)(y) &= \sum_{p\in P} t_x(yp) = \sum_{p\in P} t(xyp) = \\ &(V^*t)(x \cdot y) = (V^*t)_x(y) \leq t(xy) = t_x(y) \\ &t_x(y) = t(xy) \leq t(y)\end{aligned}$$

and so $t_x \in \mathcal{S}^*$.

If t is V^*-invariant in $\mathcal{S}^*$ then $V^*t = t$ and so

$$(V^* t_x)(y) = (V^*t)_x(y) = t_x(y), \quad \forall y \in X,$$

i.e. t_x is V^*-invariant.

Corollary 3.9 *For any $s \in \mathcal{S}^*$, V^*-invariant and $s \neq 0$ there exists $s' \in \mathcal{S}^*$, V^*-invariant, $s' \leq s$ such that $\sum_{p\in P} s'(p) < \infty$.*

A function $f : X \to \mathbb{R}$ is called *completely multiplicative* if for any $x, y \in X$ we have $f(xy) = f(x) \cdot f(y)$.

We recall that a function $f : X \to \mathbb{R}$ is called *multiplicative* if for any $x, y \in X$ such that the greatest common divizor of x, y is equal 1 we have $f(xy) = f(x)f(y)$. Obviously any completely multiplicative function is multiplicative but the converse is not true.

There are many very interesting multiplicative functions which are not completly multiplicative. For instance the function $x \to O(x)$ where $O(x)$ is the set of all divizor of x is multipicative but not completely multiplicative.

For any arbitrary function g on $\widetilde{P}$ there exists a unique multiplicative extension $\widetilde{g}$ of g on X. Indeed if

$$x = p_1^{x_1} p_2^{x_2} \dots p_n^{x_n}$$

with $p_i \in P$, $x_i \in N^*$, $i \neq j \Rightarrow p_i \neq p_j$ we have

$$\widetilde{g}(x) = \prod_{i=1}^{n} g(p_i^{x_i}).$$

We remark that if $g \geq 0$ then $\widetilde{g} \geq 0$.

For any arbitrary function g on P there exists a unique completely multiplicative extension $\widetilde{g}$ of g on X. Indeed if

$$x = p_1^{x_1} p_2^{x_2} \dots p_n^{x_n}$$

we put

$$\widetilde{g}(x) = (g(p_1))^{x_1} (g(p_2))^{x_2} \dots (g(p_n))^{x_n}.$$

Also if $g \geq 0$ then $\widetilde{g} \geq 0$.

Proposition 3.10 *If $t \in \mathcal{S}^*$, $t \neq 0$ is V^*-invariant and extremal then*

$$l := \sum_{p \in P} t(p) < \infty$$

and the function t/l is completely multiplicative.

Proof. From

$$t(x) = V^* t(x) = \sum_{p \in P} t_p(x)$$

it follows using the fact that t is extremal, that there exists, for each $p \in P$, $\alpha_p \in \mathbb{R}_+$ with

$$t_p = \alpha_p t.$$

Therefore

$$t(p) = \sum_{q \in P} t_p(q) = \alpha_p (\sum_{q \in P} t(q)) \quad \forall p \in P.$$

Since there exists $p \in P$ with $t(p) > 0$ we get $\alpha_p > 0$ and so $l = \sum_{q \in P} t(q) < \infty$. Now, $t' = t/l$ is also V^* invariant and extremal and moreover

$$\sum_{p \in P} t'(p) = 1, \quad t'_p = \alpha_p t'.$$

Hence for any $p \in P$ we have

$$t'(px) = t'_p(x) = \alpha_p t'(x) = t'(p) t'(x)$$

i.e. t' is completely multiplicative.

Theorem 3.11 *Let $t \in \mathcal{S}^*$ with $\sum_{p \in P} t(p) = 1$. Then t is V^*-invariant and extremal iff t is completely multiplicative.*

Proof. Using the above Proposition it is sufficient to show that any positive, completely multiplicative function t with $\sum_{p \in P} t(p) = 1$ is V^*-invariant and extremal. Indeed let t be a positive completely multiplicative function with $\sum_{p \in P} t(p) = 1$. Since

$$V^* t(x) = \sum_{p \in P} t(xp) = \sum_{p \in P} t(x) t(p) = t(x)$$

it follows that t is V^*-invariant. Suppose that t is not extremal. Then there exists $u, v \in \mathcal{S}^* | \{0\}$ such that $s = u + v$ and such that $u \wedge v = 0$ (where $\wedge$ is the infimum in $\mathcal{S}^*$ with respect to the specific order in $\mathcal{S}^*$).

If we put

$$\alpha := \sum_{p \in P} u(p)$$

we have $0 < \alpha < 1$ and

$$t = \alpha u' + (1-\alpha)v'$$

where $u' = u/\alpha$, $v' = v/1-\alpha$. For any $p \in P$ we have

$$t_p = \alpha u'_p + (1-\alpha)v'_p.$$

Since t is completely multiplicative we have

$$t_p = t(p) \cdot t.$$

Suppose that $t(p) > 0$. Then we have

$$t = \frac{\alpha}{t(p)} u'_p + \frac{1-\alpha}{t(p)} v'_p = \alpha u' + (1-\alpha)v'.$$

Since u', v' are V^*-invariant then from Proposition 3.8 we get $u'_p \preceq u'$ and so

$$\frac{u'_p}{t(p)} = u'.$$

Since $V^* u' = u'$ we deduce

$$\frac{u'(p)}{t(p)} = \frac{1}{t(p)} \sum_{q \in P} u'_p(q) = \sum_{q \in P} u'(q) = 1,$$

$$u'(p) = t(p), \quad u'_p = u'(p) \cdot u'.$$

If $t(p) = 0$ than $t_p = 0$ and therefore $u'_p = 0$ and so

$$u'_p = u'(p) u'.$$

Hence

$$u'(p) = t(p) \quad \forall p \in P,$$

$$u'_p = t(p) u' \quad \forall p \in P$$

i.e. u' is completely multiplicative and therefore

$$u' = t \text{ contradiction.}$$

Notation. For any $t \in \widetilde{\mathcal{S}}^*$ and any $x \in X$ we denote by t_x the function on X defined by

$$t_x(y) = \begin{cases} 0 & \text{if}(x,y) \neq 1 \\ t(y \cdot x) & \text{if}(x,y) = 1 \end{cases}$$

where (x, y) means the greatest common divisor of (x, y).

Proposition 3.12 *For any $t \in \widetilde{\mathcal{S}}^*$ and any $x \in X$ the function t_x belongs to $\widetilde{\mathcal{S}}^*$ and we have*

$$t_x \leq t,$$

$$\widetilde{V}^*(t_x) = (\widetilde{V}^* t)_x.$$

Particularly if t is $\widetilde{V}^*$-invariant then t_x is also $\widetilde{V}^*$-invariant.

Proof. We have

$$\widetilde{V}^*(t_x)(y) = \sum_{\substack{q \in P, \alpha \in N^* \\ q \nmid y}} t_x(yq^\alpha) =$$

$$= \begin{cases} \displaystyle\sum_{\substack{q \in P, \alpha \in N^* \\ q \nmid xy}} t(yxq^\alpha) & \text{if}(x,y) = 1 \\ 0 & \text{if}(x,y) \neq 1. \end{cases}$$

$$= \begin{cases} \displaystyle\sum_{\substack{q \in P, \alpha \in N^* \\ q \nmid xy}} t(yxq^\alpha) & \text{if}(x,y) = 1 \\ 0 & \text{if}(x,y) \neq 1. \end{cases}$$

$$= \begin{cases} (\widetilde{V}^* t)(yx) & \text{if}(x,y) = 1 \\ 0 & \text{if}(x,y) \neq 1 \end{cases} = (\widetilde{V}^* t)_x(y)$$

i.e.

$$\widetilde{V}^*(t_x) = (\widetilde{V}^* t)_x.$$

Obviously if, $x = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_n^{\alpha_n}$ we have

$$(x, y) = 1 \Rightarrow t_x(y) = t(y \cdot x) \leq (\widetilde{V}^*)^n t(y) \leq t(y)$$

$$(x, y) \neq 1 \Rightarrow t_x(y) = 0 \leq t(y)$$

and so

$$t_x \leq t$$

From

$$t_1 \leq t_2 \Rightarrow (t_1)_x \leq (t_2)_x$$

it follows

$$\widetilde{V}^*(t_x) = (\widetilde{V}^* t)_x \leq t_x$$

and so $t_x \in \mathcal{S}_{\widetilde{V}^*}$. Suppose now that t is $\widetilde{V}^*$-invariant. We get

$$\widetilde{V}^*(t_x) = (\widetilde{V}^* t)_x = t_x$$

i.e. t_x is also $\widetilde{V}^*$-invariant.

Proposition 3.13 *If $t \in \widetilde{\mathcal{S}}^*$ is $\widetilde{V}^*$-invariant and extremal then $t = 0$.*

Proof. Suppose that $t \in \mathcal{S}_{\widetilde{V}^*}$ is $\widetilde{V}^*$-invariant and suppose that $x \in X$ is such that $t(x) > 0$. From

$$t(x) = \widetilde{V}^* t(x) = \sum_{\substack{q \in P, q \nmid x \\ \alpha \in \mathbb{N}^*}} t(xq^\alpha)$$

it follows that there exists $q \in P$, $q \nmid x$ and $\alpha \in N^*$ with

$$t(xq^\alpha) > 0.$$

The function t_x is $\widetilde{V}^*$-invariant, $t_x \leq t$ and so $t_x \preceq t$. Since t is extremal there exists $\theta_x \geq 0$ with

$$t_x = \theta_x t.$$

We have

$$t_x(q^\alpha) = t(xq^\alpha) > 0$$

and so $\theta_x > 0$ which contradicts the relation

$$0 = t_x(x) = \theta_x t(x).$$

Theorem 3.14 *Any $t \in \widetilde{\mathcal{S}}^*$, $\widetilde{V}^*$-invariant is equal zero.*

Proof. Suppose firstly that

$$\sum_{\substack{q \in P \\ \beta \in N^*}} t(q^\beta) < \infty.$$

Hence for a suitable $\alpha \in \mathbb{R}_+^*$ we have $\alpha t \in \widetilde{L}$ where

$$\widetilde{L} = \{s \in \widetilde{\mathcal{S}}^* / \sum_{x \in \widetilde{P}} s(x) \leq 1\}.$$

Using Corollary 3.7 there exists a probability measure m on $\widetilde{L}$ carried by the set $\widetilde{E}$ of all extremal points of $\widetilde{L}$ such that

$$\alpha t = \int_{\widetilde{E}} sdm(s)$$

Since

$$\widetilde{V}^*(\alpha t) = \alpha t$$

we deduce

$$\alpha t = \int_{\widetilde{E}} \widetilde{V}^* sdm(s) \leq \int_{\widetilde{E}} sdm(s) = \alpha t$$

and so m is carried by the set of all extremal elements $s \in \widetilde{L}^*$ with $\widetilde{V}^* s = s$ and therefore, from Proposition 3.13, $m = 0$, $\alpha t = 0$, $t = 0$.

Suppose that t is arbitrary and let $p \in P$, $\alpha \in \mathbb{N}^*$. Since $t_{p\alpha}$ is $\widetilde{V}^*$-invariant and

$$\sum_{\substack{q \in P \\ q \neq p, \beta > 0}} t_{p\alpha}(q^\beta) = \sum_{\substack{q \in P \\ q \neq P, \beta > 0}} t(p^\alpha q^\beta) = \widetilde{V}^* t(p^\alpha) = t(p^\alpha) < \infty$$

it follows that $t_{p\alpha}$ satisfies the above condition and therefore $t_{p\alpha} = 0$ i.e.

$$t(xp^\alpha) = 0 \quad \forall x \in X \text{ with } p \nmid x$$

and so

$$t(x) = \widetilde{V}^* t(x) = \sum_{\substack{p \in P \\ \alpha \in \mathcal{S}^*, p \nmid x}} t(xp^\alpha) = 0 \quad \forall x \in X.$$

4. Martin compactification of the set of natural numbers and "prime numbers".

In this section we consider on the set X of all natural numbers $x \geq 2$ the order relation $x|y$. Let us denote by V and V^* the kernels associated with this order relation and let G and G^* the Green kernels associated with V and V^*. We denote by $\mathcal{S}$ (resp. $\mathcal{S}^*$) the convex cones of all V-supermedian (resp. V^*-supermedian) functions. As in the previous section for any $x \in X$ we denote by G_x (resp. G^*_x) the function given by

$$G_x = G1_{\{x\}}(\text{resp. } G^*_x = G^*1_{\{x\}}).$$

Further we denote by u the function on X given by

$$u(x) = \sum_{p \in P} G_p(x) = \sum_{p \in P} G^*_x(p).$$

Obviously $u \in \mathcal{S}$ and it is subtractible in $\mathcal{S}$, $u(x) > 0 \; \forall x \in X$.

We denote by L the convex subset of $\mathcal{S}^*$ given by

$$L := \{t \in \mathcal{S}^* / \sum_{p \in P} t(p) \leq 1\}.$$

Obviously for any $x \in X$ we have

$$K_x := \frac{G^*_x}{u(x)} \in L, \quad \sum_{p \in P} K_x(p) = 1$$

and moreover K_x is an extrem point of the convex set L.

From Proposition 3.6 the set L is compact with respect to the topology of simply convergence on L. On the other hand any extrem point $t \in L$, $t \neq 0$ is such that

$$\sum_{p \in P} t(p) = 1.$$

Since any element t of $\mathcal{S}^*$ is the sum of the form $t = G^*f + h$ where h is V^*-invariant it follows that any non zero extrem point of L is either K_x with $x \in X$ or it is V^*-invariant and therefore a completely multiplicative function h with $\sum_{p \in P} h(p) = 1$.

In the sequel we denote by P^* the set of all non zero extrem points of L which are completely multiplicative.

The elements of P^* are called "coprime numbers". This terminology we be justified in the sequel.

We denote by E the set of all non zero extrem points of L. Obviously we have

$$E = \{K_x / x \in X\} \cup P^*.$$

We consider on E the relation $\ll$ given by

$$t_1 \ll t_2 \overset{\text{def}}{\Longleftrightarrow} \exists \alpha > 0, t_1 \leq \alpha t_2.$$

Proposition 4.1 *The relation $\ll$ is an order relation on E such that*

1) $K_x \ll K_y \Leftrightarrow x|y, \forall x, y \in X$

2) $h_1 \ll h_2 \Leftrightarrow h_1 = h_2, \forall h_1, h_2 \in P^*$

3) $h \in P^*, x \in X \Rightarrow h \not\ll K_x$

4) $K_x \ll h \Leftrightarrow h(x) > 0 \;\forall\; x \in X, h \in P^*$.

Particularly for any $x \in X$ there exists $h \in P^*$ with $K_x \ll h$ and moreover for any $h \in P^*$ the set $\{x \in X / K_x \ll h\}$ is infinit.

Proof.

1) Let $x, y \in X$. We have, from Proposition 3.5,

$$x|y \Leftrightarrow G_x^* \leq G_y^* \Rightarrow K_x \leq \frac{u(x)}{u(y)} K_y.$$

Suppose now that there exists $\alpha > 0$ with $K_x \leq \alpha K_y$. We get

$$G_x^* \leq \alpha \frac{u(x)}{u(y)} G_y^*$$

and therefore $G_y^*(x) > 0$. From Proposition 3.5 we deduce $x|y$.

2) Let$h_1, h_2 \in P^*$ such that $h_1 \ll h_2$ i.e. there exists $\alpha > 0$ with $h_1 \leq \alpha h_2$. Since h_1, h_2 are V^*-invariant and extremal in $\mathcal{S}^*$ it follows that there exists $\theta \geq 0$

with $h_1 = \theta h_2$. From

$$\sum_{p\in P} h_1(p) = \sum_{p\in P} h_2(p) = 1$$

we deduce $\theta = 1$, $h_1 = h_2$.

3) Let $h \in P^*$, $x \in X$ and suppose that $h \ll K_x$ i.e. there exists $\alpha > 0$ with $h \leq \alpha K_x$. Since K_x is a V^*-potential then h is also a V^*-potential. Because h is V^*-invariant we deduce $h = 0$ contradiction.

4) Let $x \in X$ and $h \in P^*$. If $K_x \ll h$ then there exists $\alpha > 0$ with $K_x \leq \alpha h$ and so

$$h(x) \geq \frac{1}{\alpha} K_x(x) > 0.$$

Conversely if $h(x) > 0$ then there exists $\beta > 0$ with $u(x) \geq \beta K_x(x)$ and so K_x being a V^*-potential of the form

$$K_x = K_x(x) \cdot G^* 1_{\{x\}}$$

it follows

$$h \geq \beta K_x, \quad K_x \ll h.$$

Let now

$$x = p_1^{x_1} p_2^{x_2} \dots p_n^{x_n}.$$

with $p_i \in P$, $x_i \in \mathbb{N}^*$, $1 \leq i \leq n$ and let h be a completely multiplicative function on X with

$$h(p_i) = \frac{x_i}{\sum_{j=1}^{n} x_j} \quad 1 \leq i \leq n$$

and $h(p) = 0$ if $p \neq p_i$, $1 \leq i \leq n$.

Since $\sum_{p\in P} h(p) = 1$ it follows that $h \in P^*$. On the other hand we have

$$h(x) = \frac{\prod_{i=1}^{n} x_i^{x_i}}{\left(\sum_{i=1}^{n} x_i\right)^{\sum_{i=1}^{n} x_i}} > 0$$

i.e. $K_x \ll h$.

If $h \in P^*$ then from

$$\{x \in X/K_x \ll h\} = \{x \in X/h(x) > 0\}$$

and from

$$h(x) > 0, h(y) > 0 \Rightarrow h(xy) > 0$$

it follows that it $K_x \ll h$ then $K_x \ll h$ and so the set

$$\{x \in X/K_x \ll h\}$$

is infinit.

Remark. *Looking this Proposition it will be naturally to identily the ordered set* $(X, |)$ *with the subset*

$$\{K_x/x \in X\}$$

of E *endowed with the order relation* $\ll$*. Also in the ordered set* $(E, \ll)$*,* P^* *is the set of all maximal elements of* E *and* P *becomes the set of all minimal elements of* E*. Thus it is naturally to regard the elements of* P^**, in oposite of the elements of* P*, as "coprime numbers".*

In the sequel we give some results concerning the approximation of any element $h \in P^*$ by the elements K_x with $x \in X$ in terms of simply convergence.

For any $x \in X$, $x = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_n^{\alpha_n}$ where $p_i \in P$ and $\alpha_i \in \mathbb{N}^*$ we denote by $l(x)$ the number

$$l(x) := \sum_{i=1}^{n} \alpha_i$$

Proposition 4.2 *Let* $(x_m)_m$ *be a sequence in* X *with* $l(x_m) \to +\infty$ *such that for any* $p \in P$ *there exists*

$$\beta_p = \lim_{m \to \infty} K_{x_m}(p)$$

and $\sum_{p \in P} \beta_p = 1$*. Then the sequence* $(K_{x_n})_n$ *converges simply to an element* $h \in P^*$*.*

Proof. Suppose

$$x_m = \prod_{p \in P} p^{\alpha_p^m}, \alpha_p^m \in \mathbb{N}.$$

and let $y \in X$,

$$y = \prod_{p \in P} p^{\alpha_p} \quad \alpha_p \in \mathbb{N}.$$

We want to show that

$$\lim_{m\to\infty} K_{x_m}(y) = \prod_p (\lim_{m\to\infty} K_{x_m}(p))^{\alpha_p} = \prod_p \beta_p^{\alpha_p}.$$

For this it is sufficient to consider only two cases:

a) $y|x_m\ \forall\, m \in \mathbb{N}$; b) $y \nmid x_m\ \forall\, m \in \mathbb{N}$.

a) In this case we have, using Proposition 3.1,

$$K_{x_m}(y) = \frac{G^*_{x_m}(y)}{\sum\limits_{p\in P} G^*_{x_m}(p)} = \frac{G_y(x_m)}{\sum\limits_{p\in P} G_p(x_m)} =$$

$$= \frac{\left[\sum\limits_{p\in P} (\alpha_p^m - \alpha_p)\right]!}{\prod\limits_{p\in P}\left[(\alpha_p^m - \alpha_p)!\right]} \cdot \frac{\prod\limits_{p\in P}(\alpha_p^m!)}{\left(\sum\limits_{p\in P}\alpha_p^m\right)!} =$$

$$= \frac{\prod\limits_{p\in P}\left[\alpha_p^m(\alpha_p^m-1)\dots(\alpha_p^m-\alpha_p+1)\right]}{\left(\sum\limits_{q\in P}\alpha_q^m\right)\cdot\left(\sum\limits_{q\in P}\alpha_q^m - 1\right)\dots\left(\sum\limits_{q\in P}\alpha_q^m - \alpha_q + 1\right)} =$$

$$= \prod_{\substack{p\in P\\ p|y}} \frac{\alpha_p^m}{\sum\limits_{q\in P}\alpha_q^m} \cdot \frac{\alpha_p^m - 1}{\left(\sum\limits_{q\in}\alpha_q^m - 1\right)} \dots \frac{\alpha_p^m - \alpha_p + 1}{\sum\limits_{q\in P}\alpha_q^m - \alpha_q + 1}.$$

Since

$$K_{x_m}(p) = \frac{\alpha_p^m}{\sum\limits_{q\in}\alpha_q^m}$$

and since

$$l(x_m) = \sum_{q\in P}\alpha_q^m \to +\infty$$

it follows that

$$K_{x_m}(y) \to \prod_{p \in P} \beta_p^{\alpha_p} = (\lim_{m \to \infty} K_{x_m}(p))^{\alpha_p}.$$

b) In this case it is sufficient to consider only the situation in which there exists $p \in P$, $p|y$ and $p \nmid x_m \forall m \in \mathbb{N}$. In this case we have

$$y \nmid x_m \Rightarrow K_{x_m}(y) = 0$$

$$p \nmid x_m \Rightarrow K_{x_m}(p) \Rightarrow \beta_p = 0$$

and so

$$K_{x_m}(y) = 0 = \prod_{q \in P} \beta_q^{\alpha_q} \quad \forall m \in \mathbb{N}.$$

Proposition 4.3. For any $h \in P^*$ there exists an increasing sequence $(x_m)_m$ in $(X, |)$ such that $(K_{x_m})_m$ converges simply to h,

$$K_{x_m} \ll h$$

and

$$l(x_m) \to \infty.$$

Proof. Firstly suppose that

$$A := \{p \in P / h(p) \neq 0\}$$

is finite and that $h(p) \in Q$ for all $p \in P$. Then there a function

$$P \ni p \to \alpha_p \in \mathbb{N}$$

such that

$$h(p) = \frac{\alpha_p}{\sum_{q \in P} \alpha_q} \quad \forall p \in \mathbb{N}.$$

We consider

$$x := \prod_{p \in P} p^{\alpha_p}$$

and the sequence $(x_m)_m$ given by

$$x_m = x^m.$$

We have $l(x_m) \to +\infty$ and from

$$K_{x_m}(p) = K_x(p) = \frac{\alpha_p}{\sum_{q\in P} \alpha_q}$$

it follows that

$$\lim_{m\infty} K_{x_m}(p) = \frac{\alpha_p}{\sum_{q\in P} \alpha_q}.$$

Using the above proposition it follows that the sequence $(K_{x_m})_m$ converges simple to an element $h' \in P^*$ such that

$$h'(p) = \frac{\alpha_p}{\sum\limits_{q\notin P} \alpha_q} = h(p).$$

Since h, h' are completely multiplicative we get $h' = h$.

Suppose now that $h \in P^*$ is such that

$$A := \{p \in P/h(p) > 0\}$$

is finite. Let now $(h_m)_m$ be a sequence in P^* such that

$$A = \{p \in P/h_m(p) > 0\}$$

and moreover for any $p \in A$ we have $h_m(p) \in Q_x$, $h_m(p) - \dfrac{1}{m} \leq h(p) \leq h_m(p) + \dfrac{1}{m}$ $\forall\, p \in A$. Obviously $(h_m)_m$ converges simple to h.

From this fact and from the first step of the proof there exists an increasing sequence $(x_m)_n$ in $(X, |)$ such that $(K_{x_m})_m$ converges simply to h and

$$K_{x_m} \ll h \quad \forall m \in \mathbb{N},$$

$$l(x_m) \to +\infty.$$

Suppose now $h \in P^*$ such that

$$A := \{p \in P/h(p) > \circ\}$$

is infinit. We consider an increasing sequence $(F_n)_n$ of finite subsets of A and

$$\sum_{p \notin F_n} h(p) < \frac{1}{n}$$

For any $n \in \mathbb{N}$ let $h_n \in P^*$ such that

$$\{p \in P/h_n(p) > 0\} = F_n$$

and such that

$$h_n(p) = \frac{h(p)}{\sum_{q \in F_n} h(q)} \ \forall\, p \in F_n.$$

Obviously $(h_n)_n$ is simply convergent to h. Using the preceeding consideration there exists an increasing sequence $(x_n)_n$ in X such that

$$K_{x_n} \ll h_n, \quad l(x_n) \geq n$$

and

$$|K_{x_n} - h_n| < \frac{1}{n} \text{on} \{y \in X/l(y) \leq n\}$$

Obviously

$$K_{x_n} \ll h, \quad l(x_n) \to +\infty$$

and $(K_{x_n})_n$ converges simply to h.

References

[1] N. Boboc, Gh. Bucur, A. Cornea: Order and Convexity in Potential Theory; H - cones. Lecture Notes in Math. Springer Verlag 1981.

[2] N. Boboc, Gh. Bucur: Potential theory on finite ordered set. Rev. Roumain Math. Pures Appl., 38,6 (1993) 459 – 480.

[3] N. Boboc, Gh. Bucur: Potential theory on ordered sets. Totally increasing functions. Rev. Roumain. Math Pures at Appl.43, 3 – 4(1998) 277 – 298.

[4] N. Boboc, Gh. Bucur: Potential Theory on ordered sets. Exact totally increasing functions. Rev. Roumain Math. Pures at Appl. 43, 7 – 8 (1998) 685 – 720.

[5] N. Boboc, Gh. Bucur: Aspects of Number Theory in Terms of Potential Theory. Potential Analysis 14, (2001) , 31 – 51.

[6] Eckford Cohen: Arithmetic functions associated with the unitary divizors of an integer. Math. Zeisch 74, (1960) 66 – 80.

[7] P.A. Meyer: Probability and Potentials. Blaisdel Publishing Company, Waltham Massachussetts – Toronto – London – 1966.

LARGE DEVIATIONS VIA TRANSFERENCE PLANS *

Sergey G. Bobkov †

1 Introduction

Let P be a log-concave probability measure on $\mathbf{R}^n$. Equivalently, P is concentrated on some affine subspace $E \subset \mathbf{R}^n$ where it has a density p, with respect to Lebesgue measure on E, such that

$$p(tx + (1-t)y) \geq p(x)^t p(y)^{1-t}, \quad \text{for all } t \in (0,1) \text{ and } x, y \in E.$$

We refer the reader to the classical 1974 paper by C. Borell [Bor] for a general theory of such measures.

As is known, given a function f on $\mathbf{R}^n$, certain possible distributional properties of f with respect to P can be controled by the behavior of this function along lines. For example, when $f(x) = \|x\|$ is an arbitrary norm, we have an inequality for large deviations

$$P\{f > \lambda \mathbf{E} f\} \leq C e^{-c\lambda}, \quad \lambda \geq 0, \tag{1.1}$$

where C and c are positive numerical constants, and where we use probabilistic notations $\mathbf{E} f = \int f \, dP$ for the expectation with respect to P. In an equivalent form this fact first appeared in [Bor], cf. Lemma 3.1. If f is a polynomial in n real variables of degree d, we have a similar inequality

$$P\{|f| > \lambda \mathbf{E}|f|\} \leq C(d)\, e^{-c(d)\, \lambda^{r(d)}}, \quad \lambda \geq 0, \tag{1.2}$$

thus with the right hand side depending on d, but independent of the measure P. This observation, which gave an affirmative answer to a conjecture of V. D. Milman, is due to J. Bourgain [Bou] who considered for P the uniform distribution on an arbitrary convex body in $\mathbf{R}^n$.

Both (1.1) and (1.2) can be united by a more general scheme. With every continuous function f on $\mathbf{R}^n$ and $\varepsilon \in (0,1)$, we associate the quantity

$$\delta_f(\varepsilon) = \sup_{x_0, x_1 \in \mathbf{R}^n} \mathrm{mes}\{t \in (0,1) : |\, f(tx_0 + (1-t)x_1)\,| < \varepsilon |f(x_0)|\}.$$

*Key words: Large deviations, Khinchine-type inequalities, transportation of mass
†Supported in part by an NSF grant

As turns out, the behavior of δ_f near zero is connected with large deviations of f, and moreover, the corresponding inequalities can be made independent of P. To study the polynomial case, J. Bourgain established the property

$$\delta_f(\varepsilon_0) \leq \delta_0 \tag{1.3}$$

with $\delta_0 = 1/2$ and with some $\varepsilon_0 \in (0,1)$ depending upon d. This was already enough to derive a very general statement on large deviations in the form (1.2). However, the altitude of $\delta_f(\varepsilon)$ for small ε's may contain an additional information on the strength of deviations. In this note, we refine and extend Bourgain's approach to arbitrary functions f and log-concave measures P, with resulting estimates depending upon δ_f, only. In particular, we prove:

Theorem 1.1. *Let P be a log-concave probability measure on $\mathbf{R}^n$, and let f be a continuous function on $\mathbf{R}^n$. Then, for all $\lambda > 2e$ such that $\delta_f(2e/\lambda) \leq 1/2$,*

$$P\{|f| > \lambda\, \mathbf{E}|f|\} \leq \exp\left\{-\frac{1}{2\delta_f(2e/\lambda)}\right\}. \tag{1.4}$$

Once $\delta_f(\varepsilon) \to 0$, as $\varepsilon \to 0$, the assumption $\delta_f(2e/\lambda) \leq 1/2$ is fulfilled for all λ large enough. In case $\delta_f(\varepsilon) \leq C\varepsilon^r$, for all $\varepsilon \in (0,1)$ and some $C \geq 1$, $r > 0$, we thus arrive at the estimate of the form

$$P\{|f| > \lambda\, \mathbf{E}|f|\} \leq c_1 e^{-c_2 \lambda^r}, \quad \lambda \geq 0.$$

As an example, we will observe that $\delta_f(\varepsilon) = \frac{2\varepsilon}{1+\varepsilon}$ for any norm $f(x) = \|x\|$ on $\mathbf{R}^n$, and we are thus lead to (1.1). In the polynomial case, $\delta_f(\varepsilon) = O(\varepsilon^{1/d})$ that leads to (1.2) with the correct power $r(d) = 1/d$.

The main Bourgain argument based on the existence of suitable measure-preserving maps is described in section 2. In the next section 3, we study large deviations under the condition 1.3. The latter turns out to be related to a property known as Markov's inequality for polynomials in one real variable. In section 4, we consider Theorem 1.1 itself and show how it can be applied to norms and polynomials, by computing or estimating the quantity $\delta_f(\varepsilon)$. In section 5, we apply Theorem 1.1 to study deviations of convex functions f from their mean $\mathbf{E}f$. In particular, we will consider the case of the euclidean norm for which it is possible to reach exponentially decreasing tails in terms of P-variance. At the end of this note, we put an appendix devoted to triangular measure-preserving maps.

2 Bourgain's argument

As a first basic step, we prove:

Theorem 2.1. *Let f be a continuous function on $\mathbf{R}^n$, and let P be a log-concave probability measure on $\mathbf{R}^n$. Let $\varepsilon \in (0,1)$ and $\delta = \delta_f(\varepsilon) < 1$. Then, for all $\lambda \geq 0$ and $\gamma \in (0, 1-\delta]$,*

$$P\{|f| > \lambda\varepsilon\} \geq \gamma P\{|f| > \lambda\}^{\delta/(1-\gamma)}. \tag{2.1}$$

Proof. Without loss of generality, we may assume that P is absolutely continuous with respect to Lebesgue measure on $\mathbf{R}^n$. Thus, P is concentrated on an open convex set $A_0 \subset \mathbf{R}^n$ where it has a positive, log-concave density $p(x)$. We may also assume that $0 < \lambda < \operatorname{ess\,sup} f$. For such values, introduce a family of non-empty open subsets of A_0,

$$A_\lambda = \{x \in A_0 : |f(x)| > \lambda\}.$$

Fix λ and take a regular subset A of A_λ (e.g., a finite union of open balls, cf. Appendix for details). In particular, $P(A) > 0$. We follow an argument of J. Bourgain [Bou]: There exists unique continuous bijective triangular map $T : A \to A_0$ which pushes forward the normalized restriction P_A of P to A to the measure P. Moreover, the components $T_i = T_i(x_1, \dots, x_i)$ of T, $i = 1, \dots, n$, are C^1-smooth with respect to x_i-coordinates and satisfy $\frac{\partial T_i}{\partial x_i} > 0$ so that the Jacobian

$$J(x) = \prod_{i=1}^n \frac{\partial T_i(x)}{\partial x_i}, \quad x \in A,$$

is continuous and positive on A. Since P_A has density $p_A(x) = \frac{p(x)}{P(A)}$, $x \in A$, the property that T pushes forward P_A to P is equivalent to saying that

$$\frac{p(x)}{P(A)} = p(T(x))\, J(x), \quad x \in A. \tag{2.2}$$

Now, for each $t \in (0,1)$, introduce another map,

$$T_t(x) = tx + (1-t)T(x), \quad x \in A,$$

which is also continuous, triangular, with components that are C^1-smooth with respect to x_i-coordinates. Moreover, its Jacobian J_t satisfies

$$J_t(x) = \prod_{i=1}^n \left(t + (1-t)\, \frac{\partial T_i(x)}{\partial x_i}\right) \geq J(x)^{1-t}, \quad x \in A. \tag{2.3}$$

Consider the set

$$B_t = \{x \in A : |f(tx + (1-t)T(x))| > \lambda\varepsilon\}$$

and its image $B'_t = T_t(B_t)$. Clearly, if $y \in B'_t$, then $y = tx + (1-t)T(x)$, for some $x \in B_t$, hence $|f(y)| > \lambda\varepsilon$, that is, $y \in A_{\lambda\varepsilon}$. This means that $B'_t \subset A_{\lambda\varepsilon}$, and therefore

$$P(B'_t) \leq P(A_{\lambda\varepsilon}). \tag{2.4}$$

On the other hand, using the log-concavity $p(tx+(1-t)x') \geq p(x)^t p(x')^{1-t}$ (which will be needed with $x' = T(x)$), and applying (2.2)-(2.3), we get

$$\begin{aligned} P(B_t') &= \int_{T_t(B_t)} p(y)\,dy = \int_{B_t} p(T_t(x))\,J_t(x)\,dx \\ &\geq \int_{B_t} p(T_t(x))\,J(x)^{1-t}\,dx \ \geq \ \int_{B_t} p(x)^t\,p(T(x))^{1-t}\,J(x)^{1-t}\,dx \\ &= P(A)^{t-1} \int_{B_t} p(x)\,dx \ = \ P(A)^{t-1}\,P(B_t). \end{aligned}$$

Together with (2.4), this yields

$$P(A_{\lambda\varepsilon}) \geq P(A)^{t-1}\,P(B_t). \tag{2.5}$$

Now, in order to further estimate from below the last term in (2.5), it is the time to involve the function δ_f. By the definition, for any $x \in A$,

$$\mathrm{mes}\{t \in (0,1) : |f(tx+(1-t)T(x)| < \varepsilon |f(x)|\} \leq \delta.$$

Since $A \subset A_\lambda$, we have $|f(x)| > \lambda$, so,

$$\mathrm{mes}\{t \in (0,1) : |f(tx+(1-t)T(x)| \leq \varepsilon\lambda\} \leq \delta,$$

or equivalently,

$$\int_0^1 1_{\{|f(tx+(1-t)T(x)|>\lambda\varepsilon\}}\,dt \geq 1-\delta.$$

Integrating this inequality over the measure P_A and interchanging the integrals, we get

$$\int_0^1 P_A(B_t)\,dt \geq 1-\delta. \tag{2.6}$$

Thus, the function $\psi(t) = P_A(B_t)$ being bounded by 1 satisfies $\int_0^1 \psi(t)\,dt \geq 1-\delta$. This actually implies that $\psi(t) \geq \gamma$, for some $t \in (0, t_0]$ where $t_0 = \frac{\delta}{1-\gamma} \in (0,1]$. Indeed, assuming that $\psi(t) < \gamma$, whenever $t \in (0, t_0]$, we would get

$$\int_0^1 \psi(t)\,dt = \int_0^{t_0} \psi(t)\,dt + \int_{t_0}^1 \psi(t)\,dt < \gamma t_0 + (1-t_0) = 1-(1-\gamma)t_0 = 1-\delta.$$

Thus, $\int_0^1 \psi(t)\,dt < 1-\delta$ that contradicts to (2.6). We can therefore conclude that

$$\frac{P(B_t)}{P(A)} = P_A(B_t) \geq \gamma, \quad \text{for some} \quad t \in (0, t_0].$$

Applying this in (2.5), we arrive at $P(A_{\lambda\varepsilon}) \geq \gamma P(A)^t$, and since $t \leq t_0$,

$$P(A_{\lambda\varepsilon}) \geq \gamma P(A)^{t_0}. \tag{2.7}$$

At last, approximating from below the set A_λ by regular subsets A so that $P(A) \uparrow P(A_\lambda)$, we get from (2.7) in the limit $P(A_{\lambda\varepsilon}) \geq \gamma P(A_\lambda)^{t_0}$, that is, exactly (2.1).

Theorem 2.1 is proved.

Remark 2.2. The above argument still works with many other measure preserving maps. For example, one may take for T the Brenier map, i.e., of the form $T = \nabla\varphi$, for some mod(P)-uniquely defined convex function φ, cf. [Bre] and [M]. In this case, the derivative $T'(x)$ represents a positively definite matrix, and the crucial inequality (2.3) should be replaced with

$$\det(t\,\mathrm{Id} + (1-t)\,T'(x)) \geq \det^{1-t}(T'(x))$$

which is a particular (and log-concave) case of the Brunn-Minkowski-type inequality for determinants $\det^{1/n}(A+B) \geq \det^{1/n}(A) + \det^{1/n}(B)$ in the class of all positively definite $n \times n$-matrices. However, to make the argument following (2.3) absolutely rigorous (the change of the variable formula), it is desirable to require that the map T_t be in a certain sense regular. C^1-smoothness seems too strong requirement, but specializing in triangular maps, it is enough to have C^1-smoothness of the components T_i of T with respect to i-th coordinates. We provide more details in appendix.

Remark 2.3. An attempt to choose an optimal γ in (2.1) complicates this inequality, but in essense does not give an improvement. For further applications, at the sake of some loss in constants, one may use Theorem 2.1 with $\gamma = \frac{1}{2}$, for example.

It is however interesting to know how sharp the inequality (2.1) is. The definition of δ_f reflects the behavior of the function f along all lines, so one may try to derive inequalities of this kind by appealing to the localization technique going back to the papers by M. Gromov and V. D. Milman, cf. [G-M], [A], and L. Lovász and M. Simonovits [L-S], cf. also [K-L-S]. The advantage of this approach is that it allows one to reduce many problems to dimension one where it is much easier to explore extremal situations. In a recent preprint [N-S-V], F. Nazarov, M. Sodin, and A. Volberg employ the localization ideas to prove the following remarkable statement which they call the geometric KLS lemma in the spirit of [K-L-S]: Given a compact convex set K in $\mathbf{R}^n$, its closed subset F, and a number $\alpha > 1$, define

$$F_\alpha = \left\{ x \in K : \text{for every interval } J \text{ such that } x \in J \subset K,\ \frac{|F \cap J|}{|J|} \geq \frac{\alpha - 1}{\alpha} \right\}.$$

Then, if $\mathrm{vol}_n(F) > 0$,

$$\frac{\mathrm{vol}_n(F_\alpha)}{\mathrm{vol}_n(K)} \leq \left(\frac{\mathrm{vol}_n(F)}{\mathrm{vol}_n(K)} \right)^\alpha.$$

It is noted in [N-S-V] that in the definition of F_α it is enough to consider only the intervals J that have x as one of their endpoints. Moreover, the above inequality extends to arbitrary log-concave probability measures P in the form

$$P(F_\alpha) \leq P(F)^\alpha. \tag{2.8}$$

To see a connection with (2.1), take $F = \{x : |f(x)| \geq \lambda\varepsilon\}$, $G = \{x : |f(x)| \geq \lambda\}$, and assume that $\delta = \delta_f(\varepsilon) < 1$. Then, by the very definition of δ_f, we have $G \subset F_\alpha$ for $\alpha = 1/\delta$, so (2.8) turns into

$$P\{|f| \geq \lambda\varepsilon\} \geq P\{|f| \geq \lambda\}^\delta. \tag{2.9}$$

This is an improved and more correct version of (2.1): the factor γ can thus be replaced with 1 while the power $\delta/(1-\gamma)$ can be replaced with δ. The inequality (2.9) is sharp already in some special sitiuations. For example, for an arbitrary norm $f(x) = \|x\|$ on $\mathbf{R}^n$, we have $\delta_f(\varepsilon) = \frac{2\varepsilon}{1+\varepsilon}$, $\varepsilon \in (0,1)$, so $P\{|f| \geq \lambda\varepsilon\} \geq P\{|f| \geq \lambda\}^{2\varepsilon/(1+\varepsilon)}$ which is another version of the ineqiality

$$1 - P\left(\frac{1}{\varepsilon}A\right) \leq (1 - P(A))^{2\varepsilon/(1+\varepsilon)} \tag{2.10}$$

for the class of all centrally symmetric convex sets A in $\mathbf{R}^n$. The latter was proved using a localization lemma by L. Lovász and M. Simonovits [L-S] for euclidean balls A and later extended by O. Guédon [G] to the general case. He also observed that equality in (2.10) is attained in dimension one for any interval $A = (-a, a)$, $a > 0$, at the (non-symmetric) exponential measure P with $P(x, +\infty) = e^{-(x+a)}$, $x > -a$. We do not know whether the argument based on the transference plans can appropriately be modified to reach the sharp forms (2.9)-(2.10).

Remark 2.4. It follows from (2.1) by letting $\lambda \downarrow 0$ that $P\{f = 0\} = 0$, if $f \neq 0$ mod(P) and $\delta_f(\varepsilon) \to 0$, as $\varepsilon \downarrow 0$. Note also that in Theorem 2.1 one may assume that f is defined on A_0 (rather than on the whole space), and restrict the points x_0 and x_1 in the definition of δ_f to the set A_0.

3 Iteration procedure. Markov's classes

In order to apply the inequality (2.1), the weakest assumption which should be required from f is the property $\delta_f(\varepsilon) \neq 0$ (identically), that is,

$$\delta_f(\varepsilon_0) \leq \delta_0 \tag{3.1}$$

for some $\varepsilon_0 \in (0,1)$ and $\delta_0 \in (0,1)$. As shown in [Bou], already in this situation one can recover exponentially decreasing tails for f by iterating the inequality (2.1). Namely, we have:

Theorem 3.1. *Under the condition* (3.1), *there exist positive numbers C, c, r, depending on $(\varepsilon_0, \delta_0)$, only, such that, for all $\lambda \geq 0$,*

$$P\{|f| > \lambda \mathbf{E}|f|\} \leq C\, e^{-c\lambda^r}. \tag{3.2}$$

The power r appearing in (3.2) can be chosen as close to the number $r_0 = \frac{\log(1/\delta_0)}{\log(1/\varepsilon_0)}$, as we wish. In particular, f has finite moments $\mathbf{E}|f|^q$ of any order q, and, moreover, for all $0 < r < r_0$,

$$\mathbf{E} \exp\{|f|^r\} = \int \exp\{|f|^r\}\, dP < +\infty.$$

In addition, f satisfies Khinchine-type inequalities

$$(\mathbf{E}|f|^q)^{1/q} \leq C\, \mathbf{E}|f|, \qquad C = C(q, \varepsilon_0, \delta_0), \quad q \geq 1. \tag{3.3}$$

It would therefore be interesting to explore the class of all functions f possessing the property (3.1). One sufficient condition was suggested by Yu. V. Prokhorov in his study of Khinchine-type inequalities for polynomials over Gaussian and Γ-distributions on the real line, cf. [Pr1], [Pr2]. Prokhorov's proof of (3.3) is based on Markov's inequality,

$$\max_{0 \leq t \leq 1} |Q'(t)| \leq \kappa \max_{0 \leq t \leq 1} |Q(t)|, \tag{3.4}$$

which holds true for any polynomial Q in real variable t of degree d with (optimal) constant $\kappa = 2d^2$. Let us then say that a given function f on $\mathbf{R}^n$ belongs to the (Markov) class $M(\kappa)$ with constant $\kappa \geq 1$, if, for all vectors $x_0, x_1 \in \mathbf{R}^n$, the function $Q(t) = f(tx_0 + (1-t)x_1)$ is absolutely continuous on $\mathbf{R}$ and has a Radon-Nikodym derivative Q' satisfying the inequality (3.4). With this definition, we have:

Proposition 3.2. *Every function f in $M(\kappa)$ satisfies* (3.1) *with*

$$\varepsilon_0 = \frac{1}{2}, \quad \delta_0 = 1 - \frac{1}{2\kappa}.$$

Indeed, following an argument of [Pr1-2], let t_0 be a point of maximum of $|Q(t)|$ on [0,1], and assume for definiteness that $Q(t_0) > 0$. Then, by (3.4), for all $t \in [0,1]$,

$$Q(t) \geq Q(t_0)(1 - \kappa|t - t_0|) \geq \frac{1}{2} Q(t_0),$$

where the second inequality holds true in a smaller subinterval $|t - t_0| \leq 1/(2\kappa)$, $0 \leq t \leq 1$. This interval has length at least $1/(2\kappa)$, so

$$\begin{aligned} \mathrm{mes}\{|\, f(tx_0 + (1-t)x_1)\,| < \varepsilon |f(x_0)|\} \;&=\; \mathrm{mes}\Big\{t \in (0,1) : |Q(t)| < \frac{1}{2}|Q(1)|\Big\} \\ &\leq\; \mathrm{mes}\Big\{t \in (0,1) : |Q(t)| < \frac{1}{2}|Q(t_0)|\Big\} \leq 1 - \frac{1}{2\kappa}. \end{aligned}$$

Thus, according to Theorem 3.1 and Proposition 3.2, any function f in $M(\kappa)$ shares the large deviation inequality (3.2) and the Khinchine-type inequality (3.3). Moreover, for large values of κ, the critical value r_0 in (3.2) is of order at most C/κ.

According to Markov's inequality, any polynomial f on $\mathbf{R}^n$ of degree d belongs to the class $M(2d^2)$. Another important example: any norm $f(x) = \|x\|$ belongs to the class $M(2)$. Indeed, the function $Q(t) = f(tx_0 + (1-t)x_1)$ is convex and satisfies, by the triangle inequality, $Q(t) \geq Q_0(t) \equiv |\,t\|x_0\| - (1-t)\|x_1\|\,|$. Since $Q_0(0) = Q(0)$, $Q_0(1) = Q(1)$, we conclude that

$$\max_{0\leq t\leq 1} |Q'(t)| \leq \max_{0\leq t\leq 1} |Q_0'(t)| = \|x_0\| + \|x_1\| \leq 2 \max_{0\leq t\leq 1} |Q(t)|.$$

Proof of Theorem 3.1. Assume for definiteness that $f \neq 0 \mod(P)$ and write the inequality (2.1) with $\varepsilon = \varepsilon_0$, $\delta = \delta_0$ and an arbitrary fixed $\gamma \in (0, 1-\delta_0)$ as

$$P\{|f| > \lambda\} \leq \alpha^\beta \, P\{|f| > \lambda\varepsilon_0\}^\beta, \quad \lambda \geq 0, \tag{3.5}$$

where $\alpha = \frac{1}{\gamma}$, $\beta = \frac{1-\gamma}{\delta_0}$. Thus, $\alpha > 1$ and $\beta > 1$. In case $\lambda = 0$, we get in particular that

$$P\{|f| > 0\} \geq \alpha^{-\frac{\beta}{\beta-1}} \tag{3.6}$$

Now, applying (3.5) to $\lambda\varepsilon_0$, we get $P\{|f| > \lambda\} \leq \alpha^{\beta+\beta^2} \, P\{|f| > \lambda\varepsilon_0^2\}^{\beta^2}$. Similarly, on the k-th step, we will have

$$P\{|f| > \lambda\} \leq \alpha^{\beta+\ldots+\beta^k} \, P\{|f| > \lambda\varepsilon_0^k\}^{\beta^k}.$$

Using $\beta + \ldots + \beta^k \leq \frac{\beta}{\beta-1}\, \beta^k$, we obtain a simpler estimate

$$P\{|f| > \lambda\} \leq \left(\alpha^{\frac{\beta}{\beta-1}} \, P\{|f| > \lambda\varepsilon_0^k\}\right)^{\beta^k}. \tag{3.7}$$

Now denote by m a quantile of $|f|$ of order $e^{-1}\,\alpha^{-\frac{\beta}{\beta-1}}$, that is, any number such that

$$P\{|f| > m\} \leq \frac{1}{e\,\alpha^{\frac{\beta}{\beta-1}}}, \quad P\{|f| < m\} \leq \frac{1}{e\,\alpha^{\frac{\beta}{\beta-1}}} \tag{3.8}$$

By (3.6), such a number m must be positive. Furthermore, the inequality (3.7) with $\lambda\varepsilon_0^k = m$ yields, for all $k = 1, 2\ldots,$

$$P\left\{\frac{|f|}{m} > \varepsilon_0^{-k}\right\} \leq \exp\left\{-\beta^k\right\}. \tag{3.9}$$

Now take any $x \geq 1/\varepsilon_0$ and pick up a natural namber k such that $\varepsilon_0^{-k} \leq x < \varepsilon_0^{-(k+1)}$. Then, $k \geq \frac{\log x}{\log(1/\varepsilon_0)} - 1$, so, $\beta^k \geq \frac{1}{\beta} x^{\log\beta/\log(1/\varepsilon_0)}$. Since $P\left\{\frac{|f|}{m} > x\right\} \leq P\left\{\frac{|f|}{m} > \varepsilon_0^{-k}\right\}$, we derive from (3.9)

$$P\left\{\frac{|f|}{m} > x\right\} \leq \exp\left\{-\frac{1}{\beta}\, x^{\frac{\log\beta}{\log(1/\varepsilon_0)}}\right\}, \quad x \geq \frac{1}{\varepsilon_0}. \tag{3.10}$$

The power $r = \frac{\log\beta}{\log(1/\varepsilon_0)}$ in (3.10) is less than $r_0 = \frac{\log(1/\delta_0)}{\log(1/\varepsilon_0)}$ but it can be made close to this number by choosing small values of γ.

Now, in order to replace the quantile with the mean $\mathbf{E}|f|$, we may use the inequality $\mathbf{E}|f| \geq \frac{1}{e\,\alpha^{\frac{\beta}{\beta-1}}}\,m$, so (3.10) yields

$$P\left\{|f| > e\,\alpha^{\frac{\beta}{\beta-1}}\,x\,\mathbf{E}|f|\right\} \leq \exp\left\{-\frac{1}{\beta}\,x^{\frac{\log\beta}{\log(1/\varepsilon_0)}}\right\}, \quad x \geq \frac{1}{\varepsilon_0}.$$

Making the change $\lambda = e\,\alpha^{\frac{\beta}{\beta-1}}\,x$, we get the desired inequality $P\{|f| > \lambda\,\mathbf{E}|f|\} \leq e^{-c\,\lambda^r}$, $\lambda \geq \lambda_0$, with an arbitrarily chosen $\gamma \in (0, 1-\delta_0)$, and

$$r = \frac{\log\beta}{\log(1/\varepsilon_0)}, \quad c = \frac{1}{\beta\left(e\,\alpha^{\frac{\beta}{\beta-1}}\right)^{\frac{\log\beta}{\log(1/\varepsilon_0)}}}, \quad \lambda_0 = \frac{e\,\alpha^{\frac{\beta}{\beta-1}}}{\varepsilon_0}.$$

4 Theorem 1.1. Norms and polynomials

To derive the inequality of Theorem 1.1 from Theorem 2.1, assume f is normalized so that $\mathbf{E}|f| = 1$. By Chebyshev's inequaity, $P\{|f| \geq x\} \leq 1/x$, for all $x > 0$. If $\delta_f(\varepsilon) \leq 1/2$, we can take in (2.1) $\gamma = 1/2$ which leads to

$$P\{|f| > \lambda\} \leq (2\,P\{|f| > \lambda\varepsilon\})^{1/(2\delta_f(\varepsilon))} \leq \left(\frac{2}{\lambda\varepsilon}\right)^{1/(2\delta_f(\varepsilon))}, \quad \lambda \geq 0.$$

Choosing if possible $\varepsilon = 2e/\lambda$, we then arrive at the estimate (1.4), that is,

$$P\{|f| > \lambda\,\mathbf{E}|f|\} \leq \exp\left\{-\frac{1}{2\delta_f(2e/\lambda)}\right\}, \qquad \lambda > 2e, \ \ \delta_f(2e/\lambda) \leq 1/2. \tag{4.1}$$

The above inequality immediately implies:

Corollary 4.1. *If $\delta_f(\varepsilon) \leq C\varepsilon^r$, for all $\varepsilon \in (0,1)$ and some $C \geq 1$, $r > 0$, then*

$$P\left\{\frac{1}{2e}\,|f| \geq \lambda\,\mathbf{E}|f|\right\} \leq \exp\left\{-\frac{\lambda^r}{2C}\right\}, \quad \lambda^r \geq 2C. \tag{4.2}$$

As we see, the inequalities (4.1)-(4.2) may contain more precise information in comparison with the general Markov classes $M(\kappa)$. This concerns in particular such functions f as norms and polynomials for which it would be interesting to explore the behavior of δ_f near zero. We start with an arbitrary norm $f(x) = \|x\|$ on $\mathbf{R}^n$.

Proposition 4.2. *For any norm f, we always have $\delta_f(\varepsilon) = \frac{2\varepsilon}{1+\varepsilon}$, $\varepsilon \in (0,1)$.*

Proof. By the triangle inequality, for all $x_0, x_1 \in \mathbf{R}^n$, and $t \in (0,1)$,

$$\|tx_0 + (1-t)x_1\| \geq |\, t\|x_0\| - (1-t)\|x_1\|\,|.$$

Hence,

$$\mathrm{mes}\{t \in (0,1) : \|tx_0 + (1-t)x_1\| < \varepsilon\|x_0\|\} \leq$$
$$\mathrm{mes}\{t \in (0,1) : |\, t\|x_0\| - (1-t)\|x_1\|\,| < \varepsilon\|x_0\|\}.$$

Putting $a = \|x_0\|$, $b = \|x_1\|$, $c = \frac{a}{a+b} \leq 1$, and assuming $a > 0$, we obtain

$$\begin{aligned}
\mathrm{mes}\{t : |\, ta - (1-t)b\,| < \varepsilon a\} &= \mathrm{mes}\{t : -\varepsilon a < ta - (1-t)b < \varepsilon a\} \\
&= \mathrm{mes}\{t : (1-\varepsilon)\, a < t\,(a+b) < (1+\varepsilon)\, a\} \\
&= \mathrm{mes}\{t : (1-\varepsilon)\, c < t < (1+\varepsilon)\, c\} \\
&= \min\{1, (1+\varepsilon)\, c\} - (1-\varepsilon)\, c.
\end{aligned}$$

The last quantity is maximized in $0 \leq c \leq 1$ at $c = \frac{1}{1+\varepsilon}$ which gives $\frac{2\varepsilon}{1+\varepsilon}$. The optimality of this upper bound can be seen in case $x_1 = -\varepsilon x_0$, $\|x_0\| = 1$. Proposition 4.2 is proved.

Thus, the assumption made in Corollary 4.1 is fulfilled for the norm-function with $r = 1$ and $C = 2$. Hence, by (4.2), for all $\lambda \geq 8e$,

$$P\,\{|f| > \lambda\, \mathbf{E}|f|\} \leq e^{-\lambda/(16e)}.$$

The numerical constants are certainly not optimal and can be improved by virtue of (2.10).

Now let f be an arbitrary polynomial of degree at most $d \geq 1$. In this case, the maximal possible value of $\delta_f(\varepsilon)$ is completely determined in dimension one, so assume $n = 1$. In [Bou] it was shown that, for some numerical $c_0 \in (0,1)$,

$$\mathrm{mes}\left\{t \in (0,1) : |f(t)| < c_0^d\, \|f\|_{L^\infty(0,1)}\right\} \leq \frac{1}{2}.$$

Thus, we always have $\delta_f(c_0^d) \leq \frac{1}{2}$ which complements Proposition 3.2 in the polynomial case, namely, $\delta_f(1/2) \leq 1 - \frac{1}{4d^2}$. As for small values of ε, we have:

Proposition 4.3. *For any polynomial f of degree at most $d \geq 1$, for all $\varepsilon \in (0,1)$,*

1) $\delta_f(\varepsilon) \leq 2d\, \varepsilon^{1/d}$; 2) $\delta_f(\varepsilon) \leq 2\, \varepsilon^{1/d} \log \frac{1}{\varepsilon^{1/d}}$.

Proof. Let $f(t) = \prod_{i=1}^d (t - z_i)$ with $z_i \in \mathbf{C}$, $1 \leq i \leq d$. Then, on the interval (0,1),

$$\mathrm{mes}\{t \in (0,1) : |f(t)| < \varepsilon\, |f(0)|\} \quad \leq \quad \mathrm{mes} \bigcup_{i=1}^d \{|t - z_i| < \varepsilon^{1/d}\, |z_i|\}$$

$$\leq \sum_{i=1}^{d} \mathrm{mes}\{|t - z_i| < \varepsilon^{1/d}\,|z_i|\}.$$

Since $|t - z_i| \geq |t - |z_i||$, the roots z_i may be assumed to be real non-negative numbers. But, for any $c \in (0,1)$ and $z > 0$, the quantity $\mathrm{mes}\{t \in (0,1) : |t - z| < cz\}$ is maximized at $z = \frac{1}{1+c}$ and is equal to $\frac{2c}{1+c}$. This gives the first inequality.

To get the second one, we follow an argument of [Bou]. Fix $\alpha \in (0,1)$ and put $u_i = 1/z_i$ ($z_i > 0$). By Chebyshev's and Hölder's inequalities,

$$\begin{aligned}
\mathrm{mes}\{t \in (0,1) : |f(t)| < \varepsilon\,|f(0)|\} &= \mathrm{mes}\left\{\prod_{i=1}^{d} |u_i t - 1|^{-\alpha/d} > \varepsilon^{-\alpha/d}\right\} \\
&\leq \varepsilon^{\alpha/d} \int_0^1 \prod_{i=1}^{d} |u_i t - 1|^{-\alpha/d}\, dt \\
&\leq \varepsilon^{\alpha/d} \left(\prod_{i=1}^{d} \int_0^1 |u_i t - 1|^{-\alpha}\, dt\right)^{1/d} \leq \frac{2\varepsilon^{\alpha/d}}{1-\alpha}
\end{aligned}$$

where we used a simple inequality $\int_0^1 |ut - 1|^{-\alpha}\, dt \leq \frac{2}{1-\alpha}$ ($u \geq 0$) on the last step. It remains to optimize over all $\alpha \in (0,1)$.

Thus, the condition of Corollary 4.1 is fulfilled with $r = 1/d$ and $C = 2d$. Hence:

Corollary 4.4. *For all* $\lambda \geq (4d)^d$,

$$P\left\{|f| > \lambda\,\mathbf{E}|f|\right\} \leq e^{-\lambda^{1/d}/(8ed)}.$$

The upper bound can further be sharpened with the help of the localization lemma of Lovász-Simonovits [L-S] which allows one to get in Khinchine-type inequalities for polynomials a correct order of constants as functions of degree d. As shown in [B1-2], for all $p \geq 1$,

$$(\mathbf{E}|f|^p)^{1/p} \leq (cp)^d\,\mathbf{E}|f|,$$

where $c > 1$ is a universal constant. Hence, by Chebyshev's inequality, for all $\lambda > 0$, $P\{|f| > \lambda\,\mathbf{E}|f|\} \leq \frac{(cp)^{pd}}{\lambda^p}$. Optimizing the right hand side over $p \geq 1$, we arrive at

$$P\{|f| > \lambda\,\mathbf{E}|f|\} \leq e^{-d\lambda^{1/d}/(ce)},$$

provided that $\lambda \geq (ce)^d$.

5 Deviations from the mean

Large deviations of f from the mean $\mathbf{E}f = \int f\,dP$ can be controled once we know how to estimate the quantity δ_{f-c} uniformly over all $c \in \mathbf{R}$. For example, since the class of polynomials f of degree d is closed under translations $f \to f + \mathrm{const}$, (1.2) implies the bound

$$P\{|f - \mathbf{E}f| > \lambda\sigma\} \le C(d)\, \exp\left\{-c(d)\, \lambda^{1/d}\right\}, \quad \lambda \ge 0,$$

in terms of the variance $\sigma^2 = \mathbf{E}(f - \mathbf{E}f)^2$. One may therefore hope to reach similar dimension-free inequalities for other classes of functions. The question is stimulated by the observation (typical in concentration problems, cf. [M-S], [L]) that many interesting f's have very large expectations $\mathbf{E}f$, but relatively small variances σ^2. In this situation, bounds for $P\{|f - \mathbf{E}f| > \lambda\sigma\}$ are certainly more delicate and preferable in comparison with those for $P\{|f| > \lambda \mathbf{E}|f|\}$. However, if we wish to involve into consideration arbitrary norms, the desired extension of (1.1) to the larger class $f(x) = \|x\| - c$ is no longer valid, and some extra condition on the norm like the uniform convexity is required. To illustrate these ideas, we will consider here the example of the euclidean norm $f(x) = \|x\|_2$ on $\mathbf{R}^n$.

To start with, it might be reasonable to find an appropriate form of Theorem 1.1 for the case of devations from constants. To every continuous function f on $\mathbf{R}^n$ and $\varepsilon > 0$, we may associate another quantity $\Delta_f(\varepsilon)$ defined to be the least number $\Delta \in [0,1]$ such that, for all $x_0, x_1 \in \mathbf{R}^n$, the function $Q(t) = f(tx_0 + (1-t)x_1)$ satisfies

$$\mathrm{mes}\left\{t \in [0,1] : Q(t) - \min_{0 \le s \le 1} Q(s) < \varepsilon \left[\max_{0 \le s \le 1} Q(s) - \min_{0 \le s \le 1} Q(s)\right]\right\} \le \Delta.$$

Theorem 5.1. *Let P be a log-concave probability measure on $\mathbf{R}^n$, and let f be a convex function on $\mathbf{R}^n$ with mean $\mathbf{E}f$ and variance σ^2. Then, for all $\lambda > 2e$ such that $\Delta_f(4e/(\lambda + 2e)) \le 1/2$,*

$$P\{|f - \mathbf{E}f| > \lambda\,\sigma\} \le \exp\left\{-\frac{1}{2\delta_f(4e/(\lambda + 2e))}\right\}.$$

The statement follows immediately from Theorem 1.1 and

Lemma 5.2. *For every convex f on $\mathbf{R}^n$, for all $\varepsilon \in (0,1)$,*

$$\sup_{c \in \mathbf{R}} \delta_{f-c}(\varepsilon) = \Delta_f\left(\frac{2\varepsilon}{1+\varepsilon}\right). \tag{5.1}$$

Proof. Fix $x_0, x_1 \in \mathbf{R}^n$ and the corresponding function $Q(t) = f(tx_0+(1-t)x_1)$, $0 \le t \le 1$ (not identically a constant on [0,1]). Since Q is convex, it attains its maximum at $t = 0$ or $t = 1$. By homogeneity and translation invariance of (5.1), and replacing x_0 with x_1 if necessary, we may assume that $Q(1) = \max_{0 \le s \le 1} Q(s) = 1$ and $Q(t_0) = \min_{0 \le s \le 1} Q(s) = 0$, for some $t_0 \in [0, 1)$. Consider the quantity

$$\varphi(c) = \mathrm{mes}\{t \in [0, 1] : |Q(t) - c| < \varepsilon|1 - c|\}$$

appearing in the definition of $\delta_{f-c}(\varepsilon)$. In view of $\sup_{c\in\mathbf{R}} \delta_{f-c}(\varepsilon) = \sup_{x_0,x_1} \sup_{c\in\mathbf{R}} \varphi(c)$, we need to maximize the latter function over all c.

If $c > 1$, $|Q(t) - c| < \varepsilon|1 - c|$ implies $Q(t) > 1$ that is not possible. So, assume $c \le 1$ in which case the definition becomes

$$\varphi(c) = \mathrm{mes}\{t \in [0, 1] : (1 + \varepsilon)c - \varepsilon < Q(t) < (1 - \varepsilon)c + \varepsilon\}. \tag{5.2}$$

In the range $\{c : (1 + \varepsilon)c - \varepsilon < 0\} = (-\infty, \frac{\varepsilon}{1+\varepsilon})$, the first inequality in (5.2) is fulfilled automatically, while the upper bound $(1 - \varepsilon)c + \varepsilon$ increases with c. Thus, we may also assume $c \ge c_0 \equiv \frac{\varepsilon}{1+\varepsilon}$.

As c varies in $[c_0, 1]$, the interval $((1 + \varepsilon)c - \varepsilon, (1 - \varepsilon)c + \varepsilon)$ moves to the right and its length $2\varepsilon(1 - c)$ decreases from $2c_0$ to 0. By the convexity of Q, this implies that the length of the interval $\{t \in [t_0, 1] : (1 + \varepsilon)c - \varepsilon < Q(t) < (1 - \varepsilon)c + \varepsilon\}$ decreases as a function of c. Indeed, if Q is not a constant in any neighborhood of t_0, then it increases in $[t_0, 1]$, the inverse function $Q^{-1} : [0, 1] \to [0, t_0]$ is concave, so, for any positive decreasing function $h = h(u)$, the function $Q^{-1}(u + h) - Q^{-1}(u)$ is decreasing in u, as well. A similar argument applies to Q restricted to the interval $[0, t_0]$. Therefore, $c = c_0$ is the point of minimum to φ. To involve a possible "degenerate" case, we should write

$$\sup_{c\in\mathbf{R}} \varphi(c) = \lim_{c\uparrow c_0} \varphi(c) = \mathrm{mes}\{t \in [0, 1] : Q(t) < (1 - \varepsilon)c_0 + \varepsilon\}.$$

It remains to note that $(1 - \varepsilon)c_0 + \varepsilon = 2c_0 = \frac{2\varepsilon}{1+\varepsilon}$, and the lemma follows.

Now, let us turn to the particular case $f(x) = \|x\|_2$. The euclidean norm can be related to the polynomial f^2 of degree $d = 2$ via the following observation: For every convex $f \ge 0$ on $\mathbf{R}^n$, for all $\varepsilon > 0$ and $q \ge 1$, we have $\Delta_f(\varepsilon) \le \Delta_{f^q}(\varepsilon)$. The latter statement easily follows from the definition and a simple inequality $\frac{b^q-a^q}{c^q-a^q} \le \frac{b-a}{c-a}$, $0 \le a \le b \le c$ ($a \ne c$). Now, appropriate computations show that

$$\Delta_{\|x\|_2^2}(\varepsilon) = \frac{2\sqrt{\varepsilon}}{1 + \sqrt{\varepsilon}}, \quad \varepsilon \in (0, 1).$$

Hence, $\Delta_{\|x\|_2}(\varepsilon) \le \Delta_{\|x\|_2^2}(\varepsilon) \le 2\sqrt{\varepsilon}$. Thus, from Theorem 5.1, we obtain

Corollary 5.3. *Let $X = (X_1, \ldots, X_n)$ be a random vector in $\mathbf{R}^n$ with a log-concave distribution. Let σ^2 be the variance of $\|X\|_2$. Then,*

$$\mathrm{Prob}\{|\,\|X\|_2 - \mathbf{E}\|X\|_2\,| > \lambda\sigma\} \le Ce^{-c\sqrt{\lambda}}, \quad \lambda \ge 0,$$

where C and c are positive numerical constants.

Since one can relate the strength of concentration of $\|X\|_2$ about its mean to the standard deviation σ, one may wonder how to bound the variance itself. For normalization, let the covariances of the components of X satisfy

$$\mathrm{cov}(X_i, X_j) \equiv \mathbf{E}X_iX_j - \mathbf{E}X_i\mathbf{E}X_j = \delta_{ij}, \quad 1 \le i, j \le n, \tag{5.3}$$

where δ_{ij} is the Kronecker symbol. Under this (isotropy) assumption, the question of whether or not $\sigma^2 = \mathrm{Var}(\|X\|_2)$ does not exceed a universal constant represents a week form of a conjecture of R. Kannan, L. Lovász and M. Simonovits, cf. [K-L-S]. One simple sufficient condition of dimension free boundedness of σ^2, namely, the property

$$\mathrm{cov}(X_i^2, X_j^2) \le 0, \quad 1 \le i < j \le n, \tag{5.4}$$

was recently proposed by K. Ball and I. Perissinaki [B-P]. Indeed, for positive random variables ξ's, there is a general estimate $\mathrm{Var}(\xi) \le \frac{\mathrm{Var}(\xi^2)}{\mathbf{E}\xi^2}$, which for $\xi = \|X\|_2$ in view of (5.3) becomes $\mathrm{Var}(\|X\|_2) \le \frac{\mathrm{Var}(\|X\|_2^2)}{n}$. On the other hand,

$$\mathrm{Var}(\|X\|_2^2) = \sum_{i=1}^n \mathrm{Var}(X_i^2) + 2\sum_{i<j} \mathrm{cov}(X_i^2, X_j^2) \le \sum_{i=1}^n \mathbf{E}X_i^4 \le Cn,$$

where we used (5.4) and Khinchine-type inequality $\mathbf{E}X_i^4 \le C\,(\mathbf{E}X_i^2)^2 = C$.

In [B-P], a property implying (5.3) was verified for random vectors X uniformly distributed in ℓ_p^n balls in $\mathbf{R}^n$.

6 Appendix: Triangular maps

Here we recall some facts about triangular maps which are needed for the proof of Theorem 2.1. A map $T = (T_1, \dots, T_n) : G \to \mathbf{R}^n$ defined on an open non-empty set G in $\mathbf{R}^n$ is called triangular if its components are of the form

$$T_i = T_i(x_1, \dots, x_i), \qquad x \in G, \quad 1 \le i \le n.$$

The triangular map T will be called increasing if, for all $i \le n$, the component T_i is a (strictly) increasing function with respect to x_i-coordinate while the rest coordinates are fixed (x_i may vary within an open interval which depends on the rest coordinates x_j, $j < i$).

Such maps were used by H. Knothe [Kn] to reach some generalizations of the Brunn-Minkowski inequality. The following statement is often refered to as the construction of the Knothe mapping [Kn].

Theorem 6.1. *Let A and B be open, bounded, non-empty convex sets in $\mathbf{R}^n$. There exists a continuous, bijective, triangular map $T : A \to B$ such that*

a) *the partial derivatives $\frac{\partial T_i}{\partial x_i}$ are continuous and positive on A ;*

b) *the Jacobian $J(x) = \mathrm{Det}(T'(x)) = \prod_{i=1}^n \frac{\partial T_i}{\partial x_i}$ is constant on A and satisfies*

$$J(x) = \frac{\mathrm{Vol}_n(B)}{\mathrm{Vol}_n(A)}, \qquad x \in A;$$

c) *the map T pushes forward the uniform distribution on A to the uniform distribution on B.*

Note that T is not required to be C^1-smooth, so the property b) first defines a function "Jacobian" and then pustulates that it is a constant.

To complete Bourgain's argument, we need an appropriate generalization of Theorem 6.1 for measures. In [Bou], Theorem 6.1 is stated without convexity assumption on A which might lead to singularity problems. Indeed, consider, for example, the sets $B = (0,1)\times(0,1)$ and $A = (0,1)\times(0,2)\cup(0,2)\times(0,1) \subset \mathbf{R}^2$. The set A is open, bounded and has Lebesgue measure $|A| = 3$. Let P be a probability measure which has density $p(x) = 1/3$, for $x \in A$, and $p = 0$ outside A. Then, the distrubution P_1 of x_1-coordinate under P is concentrated on the interval $A_1 = (0,2)$ and has there density

$$p_1(x_1) = \begin{cases} 2/3, & \text{if} \quad 0 < x_1 < 1 \\ 1/3, & \text{if} \quad 1 < x_1 < 2 \end{cases}$$

That is, P_1 does not have any continuous density on $(0,2)$. But the property that P_1 has a continuous density is necessary for smoothness of triangular maps which push forward P to the uniform measure Q on B.

Thus, to save the property a) in the general non-convex case, some extra condition is required. First note that, given random vectors $X = (X_1,\dots,X_n)$ and $Y = (Y_1,\dots,Y_n)$ with values in open sets A and B and distributed according to P and Q, respectively, the first i coordinates $(X_1,\dots,X_i)$ and $(Y_1,\dots,Y_i)$ $(1 \le i \le n)$ have distributions P_i and Q_i supported on the open sets

$$A_i = \{x \in \mathbf{R}^i : \exists\, t \in \mathbf{R}^{n-i} \ \ (x,t) \in A\},$$

$$B_i = \{x \in \mathbf{R}^i : \exists\, t \in \mathbf{R}^{n-i} \ \ (x,t) \in B\},$$

which are projections of A and B to $\mathbf{R}^i$. In particular, $A_n = A$, $B_n = B$.

We will say that P is regular if it has a (necessarily continuous) density p on A such that the following two conditions are satisfied. The first condition is that, for each $i \le n$, the measure P_i has a positive continuous density p_i on A_i. This is equivalent to saying that the integral

$$p_i(x_1,\dots,x_i) = \int_{\mathbf{R}^{n-i}} p(x_1,\dots,x_i,t_{i+1},\dots,t_n)\,dt_{i+1}\dots dt_n$$

is finite for every $(x_1, \dots, x_i) \in A_i$ and represents a continuous function on A_i (when $i = n$, we just have $p_i = p$). The second condition is that, for each $i \le n$, the conditional distribution function

$$\mathrm{Prob}\{X_i \le x_i \,|\, X_j = x_j,\ j = 1, \dots, x_{i-1}\}$$

is continuous in $(x_1, \dots, x_i) \in A_i$. Under the first condition, this is equivalent to saying that the integral

$$\int_{-\infty}^{x_i} \int_{\mathbf{R}^{n-i}} p(x_1, \dots, x_{i-1}, t_i, \dots, t_n)\, dt_i \dots dt_n$$

defines a continuous function on A_i. We can now state a corresponding generalization of Theorem 6.1.

Theorem 6.2. *For all regular probability measures P and Q supported on an open set A and on an open convex set B, respectively, there exists unique increasing, continuous, triangular, bijective map $T : A \to B$ which pushes forward P to Q. Moreover, the components T_i are C^1-smooth with respect to x_i-coordinates and satisfy $\frac{\partial T_i}{\partial x_i} > 0$ on A_i.*

Some examples of regular measures will be described at the end of this section, and we turn to the study of triangular maps T themselves. Many properties of such maps are determined by behavior of functions T_i with respect to x_i-coordinates. We collect some properties in the two lemmas below.

Lemma 6.3. *If $T = (T_1, \dots, T_n) : G \to \mathbf{R}^n$ is a continuous, increasing, triangular map, then the image $T(G)$ is an open set, and T represents a homeomorphism between G and $T(G)$.*

Lemma 6.4. *Let $T = (T_1, \dots, T_n) : G \to \mathbf{R}^n$ be a continuous triangular map whose components T_i have continuous positive partial derivatives $\frac{\partial T_i}{\partial x_i}$ on G. Then, for every integrable function f on $\mathbf{R}^n$,*

$$\int_G f(T(x))\, J(x)\, dx = \int_{T(G)} f(y)\, dy, \tag{6.1}$$

where $J(x) = \prod_{i=1}^n \frac{\partial T_i(x)}{\partial x_i}$.

Note that, by Lemma 6.3, the map T from Lemma 6.4 is increasing, so T is a bijection from G to the open set $T(G)$. The topological Lemma 6.3 can easily be proved by virtue of Brauer's theorem, so we omit the proof.

Proof of Lemma 6.4. If T is C^1-smooth on G, i.e., T has a continuous derivative $T' = (\frac{\partial T_i}{\partial x_j})_{1 \le i,j \le n}$, the "true" Jacobian $J(x) = \mathrm{Det}(T'(x)) = \prod_{i=1}^n \frac{\partial T_i(x)}{\partial x_i}$

is well-defined and is everywhere positive on G. Hence, the equality (6.1) holds true by the well-known theorem on the change of the variable in the Lebesgue integral.

In general, we approximate T by smooth triangular maps T^ε. Recall that the domain of T_i is the open set

$$G_i = \{x = (x_1, \dots, x_i) \in \mathbf{R}^i : \exists y \in \mathbf{R}^{n-i}\ (x, y) \in \mathbf{R}^n\}.$$

Now, take C_0^∞-functions $K_i \geq 0$, $i = 1, \dots, n$, supported on the unit ball $D_i(0,1) \subset \mathbf{R}^i$ and such that $\int_{D_i(0,1)} K_i(y)\, dy = 1$, and introduce convolutions T_i^ε of T_i with $K_i^\varepsilon(x) = \frac{1}{\varepsilon^i} K_i(x/\varepsilon)$, $\varepsilon > 0$:

$$T_i^\varepsilon(x) = \int_{G_i} K_i^\varepsilon(x - z)\, T_i(z)\, dz, \quad x \in \mathbf{R}^i.$$

The above integral is well-defined and represents a C^∞-function on $\mathbf{R}^i$. Differentiating over x_i, we obtain a C^∞-function

$$\frac{\partial T_i^\varepsilon(x)}{\partial x_i} = \int_{G_i} \frac{\partial K_i^\varepsilon(x - z)}{\partial x_i}\, T_i(z)\, dz, \quad x \in \mathbf{R}^i.$$

The kernel K_i^ε is supported on $D_i(0, \varepsilon)$. Hence, this integral may be taken over the whole space $\mathbf{R}^i$ as soon as $D_i(x, \varepsilon)$ lies in G_i:

$$\frac{\partial T_i^\varepsilon(x)}{\partial x_i} = \int_{\mathbf{R}^i} \frac{\partial K_i^\varepsilon(x - z)}{\partial x_i}\, T_i(z)\, dz, \quad x \in \mathbf{R}^i.$$

Let $\varepsilon_i(x)$ be the supremum of such ε's. Integrating by parts, for $\varepsilon \in (0, \varepsilon_i(x)]$, $\varepsilon < +\infty$, we get

$$\frac{\partial T_i^\varepsilon(x)}{\partial x_i} = \int_{\mathbf{R}^i} K_i^\varepsilon(x - z)\, \frac{\partial T_i(z)}{\partial z_i}\, dz = \int_{G_i} K_i^\varepsilon(x - z)\, \frac{\partial T_i(z)}{\partial z_i}\, dz, \quad x \in \mathbf{R}^i.$$

The integral on the right is again well-defined and represents a C^∞-function as the convolution of $\frac{\partial T_i}{\partial z_i}$ with K_i^ε. Moreover, it is positive, by the assumptions on T_i and K_i.

Thus, the map $T^\varepsilon = (T_1^\varepsilon, \dots, T_n^\varepsilon)$ is triangular, C^∞-smooth, with positive Jacobian

$$J_\varepsilon(x) = \prod_{i=1}^n \frac{\partial T_i^\varepsilon(x_1, \dots, x_i)}{\partial x_i}$$

at every point $x \in G$ and for all $\varepsilon \in (0, +\infty)$ such that $0 < \varepsilon \leq \varepsilon_n(x)$ (note that the numbers $\varepsilon_i(x_1, \dots, x_i)$ decrease when i increases from 1 to n). Moreover, $J_\varepsilon(x) > 0$ on the set $G_\varepsilon = \{x \in G : D_n(x, \varepsilon) \subset G\}$, the open ε-interior of G. Therefore, we can apply (6.1) to any open set $A \subset G_\varepsilon$ and every intergrable function f on $\mathbf{R}^n$ to get

$$\int_A f(T^\varepsilon(x))\, J_\varepsilon(x)\, dx = \int_{T^\varepsilon(A)} f(y)\, dy. \tag{6.2}$$

Now, by the choice of K_i's, and since T and $\frac{\partial T_i}{\partial x_i}$ are continuous, $T^\varepsilon(x) \to T(x)$ and $\frac{\partial T_i^\varepsilon(x)}{\partial x_i} \to \frac{\partial T_i(x)}{\partial x_i}$, as $\varepsilon \downarrow 0$ uniformly over all $x \in A$, for every A whose closure $\mathrm{clos}(A)$ is compact and lies in G. Similarly, $J_\varepsilon(x) \to J(x)$.

Let A be open with compact closure $\mathrm{clos}(A) \subset G$. Since $G_\varepsilon \uparrow G$, as $\varepsilon \downarrow 0$, there is $\varepsilon_0 > 0$ such that $\mathrm{clos}(A) \subset G_{\varepsilon_0}$. Hence, we can apply the Lebesgue dominated convergence theorem: for every continuous function f on $\mathbf{R}^n$,

$$\int_A f(T^\varepsilon(x))\, J_\varepsilon(x)\, dx \to \int_A f(T(x))\, J(x)\, dx, \quad \text{as} \quad \varepsilon \downarrow 0. \tag{6.3}$$

To find the limit of the right hand side of (6.2), first note that

$$\limsup_{\varepsilon \downarrow 0} T^\varepsilon(A) \subset \mathrm{clos}(T(A)) = T(\mathrm{clos}(A)).$$

On the other hand, by a topological argument, we have $T(A) \subset \liminf_{\varepsilon\downarrow 0} T^\varepsilon(A)$, that is, whenever $a \in A$, the point $b = T(a)$ is contained in $T^\varepsilon(A)$, for all $\varepsilon > 0$ small enough. As a result, $1_{T(A)}(y) \leq \liminf_{\varepsilon\downarrow 0} 1_{T^\varepsilon(A)}(y) \leq \limsup_{\varepsilon\downarrow 0} 1_{T^\varepsilon(A)}(y) \leq 1_{T(\mathrm{clos}(A))}(y)$, for every $y \in \mathbf{R}^n$. Hence, for every non-negative bounded continuous function f on $\mathbf{R}^n$,

$$\int_{T(A)} f(y)\, dy \leq \liminf_{\varepsilon\downarrow 0} \int_{T^\varepsilon(A)} f(y)\, dy \leq \limsup_{\varepsilon\downarrow 0} \int_{T^\varepsilon(A)} f(y)\, dy \leq \int_{T(\mathrm{clos}(A))} f(y)\, dy.$$

Together with (6.2)-(6.3) we get

$$\int_{T(A)} f(y)\, dy \leq \int_A f(T(x))\, J(x)\, dx \leq \int_{T(\mathrm{clos}(A))} f(y)\, dy.$$

This already easily implies the equality (6.1). Lemma 6.4 follows.

In order to turn to the proof of Theorem 6.2, let us first emphasize what exactly we need to prove. Assume we have two absolutely continuous probability measures P and Q on $\mathbf{R}^n$ which are supported on some open sets A and B and have there densities $p(x)$ and $q(y)$, respectively. We wish to construct a continuous bijective map $T = (T_1, \dots, T_n) : A \to B$ which pushes forward P to Q. This property is denoted $Q = PT^{-1}$ or $Q = T(P)$ and can be defined via the equality $\int_B f\, dQ = \int_A f(T)\, dP$ or, in terms of densities, as

$$\int_B f(y) q(y)\, dy = \int_A f(T(x)) p(x)\, dx, \tag{6.4}$$

holding for every bounded measurable function f on B. If T is C^1-smooth and has at every point $x \in A$ an invertible matrix $T'(x) = (\frac{\partial T_i(x)}{\partial x_j})_{1 \leq i,j \leq n}$ of the first derivatives, one can make in the first integral the change of variable $y = T(x)$, and (6.4) becomes

$$\int_A f(T(x))\, q(T(x))\, |\mathrm{Det}(T'(x))|\, dx = \int_A f(T(x))\, p(x)\, dx.$$

Thus, if T is bijective, C^1-smooth, and the Jacobian $J(x) = \mathrm{Det}(T'(x))$ is everywhere positive, the necessary and sufficient condition for $Q = PT^{-1}$ is that, for almost all $x \in A$,

$$q(T(x))J(x) = p(x). \tag{6.5}$$

In the case where the map T is increasing and triangular, one can weaken the smoothness requirement and just assume that the components T_i have positive continuous derivatives $\frac{\partial T_i}{\partial x_i}$. Indeed, if $B = T(A)$, then, by Lemma 6.4, the equality

$$\int_B g(y)\,dy = \int_A g(T(x))\,J(x)\,dx$$

holds true for every integrable function g on B with $J(x) = \prod_{i=1}^n \frac{\partial T_i(x)}{\partial x_i}$. Applying this equality to $g(y) = f(y)q(y)$, we get

$$\int_B f(y)q(y)\,dy = \int_A f(T(x))q(T(x))\,J(x)\,dx.$$

Therefore, (6.4) would immediately follow from (6.5). Thus, we may conclude:

Lemma 6.5. *Let $T = (T_1, \dots, T_n) : A \to \mathbf{R}^n$ be a continuous triangular map whose components T_i have continuous positive partial derivatives $\frac{\partial T_i}{\partial x_i}$ on A. Let $B = T(A)$. If the equality* (6.5) *holds true for almost all $x \in A$, then the map T pushes forward P to Q.*

However, the existence of the triangular map T satisfying (6.5) requires more properties such as regularity of P and Q.

Proof of Theorem 6.2. We use induction over n, and prove at the same time that the components T_i, $1 \le i \le n$, satisfy, for all $(x_1, \dots, x_i) \in A_i$, the relation

$$\frac{\int_{-\infty}^{x_i} \int_{\mathbf{R}^{n-i}} p(x_1, \dots, x_{i-1}, t_i, \dots, t_n)\,dt_i \dots dt_n}{\int_{-\infty}^{+\infty} \int_{\mathbf{R}^{n-i}} p(x_1, \dots, x_{i-1}, t_i, \dots, t_n)\,dt_i \dots dt_n} =$$

$$\frac{\int_{-\infty}^{T_i} \int_{\mathbf{R}^{n-i}} q(T_1, \dots, T_{i-1}, t_i, \dots, t_n)\,dt_i \dots dt_n}{\int_{-\infty}^{+\infty} \int_{\mathbf{R}^{n-i}} q(T_1, \dots, T_{i-1}, t_i, \dots, t_n)\,dt_i \dots dt_n}\,, \tag{6.6}$$

where it is also claimed that all the integrals are finite and positive. For $i = 1$, the above formula becomes

$$\int_{-\infty}^{x_1} \int_{\mathbf{R}^{n-1}} p(t_1, \dots, t_n)\,dt_1 \dots dt_n = \int_{-\infty}^{T_1} \int_{\mathbf{R}^{n-1}} q(t_1, \dots, t_n)\,dt_1 \dots dt_n\,, \tag{6.7}$$

while for $i = n$, it reads as

$$\frac{\int_{-\infty}^{x_n} p(x_1, \dots, x_{n-1}, t_n)\,dt_n}{\int_{-\infty}^{+\infty} p(x_1, \dots, x_{n-1}, t_n)\,dt_n} = \frac{\int_{-\infty}^{T_n} q(T_1, \dots, T_{n-1}, t_n)\,dt_n}{\int_{-\infty}^{+\infty} q(T_1, \dots, T_{n-1}, t_n)\,dt_n}\,. \tag{6.8}$$

Note that the formulas (6.6)-(6.8) may be written in a more compact probabilistic form as

$$\mathrm{Prob}\{X_i \le x_i \,|\, X_1 = x_1, \dots, X_{i-1} = x_{i-1}\} = \mathrm{Prob}\{Y_i \le T_i \,|\, Y_1 = T_1, \dots, Y_{i-1} = T_{i-1}\},$$

where for $i \le n$ we write for short $T_i = T_i(x_1, \dots, x_i)$.

The case $n = 1$ is obvious: the desired map $T = T_1(x_1)$ is unique and is determined by

$$\int_{-\infty}^{x_1} p(t_1)\, dt_1 = \int_{-\infty}^{T_1(x_1)} q(t_1)\, dt_1. \tag{6.9}$$

Clearly, T_1 is a C^1-smooth increasing bijection from A to B (B is an interval).

Now, to perfom the induction step, assume $n \ge 2$ and recall that P_i and Q_i denote the distrubution of the first i variables $(x_1, \dots, x_i)$ under P and Q, respectively. By the induction hypothesis, there is unique continuous, increasing, triangular bijective map $(T_1, \dots, T_{n-1}) : A_{n-1} \to B_{n-1}$ which transports P_{n-1} to Q_{n-1}, and moreover, the equality (6.6) holds true on A_i for all $i \le n-1$.

According to (6.9) for the case $n = 1$, the equality (6.7) expresses the fact that the measure P_1 is transported to Q_1 by the map T_1. Similarly and more generally, the equality (6.6) expresses the fact that, given a vector $(x_1, \dots, x_{i-1}) \in A_{i-1}$, the function

$$x_i \to T_i(x_1, \dots, x_{i-1}, x_i)$$

transports the corresponding conditional measure $P_{x_1,\dots,x_{i-1}}$ of P on the line in $\mathbf{R}^i$ with these first $i-1$ coordinates to the conditional measure $Q_{T_1(x_1),\dots,T_{i-1}(x_1,\dots,x_{i-1})}$ of Q on the line with fixed coordinates $T_1(x_1), \dots, T_{i-1}(x_1, \dots, x_{i-1})$. In order to make the same to be valid when $i = n$, we just postulate equality (6.8) as the definition of T_n. Note that $P_{x_1,\dots,x_{n-1}}$ represents a probability measure which is supported on the open one dimensional set

$$A(x_1, \dots, x_{n-1}) = \{x \in \mathbf{R} : (x_1, \dots, x_{n-1}, x) \in A\},$$

while $Q_{T_1(x_1),\dots,T_{n-1}(x_1,\dots,x_{n-1})}$ is a probability measure supported (by convexity of B) on the open segment

$$B(x_1, \dots, x_{n-1}) = \{y \in \mathbf{R} : (T_1(x_1), \dots, T_{n-1}(x_1, \dots, x_{n-1}), y) \in B\}.$$

In addition, by the regularity assumption made on P and Q, these measures have positive continuous densities on $A(x_1, \dots, x_{n-1})$ and $B(x_1, \dots, x_{n-1})$, respectively. Hence, as well as in the case $n = 1$, for all $(x_1, \dots, x_{n-1}) \in A_{n-1}$, the function

$$x_n \to T_n(x_1, \dots, x_{n-1}, x_n)$$

represents a C^1-smooth increasing bijection from $A(x_1, \dots, x_{n-1})$ to $B(x_1, \dots, x_{n-1})$. This proves that $(T_1, \dots, T_{n-1}, T_n)$ is an increasing bijection from A to B together with the fact that all components T_i are C^1-smooth with respect to x_i-coordinates.

It should also be clear that, for each $i \le n$, the function T_i continuously depends on $(x_1, \dots, x_i)$. Indeed, the case $i = 1$ does not need to be verified, while for $i \ge 2$ we may argue using induction over i. Assuming that $T_1, \dots, T_{i-1}$ are continuous, introduce the function

$$\psi(x_1, \dots, x_{i-1}, y) = \int_{-\infty}^{y} \int_{\mathbf{R}^{n-i}} q(T_1, \dots, T_{i-1}, t_i, \dots, t_n)\, dt_i \dots dt_n$$

and write the equality (6.6) as

$$R(x_1, \dots, x_i) = \psi(x_1, \dots, x_{i-1}, T_i).$$

By the regularity assumption on P and Q and the induction hypothesis, both R and ψ are continuous functions defined respectively on the open sets A_i and

$$\{(x_1, \dots, x_{i-1}, y) : (x_1, \dots, x_{i-1}) \in A_{i-1},\ (T_1, \dots, T_{i-1}, y) \in B_i\}.$$

In particular, if $x'_j \to x_j$ for all $j = 1, \dots, i$, and $y = T_i(x_1, \dots, x_i)$, $y' = T_i(x'_1, \dots, x'_i)$, we get that

$$\psi(x'_1, \dots, x'_{i-1}, y') \to \psi(x_1, \dots, x_{i-1}, y).$$

The function ψ increases with respect to y. So, if y' does not converge to y, and for definiteness $y' \le y - \varepsilon$ for some $\varepsilon > 0$, then for some $\delta > 0$,

$$\begin{aligned} \psi(x'_1, \dots, x'_{i-1}, y') &\le \psi(x'_1, \dots, x'_{i-1}, y - \varepsilon) \to \psi(x_1, \dots, x_{i-1}, y - \varepsilon) \\ &< \psi(x_1, \dots, x_{i-1}, y) - \delta \end{aligned}$$

which is a contradiction. Hence, $y' \to y$, and thus T_i is continuous.

Now, differentiating (6.7) over x_1, (6.6) over x_i, where we assume that $2 \le i \le n-1$, and (6.8) over x_n, we get respectively,

$$\int_{\mathbf{R}^{n-1}} p(x_1, t_2, \dots, t_n)\, dt_2 \dots dt_n = \int_{\mathbf{R}^{n-1}} q(T_1, t_2, \dots, t_n)\, dt_2 \dots dt_n \ \frac{\partial T_1}{\partial x_1}, \quad (6.10)$$

$$\begin{aligned} &\frac{\int_{\mathbf{R}^{n-i}} p(x_1, \dots, x_i, t_{i+1}, \dots, t_n)\, dt_{i+1} \dots dt_n}{\int_{\mathbf{R}^{n-i+1}} p(x_1, \dots, x_{i-1}, t_i, \dots, t_n)\, dt_i \dots dt_n} = \\ &\frac{\int_{\mathbf{R}^{n-i}} q(T_1, \dots, T_i, t_{i+1}, \dots, t_n)\, dt_{i+1} \dots dt_n}{\int_{\mathbf{R}^{n-i+1}} q(T_1, \dots, T_{i-1}, t_i, \dots, t_n)\, dt_i \dots dt_n} \ \frac{\partial T_i}{\partial x_i}, \end{aligned} \quad (6.11)$$

$$\frac{p(x_1, \dots, x_n)}{\int_{-\infty}^{+\infty} p(x_1, \dots, x_{n-1}, t_n)\, dt_n} = \frac{q(T_1, \dots, T_n)}{\int_{-\infty}^{+\infty} q(T_1, \dots, T_{n-1}, t_n)\, dt_n} \ \frac{\partial T_n}{\partial x_n}. \quad (6.12)$$

Myltiplying (6.10)-(6.11)-(6.12) by each other, we arrive at

$$p(x_1, \dots, x_n) = q(T_1, \dots, T_n) \prod_{i=1}^{n} \frac{\partial T_i}{\partial x_i}$$

which is exactly (6.5). It remains to apply Lemma 6.5, and the existence part of Theorem 6.2 immediately follows.

The uniqueness follows from the requirement that, given a vector $(x_1, \dots, x_{i-1}) \in A_{i-1}$, the function $x_i \to T_i(x_1, \dots, x_{i-1}, x_i)$ must transport the conditional measure $P_{x_1,\dots,x_{i-1}}$ to the conditional measure $Q_{T_1(x_1),\dots,T_{i-1}(x_1,\dots,x_{i-1})}$.

Theorem 6.2 is now proved.

To give some examples of regular measures (in the above sense), we need another definition. In addition to the projections A_i, with every set $A \subset \mathbf{R}^n$, we also associate its sections

$$A_{x_1,\dots,x_i} = \{t \in \mathbf{R}^{n-i} : (x_1, \dots, x_i, t) \in A\}, \qquad (x_1, \dots, x_i) \in \mathbf{R}^i,\ 1 \leq i \leq n-1.$$

We say that A is regular, if for all $i \leq n-1$ and for all $(x_1, \dots, x_i) \in A_i$, the section $(\partial A)_{x_1,\dots,x_i}$ of the boundary of A has the $(n-i)$-dimensional Lebesgue measure zero.

For example, a finite union of balls represents a regular set. Another simple example is provided by an arbitrary open convex set in $\mathbf{R}^n$. As for regularity of measures, the following lemma covers most interesting cases.

Lemma 6.6. *Assume that a probability measure P is concentrated on an open set $B \subset \mathbf{R}^n$ where it has a positive continuous density p such that, for each $i \leq n-1$,*

$$\int_{\mathbf{R}^{n-i}} \sup_{x \in B_i} p(x,t)\, dt < +\infty \tag{6.13}$$

(where it is assumed that $p = 0$ outside B). Then, the normalized restriction of P to any regular set $A \subset B$ is a regular measure.

The condition (6.13) is fulfilled, for example, if with some positive constants C and c, the density p satsfies an inequality

$$p(x) \leq C e^{-c|x|}, \quad x \in B. \tag{6.14}$$

Proof of Lemma 6.7. By the assumption, the function

$$p_i(x) = \int_{\mathbf{R}^{n-i}} p(x,t)\, 1_A(x,t)\, dt$$

is finite for every $x \in B_i$, and moreover the function under the integral sign is bounded by an integrable function. We should show that p_i is continuous on A_i.

So, take a sequence $x^{(k)} \in A_i$ converging to a point $x \in A_i$, as $k \to \infty$. Then, for every $t \in \mathbf{R}^{n-1}$,

$$1_A(x,t) \le \liminf_{k\to\infty} 1_A(x^{(k)},t) \le \limsup_{k\to\infty} 1_A(x^{(k)},t) \le 1_{\mathrm{clos}(A)}(x,t).$$

Since $p_i(x^{(k)},t) \to p(x,t)$, as $k \to \infty$, and since A is open, we may apply Lebesgue dominated converging theorem which gives

$$\begin{aligned} p_i(x) &\le \liminf_{k\to\infty} \int_{\mathbf{R}^{n-i}} p(x^{(k)},t)\, 1_A(x^{(k)},t)\, dt \\ &\le \limsup_{k\to\infty} \int_{\mathbf{R}^{n-i}} p(x^{(k)},t)\, 1_A(x^{(k)},t)\, dt \ \le \ \int_{\mathbf{R}^{n-i}} p(x,t)\, 1_{\mathrm{clos}(A)}(x,t)\, dt. \end{aligned}$$

Now note that, by regularity of A, for any $x \in A_i$

$$1_{\mathrm{clos}(A)}(x,t) - 1_A(x,t) = 1_{\partial A}(x,t) = 0, \quad \text{for almost all} \ \ t \in \mathbf{R}^{n-i}$$

with respect to Lebesgue measure on $\mathbf{R}^{n-i}$. Hence, $\int_{\mathbf{R}^{n-i}} p(x,t)\, 1_{\mathrm{clos}(A)}(x,t)\, dt = p_i(x)$, and thus p_i is continuous. The first condition involved in the definition of regularity of a measure is therefore fulfilled. The second condition requires to verify that, for each $i \le n$, the function

$$\begin{aligned} r_i(x,x_i) &= \int_{-\infty}^{x_i} \int_{\mathbf{R}^{n-i}} p(x,t_i,t)\, 1_A(x,t_i,t)\, dt_i\, dt \\ &= \int_{-\infty}^{+\infty} \int_{\mathbf{R}^{n-i}} p(x,t_i,t)\, 1_A(x,t_i,t)\, 1_{(-\infty,x_i]\times\mathbf{R}^{n-i}}(t_i,t)\, dt_i\, dt \end{aligned}$$

is continuous in $(x,x_i) \in A_i$, as well, where for short we write $x = (x_1,\dots,x_{i-1})$, $t = (t_{i+1}\ \dots,t_n)$. In case $i = 1$, the above expression depends on x_1, only,

$$r_1(x_1) = \int_{-\infty}^{+\infty} \int_{\mathbf{R}^{n-1}} p(t_1,t)\, 1_A(t_1,t)\, 1_{(-\infty,x_1]\times\mathbf{R}^{n-1}}(t_1,t)\, dt_1\, dt,$$

and is clearly continuous on A_1. In the case $i \ge 2$, we use the property that, for every $(t_i,t) \in \mathbf{R}\times\mathbf{R}^{n-i}$, the function $(x,x_i) \to p(x,t_i,t)\, 1_{(-\infty,x_i]\times\mathbf{R}^{n-i}}(t_i,t)$ is continuous on A_i, and then argue as before: for any $x \in A_{i-1}$,

$$1_{\mathrm{clos}(A)}(x,t_i,t) - 1_A(x,t_i,t) = 1_{\partial A}(x,t_i,t) = 0, \quad \text{for almost all} \ \ (t_i,t) \in \mathbf{R}\times\mathbf{R}^{n-i}$$

with respect to Lebesgue measure on $\mathbf{R}^{n-i+1}$, and therefore, once more by the Lebesgue dominated convergence theorem, r_i is continuous on A_i. Lemma 6.6 follows.

Corollary 6.7. *Uniform distrubution on a bounded regular set is a regular measure.*

This statement appears as a particular case of Lemma 6.6 with $A = B$ and $p = 1/|A|$ on A (where $|A|$ stands for the Lebesgue measure).

At last, since absolutely continuous log-concave measure on $\mathbf{R}^n$ are known to satisfy (6.14), we also obtain:

Corollary 6.8. *Every absolutely continuous log-concave measure P on $\mathbf{R}^n$ is regular. Moreover, the normalized restriction of P to an arbitrary regular set A of positive Lebesgue measure in the support of P represents a regular measure.*

REFERENCES

[A] Alesker, S. Localization technique on the sphere and the Gromov-Milman theorem on the concentration phenomenon on uniformly convex sphere. *Convex Geometric Analysis* (Berkeley, CA, 1996). *Math. Sci. Res. Inst. Publ.*, 34, Cambridge Univ. Press, Cambridge, 1999.

[B-P] Ball, K., Perissinaki, I. Subindependence of coordinate slabs in ℓ_p^n balls. *Israel J. of Math.*, 107 (1998), 289–299.

[B1] Bobkov, S.G. On the growth of L^p-norms of polynomials. *Geom. Aspects of Func. Anal., Lecture Notes in Math.*, 1745 (2000), 27–35.

[B2] Bobkov, S.G. Some generalizations of the results of Yu.V. Prokhorov on Khinchine-type inequalities for polynomials. *Probab. Theory and its Appl.*, 45 (2000), No. 4, 745–748.

[Bor] Borell, C. Convex measures on locally convex spaces. *Ark. Math.*, 12 (1974), 239–252.

[Bou] Bourgain, J. On the distribution of polynomials on high dimensional convex sets. Israel Seminar (GAFA) 1989–90, Lecture Notes in Math., 1469 (1991), 127–137.

[Bre] Brenier, Y. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44 (1991), 375–417.

[G-M] Gromov, M., Milman, V.D. Generalization of the spherical isoperimetric inequality to uniformly convex Banach spaces. *Compositio Math.*, 62 (1987), 263–282.

[G] Guédon, O. Kahane-Khinchine type inequalities for negative exponents. *Mathematika*, 46 (1999), 165–173.

[K-L-S] Kannan, R., Lovász, L., Simonovits, M. Isoperimetric problems for convex bodies and a localization lemma. *Discrete and Comput. Geom.*, 13 (1995), 541–559.

[Kn] Knothe, H. Contributions to the theory of convex bodies. Michigan Math. J., 4 (1957), 39–52.

[L] Ledoux, M. The concetration of measure phenomenon. Math. Surveys and Monographs, AMS, vol. 89, 2001.

[L-S] Lovász, L., Simonovits, M. Random walks in a convex body and an improved volume algorithm. *Random Structures and Algorithms*, 4 (1993), No. 3, 359–412.

[M-S] Milman, V. D., Schechtman, G. Asymptotic theory of finite dimensional normed spaces. *Lecture Notes in Math.*, vol. 1200, 1986, Springer-Verlag.

[M] McCann, R. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80 (1995), No. 2, 309–323.

[N-S-V] Nazarov, F., Sodin, M., Voberg, A. The geometric KLS lemma, dimension-free estimates for the distribution of values of polynomials, and distribution of zeroes of random analytic functions. Preprint (2001).

Sergey G. Bobkov
School of Mathematics
University of Minnesota
127 Vincent Hall, 206 Church St. S.E.
Minneapolis, MN 55455

ON THE DEPENDENCE STRUCTURE OF A SYSTEM OF COMPONENTS WITH A MULTIVARIATE SHOT-NOISE HAZARD RATE PROCESS

Haijun Li*
Department of Mathematics
Washington State University
Pullman, WA 99164
lih@haijun.math.wsu.edu

Abstract

A dynamic random environment usually induces statistical dependence among components of a system operating in such an environment, which makes explicit computation of the joint survival function of component lifelengths very difficult or even intractable. In this paper, we introduce a multivariate shot-noise process to model the situations where the imposed environmental stresses have a correlated residual effect on the failure rates of various components. We then obtain some positive dependence properties (such as association, and orthant dependence) of component lifelengths of a system operating in the multivariate shot-noise environment, and investigate how the dependence structure of component lifelengths varies in response to the environmental change by using the orthant comparison method. Some computable bounds for the joint survival function of component lifelengths are also obtained.

Key words and phrases: Survival in random environments, multivariate

*Supported in part by the NSF grant DMI 9812994 shot-noise process, hazard rate process, association of probability measures on partially ordered spaces, orthant dependence comparison

1 Introduction

Consider the lifelengths $T_1, \dots, T_n$ (assume that they are positive almost surely) of n components of a system that operates in a randomly varying environment. For such a system, the simplifying assumption of components independence has resulted in inadequate assessments of system reliability. In reality, there are a number of situations where some form of dependence exists among various components. A usual factor inducing correlation is the common random environment that affects all the components of the system. The objective of this paper is to introduce a multivariate shot-noise process to model the situations where the imposed stresses from a random environment have a correlated residual effect on the failure rates of various components, and characterize the dependence structures of the systems operating in such an environment.

Failure models for system reliability that incorporate component dependence from the operating environment were proposed and studied by a number of researchers. The reader is referred to two survey papers by Singpurwalla (1995) and by Kijima, Li and Shaked (2000) for an extensive review on the modeling methodologies and a list of the related references. A common approach (Çinlar and Özekici (1987) and Lefèvre and Milhaud (1990)) is to represent the dynamic random environment by a stochastic cadlag process $X = \{X(t), t \geq 0\}$ with an appropriate state space $\mathcal{S}$, and given X, view lifelengths $T_1, \dots, T_n$ as independent random variables, and modulate their failure rates as follows,

$$\lim_{u \to 0} \frac{1}{u} P(t < T_i \leq t + u \mid T_i > t, X) = r_i(t, X(t)), \quad i = 1, \dots, n, \tag{1.1}$$

where each $r_i(t, x)$ is a positive continuous function on $\mathcal{R}_+ \times \mathcal{S}$. Observe that given the external environmental process X, the failure rates depend on X only through the current state $X(t)$, rather than on the whole history of X as described in Çinlar and Özekici (1987). Therefore, the failure model (1.1) is similar to that of Lefèvre and Milhaud (1990). From (1.1), each conditional survival function $P(T_i > t \mid X)$

satisfies the ordinary differential equation,

$$dP(T_i > t \mid X) = -P(T_i > t \mid X) r_i(t, X(t))dt, \quad i = 1, \ldots, n.$$

These equations can be solved 'path by path', and we have

$$P(T_i > t \mid X) = \exp[-\int_0^t r_i(u, X(u))du], \quad i = 1, \ldots, n. \tag{1.2}$$

Note that the right-hand side of (1.2) is well defined as the exponential of a Lebesgue integral under our assumption (see Lefèvre and Milhaud 1990). From (1.2), the marginal survival function of T_i can be represented as

$$P(T_i > t) = E(\exp[-\int_0^t r_i(u, X(u))du]), \quad i = 1, \ldots, n. \tag{1.3}$$

An alternative way to express the lifelength vector $(T_1, \ldots, T_n)$ is to use the cumulative hazard processes $\Lambda_i(t) = \int_0^t r_i(u, X(u))du$, $i = 1, \ldots, n$. Let $S_1, \ldots, S_n$ be independent of X and of each other and have the standard (that is, mean 1) exponential distribution. Then the lifelength of component i is modeled by

$$T_i = \inf\{t \geq 0 : \Lambda_i(t) > S_i\}, \quad i = 1, \ldots, n. \tag{1.4}$$

It is now evident, from (1.2) and (1.4), that the joint distribution of lifelengths $T_1, \ldots, T_n$ depends on the environment through the environmental process X and, in general, is difficult to evaluate explicitly. Çinlar, Shaked and Shanthikumar (1989) and Lefèvre and Milhaud (1990) explored the dependence of lifelengths $T_1, \ldots, T_n$ on the environmental process X and showed that if the environmental process X is associated in time (see Definition 2.2) then the lifelengths $T_1, \ldots, T_n$ are associated in the sense of Esary, Proschan and Walkup (1967). In this paper, we use a similar approach to characterize the dependence structure of component lifelengths of a system operating in a random environment where the imposed stresses have a correlated residual effect on the component failure rates.

It is well-known that the univariate shot-noise process provides a natural setting for describing the *time-dependent effects* of damage due to nontraumatic events (Cox and Isham 1980). As pointed out in Lemoine and Wenocur (1986) and in Singpurwalla and Youngren (1993), the shot-noise process model is meaningful if the imposed stresses have a residual effect on the hazard rate of an item, such as healing after a heart attack, or cracks due to fatigue which tend to close up after the material has borne a load. However, for a system of components operating in a dynamic random environment, the dependence nature of the component lifelengths is inherited from both *temporal dependence* and *spatial dependence*; that is, dependence over different time instants introduced by the event arrival process, and dependence among various components due to 'simultaneous damage' caused by an event on these components. In order to describe such correlated residual effect on the failure rates of various components, we introduce a multivariate version of the shot-noise process that incorporates both the temporal dependence and the spatial dependence effects into consideration. We then formulate the conditions on a multivariate shot-noise process under which the component lifelengths of the system operating in such an environment are positively associated or orthant dependent. Although sharing the same spirit with Çinlar, Shaked and Shanthikumar (1989) and Lefèvre and Milhaud (1990), a distinction of this work from the others in the literature is that we focus on the temporal dependence as well as spatial dependence effects and, using the orthant comparison method, we are able to compare the systems in different environments in terms of the strengths of correlations among their components.

The organization of this paper is as follows. Section 2 discusses some preliminaries on some stochastic comparison methods. A multivariate shot-noise process is introduced in Section 3. The dependence structures of this multivariate process and the related systems are discussed in Sections 3, and 4. Finally, as the byproducts of our main results, some computable bounds for certain systems are given in Section 5. Throughout this paper, the terms 'increasing' and 'decreasing' mean 'nondecreasing' and 'nonincreasing' respectively, and the measurability of sets and functions as well as the existence of expectations are assumed without explicit mention.

2 Preliminaries

In this section, we review some notions of stochastic orders and positive dependence that are relevant to our research. Most of the definitions and results discussed here can be found in Tong (1980), Lindqvist (1988) and in Shaked and Shanthikumar (1994). Let $\mathcal{E}$ be a partially ordered Polish space (that is, a complete separable metric space) with a closed partial ordering $\leq$.

Definition 2.1 1. A probability measure P on $\mathcal{E}$ is called associated if for all upper subsets U_1 and U_2 of $\mathcal{E}$, $P(U_1 \cap U_2) \geq P(U_1)P(U_2)$ (A subset $U \subseteq \mathcal{E}$ is called upper (or lower) if $x \in U$ and $x \leq y$ (or $x \geq y$) imply that $y \in U$).

2. An $\mathcal{E}$-valued random variable X is said to be larger than Y in the usual stochastic order (denoted as $X \geq_{st} Y$) if $P(X \in U) \geq P(Y \in U)$ for all upper sets $U \subseteq \mathcal{E}$.

Two $\mathcal{E}$-valued random variables X and Y have the same distribution (denoted as $X =_{st} Y$) if and only if $X \geq_{st} Y$ and $X \leq_{st} Y$.

Let $X = \{X_n, n \geq 0\}$ be a discrete-time stochastic process where X_n is $\mathcal{E}$-valued for all $n \geq 0$. Let $\mathcal{E}^\infty = \mathcal{E} \times \mathcal{E} \times \ldots$ be the product space of infinitely many $\mathcal{E}$'s with the usual product topology and the coordinate-wise partial ordering; that is, for any $x = (x_1, x_2, \ldots,) \in \mathcal{E}^\infty$ and $y = (y_1, y_2, \ldots,) \in \mathcal{E}^\infty$, $x \leq y$ if and only if $x_i \leq y_i$ for all $i \geq 1$. The product space $\mathcal{E}^\infty$ is again a partially ordered Polish space (see Billingsley 1968, Page 218). A process X is said to be *associated* if the probability measure P_X on $\mathcal{E}^\infty$ induced by X is associated in the sense of Definition 2.1 (1).

For a continuous-time $\mathcal{E}$-valued process $X = \{X(t), t \geq 0\}$, we need to consider the space $D_{\mathcal{E}}[a, b]$, the space of all functions from the real interval $[a, b]$ to $\mathcal{E}$ which are right continuous and have left limits. The space $D_{\mathcal{E}}[a, b]$ is a partially ordered Polish space with the Skorohod metric and the partial order $\leq$ defined as $x \leq y$ if $x(t) \leq y(t)$ for all $t \in [a, b]$ where $x, y \in D_{\mathcal{E}}[a, b]$. Note that the $D_{\mathcal{E}}[0, \infty)$ is still Polish, where we use the Stone (1963) modification of Skorohod's metric (Kamae, Krengel and O'Brien 1977). A continuous-time process $X = \{X(t), t \geq 0\}$ is said to be *associated* if the probability measure P_X on $D_{\mathcal{E}}[0, \infty)$ induced by X is associated in the sense of Definition 2.1 (1).

In order to verify the association property of a stochastic process, we first need the following,

Definition 2.2 Let $\mathcal{R}$ be the set of all real numbers.

1. The $\mathcal{R}^n$-valued random vectors $\mathbf{Z}_1, \ldots, \mathbf{Z}_k$ are said to be associated if the probability measure P on the space $\mathcal{R}^{kn}$ induced by the random vector $(\mathbf{Z}_1, \ldots, \mathbf{Z}_k)$ is associated in the sense of Definition 2.1 (1).

2. (Esary and Proschan 1970) An $\mathcal{R}^n$-valued stochastic process $\{\mathbf{Z}(t), t \in \mathcal{I}\}$ is said to be associated in time if for any set $\{t_1, \dots, t_k\} \subseteq \mathcal{I}$, the random vectors $\mathbf{Z}(t_1), \dots, \mathbf{Z}(t_k)$ are associated in the sense of (1) (Here $\mathcal{I}$ is either $\mathcal{R}_+ = [0, \infty)$ or $\mathcal{Z}_+ = \{0, 1, 2, \dots, \}$).

It was shown in Lindqvist (1988) that a probability measure P is associated on a partially ordered Polish space $\mathcal{E}$ if and only if $\int fgdP \geq (\int fdP)(\int gdP)$ for all real bounded increasing functions f and g defined on $\mathcal{E}$. Therefore, the probability measure P induced by an $\mathcal{R}^n$-valued random vector $\mathbf{Z} = (Z_1, \dots, Z_n)$ is associated if and only if the covariance $Cov(f(\mathbf{Z}), g(\mathbf{Z})) \geq 0$, for all increasing functions f and g (Esary, Proschan and Walkup 1976). Note also that if the probability measure $P_{\mathbf{Z}}$ induced by a process $\mathbf{Z} = \{\mathbf{Z}(t), t \in \mathcal{I}\}$ on an appropriate state space is associated if and only if

$$E(f(\mathbf{Z})g(\mathbf{Z})) \geq Ef(\mathbf{Z})Eg(\mathbf{Z}) \tag{2.1}$$

for all bounded increasing functionals f and g. This (2.1) implies that the process $\mathbf{Z}$ is associated in time. Conversely, Lindqvist (1987, 1988) has showed that if $\mathbf{Z}$ is associated in time, then (2.1) holds under certain conditions.

Some properties of association are summarized below.

Lemma 2.3 Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be partially ordered Polish spaces.

1. If a probability measure P on $\mathcal{E}_1$ is associated and $f : \mathcal{E}_1 \to \mathcal{E}_2$ is increasing, then the induced measure Pf^{-1} on $\mathcal{E}_2$ is associated.

2. If a probability measure P_1 on $\mathcal{E}_1$ is associated and a probability measure P_2 on $\mathcal{E}_2$ is associated, then the usual product measure $P_1 \times P_2$ on $\mathcal{E}_1 \times \mathcal{E}_2$ is associated.

3. Let X be an $\mathcal{E}_1$-valued random variable and Y be an $\mathcal{E}_2$-valued random variable. If X is associated, $(Y \mid X = x)$ is associated for all x and $E[f(Y) \mid X = x]$ is increasing in x for all increasing functional f, then Y is associated.

Proof: The first two results can be found in Lindqvist (1988). Jogdeo (1978) obtained (3) for the real space. Lindqvist (1988) obtained a stronger version of (3) using monotone and associated kernels. We now prove (3) directly. Note that since $\mathcal{E}_1$ and $\mathcal{E}_2$ are Polish, there exists a regular conditional probability measure $P(Y \in B \mid X = x)$ on $\mathcal{E}_2$. Let P_X be the probability measure on $\mathcal{E}_1$ induced by X. For any two upper sets U and V of $\mathcal{E}_2$, we have,

$$P(Y \in U \cap V) = \int P(Y \in U \cap V \mid X = x) P_X(dx)$$

$$\geq \int P(Y \in U \mid X = x) P(Y \in V \mid X = x) P_X(dx)$$

$$\geq \int P(Y \in U \mid X = x) P_X(dx) \int P(Y \in V \mid X = x) P_X(dx) = P(Y \in U) P(Y \in V),$$

where the first inequality follows from the association property of $(Y \mid X = x)$, and the second inequality follows from the association property of X and the monotonicity of $P(Y \in B \mid X = x)$ for any upper set B. ■

As indicated in Lindqvist (1988), Definition 2.1 (1) is equivalent to the corresponding statement with the upper set being replaced by the lower set. Thus if the mapping f in Lemma 2.3 (1) and (3) is replaced by a decreasing mapping, the results still hold.

Many different weaker notions of stochastic orders and positive dependence have been also introduced and studied in the literature, see, for example, Tong (1980) and Shaked and Shanthikumar (1994) for more detail. The following concepts that are weaker than Definition 2.1 are frequently used later.

Definition 2.4 Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ where X_i and Y_i are $\mathcal{R}$-valued random variables, $i = 1, \ldots, n$.

1. $\boldsymbol{X}$ is said to be larger (smaller) than $\boldsymbol{Y}$ in the upper (lower) orthant order, denoted as $\boldsymbol{X} \geq_{uo} (\leq_{lo}) \boldsymbol{Y}$, if $P(\boldsymbol{X} > \boldsymbol{x}) \geq P(\boldsymbol{Y} > \boldsymbol{x})$ $(P(\boldsymbol{X} \leq \boldsymbol{x}) \geq P(\boldsymbol{Y} \leq \boldsymbol{x}))$ for all $\boldsymbol{x} \in \mathcal{R}^n$. Here and in the sequel, an inequality of two vectors means component-wise inequalities.

2. $\boldsymbol{X}$ is said to be more positively upper (lower) orthant dependent than $\boldsymbol{Y}$ if $\boldsymbol{X} \geq_{uo} \boldsymbol{Y}$ $(\boldsymbol{X} \leq_{lo} \boldsymbol{Y})$ and $X_i =_{st} Y_i$ for each i.

3. $\mathbf{X}$ is said to be positively upper (lower) orthant dependent (PUOD, PLOD) if $\boldsymbol{X} \geq_{uo} \boldsymbol{X}^I$ $(\boldsymbol{X} \leq_{lo} \boldsymbol{X}^I)$, where $\boldsymbol{X}^I = (X_1^I, \ldots, X_n^I)$ denotes a vector of real random variables such that $X_i =_{st} X_i^I$ for each i and $X_1^I, \ldots, X_n^I$ are independent.

It is well known (see, for example, Shaked and Shanthikumar 1994) that

$$\boldsymbol{X} \geq_{st} \boldsymbol{Y} \Longrightarrow \boldsymbol{X} \geq_{uo} \boldsymbol{Y} \text{ and } \boldsymbol{X} \geq_{lo} \boldsymbol{Y}. \tag{2.2}$$

Note that the dependence comparisons described in Definition 2.4 (2) emphasize the comparison on the dependence structures by separating the marginals from the consideration. Some properties regarding orthant comparisons are summarized below and will be used later.

Lemma 2.5 Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two n-dimensional random vectors.

1. $\boldsymbol{X} \geq_{uo} \boldsymbol{Y}$ ($\boldsymbol{X} \leq_{lo} \boldsymbol{Y}$) if and only if $E[\prod_{i=1}^n f_i(X_i)] \geq E[\prod_{i=1}^n f_i(Y_i)]$ for every collection $\{f_1, \dots, f_n\}$ of univariate nonnegative increasing (decreasing) functions.

2. The orthant orders are closed under increasing transformations. That is, if $\boldsymbol{X} \leq_{lo} (\leq_{uo}) \boldsymbol{Y}$ and f_i is a right continuous increasing function, $i = 1, \dots, n$, then $(f_1(X_1), \dots, f_n(X_n)) \leq_{lo} (\leq_{uo}) (f_1(Y_1), \dots, f_n(Y_n))$.

3. Let $\boldsymbol{U}$ and $\boldsymbol{V}$ be another two n-dimensional random vectors such that $\boldsymbol{X} \leq_{lo}$ $(\leq_{uo})$ $\boldsymbol{Y}$ and $\boldsymbol{U} \leq_{lo} (\leq_{uo}) \boldsymbol{V}$. In addition, $\boldsymbol{X}$ and $\boldsymbol{U}$ are independent and $\boldsymbol{Y}$ and $\boldsymbol{V}$ are independent. Let $f_i : \mathcal{R}^2 \to \mathcal{R}$ be a right continuous increasing function, $i = 1, 2, \dots, n$. Then $(f_1(X_1, U_1), \dots, f_n(X_n, U_n)) \leq_{lo} (\leq_{uo}$ $) (f_1(Y_1, V_1), \dots, f_n(Y_n, V_n))$.

The following facts are easy to verify (see, for example, Tong 1980, and Szekli 1995):

$$\boldsymbol{X} \text{ is (positively) associated} \quad \Longrightarrow \quad \boldsymbol{X} \text{ is PUOD and PLOD} \Longrightarrow \mathrm{Cov}(\boldsymbol{X}) \geq 0, \tag{2.3}$$

where $\mathrm{Cov}(\boldsymbol{X})$ is the covariance matrix of $\boldsymbol{X}$. Note that the association property of a random vector implies that its joint distribution and joint survival functions can be bounded below by the products of its marginal distribution and marginal survival functions respectively. While the PLOD (PUOD) property of a random vector means that its joint distributions (joint survival functions) can be bounded below by the product of its marginal distributions (marginal survival functions).

Fortuin, Kasteleyn and Ginibre (1971) introduced the so-called FKG-inequality which implies the association property of a probability measure. A probability measure P on a finite distributive lattice $\mathcal{E}$ is said to satisfy the FKG inequality if for all $x, y \in \mathcal{E}$,

$$P(x)P(y) \leq P(x \wedge y)P(x \vee y), \tag{2.4}$$

where $x \wedge y$ is the greatest lower bound of x and y, and $x \vee y$ is the least upper bound of x and y. Fortuin, Kasteleyn and Ginibre (1971) proved that if P satisfies (2.4), then P is associated. Also see Karlin and Rinott (1980) for a generalization of this result.

Xu and Li (2000), and Li and Xu (2001) introduced the so-called IE-transforms that can be used to construct a sequence of n-dimensional PLOD (or PUOD) random vectors from a vector of n independent random variables. The reader is referred to these papers for more detail.

3 A Multivariate Shot-Noise Process

In this section, we introduce a multivariate shot-noise process and use such a process to model (multivariate) hazard rates of lifelengths of components of a system operating in dynamic environments. Some properties on dependence structures of such multivariate shot-noise processes and related systems are also given here.

We first describe the univariate shot-noise process. Suppose an item operates in a random environment consisting of a series of events ('shots') whose effect is to induce stresses of unknown varying magnitudes A_k, $k \geq 0$, on the item. The shot are assumed to occur over time according to a Poisson process with a known rate μ, and whenever a stress of magnitude A_k is induced at an epoch τ_k, then its contribution to the item's hazard rate $X(\tau_k + u)$ at time $\tau_k + u$ is $A_k h(u)$, where the *attenuation function* $h : \mathcal{R} \to \mathcal{R}$ is positive and decreasing for nonnegative arguments with $h(u) = 0, u < 0$. Thus the item's hazard rate at time t is given by

$$X(t) = \sum_{k=0}^{\infty} A_k h(t - \tau_k), \tag{3.1}$$

where $\{\tau_k, k \geq 0\}$ are the epochs at which the shots occur with respective magnitudes $\{A_k, k \geq 0\}$. The process $\{X(t),\, t \geq 0\}$ is called a shot-noise process. See Cox and Isham (1980) for more details on the univariate shot-noise process. Under some conditions on an item with shot-noise hazard rate process, Lemoine and Wenocur (1986) (and also see Singpurwalla and Youngren (1993)) obtained the survival function of the item in terms of its Laplace transform. Special cases of their results provide us with an alternative motivation for some well-known distributions (such as Pareto) as failure models, and also yield some new families of failure distributions.

In order to model the hazard rates of lifelengths of components in a system sharing the same randomly varying environment, one has to incorporate the shot's 'simultaneous' effect on various components in the system. Singpurwalla and Youngren (1991) introduced the *environmental factor function* to handle such effect and their idea is the following. Consider the lifelengths T_1, T_2 of a two component system. When they operate in the laboratory (design) environment, T_i has a specified failure rate $\lambda_i(t)$, $t \geq 0, i = 1, 2$. Suppose that the system is installed in an operating (use) environment comprising several covariates (stresses) whose presence and intensities change over time. Assume that the net effect of the operating environment is to modulate $\lambda_i(t)$ to $\lambda_i(t)X(t)$, where $X(t)$ is a suitable stochastic process (called the environmental factor function). As indicated in Singpurwalla and Youngren (1993), one possible choice for $X(t)$ is the shot-noise process as described in (3.1).

However, in general, the environmental stress and its residual effect may not always have the same impact on the various components in a system. This motivates us to introduce a multivariate shot-noise process, and to extend the above model to handle a more general situation. We describe our model in the context of nontraumatic shock environment. Let $\tau = \{\tau_k, k \geq 0\}$ be a point process on $\mathcal{R}_+$, where random variables $0 \leq \tau_0 < \tau_1 < \ldots$ are defined on the same probability space. The point processes considered in this paper are assumed to have no multiplicities and have no limiting points (non-explosive). Consider a system of n components that operates in a dynamic environment whose net effect is to inflict shocks of varying magnitudes. Suppose that shocks occur according to the point process τ, and the consequence of each shock is a set of simultaneous stresses on the components contributing to their failure rates. Upon an arrival epoch τ_k, the contribution of this shock to the hazard rate of component i at time $\tau_k + u$ is $A^i_k h_i(u)$, for each $i = 1, \ldots, n$, where the attenuation function $h_i : \mathcal{R} \to \mathcal{R}$ is positive and decreasing for nonnegative arguments with $h_i(u) = 0, u < 0$. As such, the total contribution to the hazard rate of component i, $i = 1, \ldots, n$, by time t from the environmental shocks is given by

$$X_i(t) = \sum_{k=0}^{\infty} A_k^i h_i(t - \tau_k), \quad i = 1, \ldots, n. \tag{3.2}$$

Note that for each k, $A_k^1, \ldots, A_k^n$ may be dependent, but we assume in the sequel that the sequence of nonnegative random vectors $\mathbf{A} = \{\mathbf{A}_k = (A_k^1, \ldots, A_k^n), k = 0, 1, \ldots,\}$ are independent and identically distributed, and also independent of the point process τ. The n-dimensional process $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))$ is called a multivariate shot-noise process.

Since $A_k^1, \ldots, A_k^n$ are nonnegative and $h_1(u), \ldots, h_n(u)$ are nonnegative functions, it is sufficient in the sequel to assume that the sums in (3.2) converge. Daley (1971) introduced a multivariate version of shot-noise process where τ is a homogeneous Poisson process in $\mathcal{R}^n$ and $A_k^j = W_k$ for all $j = 1, \ldots, n$. Daley (1971) also discussed the conditions under which the multivariate shot-noise process is absolutely convergent. Daley's motivation was in the description of gravitational fields, while ours is in modeling components of a system sharing the same dynamic shock environment.

Let T_i be the lifelength of component i of a system in a multivariate shot-noise process environment described as in (3.2). Given $\mathbf{X} = \{\mathbf{X}(t), t \geq 0\}$, the lifelengths $T_1, T_2, \ldots, T_n$ are independent and

$$\lim_{u \downarrow 0} \frac{1}{u} P(t < T_i \leq t + u \mid T_i > t, \mathbf{X}) = r_i(t, X_i(t)), \qquad i = 1, 2, \ldots, n, \tag{3.3}$$

where each r_i is a positive continuous function on $\mathcal{R}_+ \times \mathcal{R}$. As we commented in Section 1, our failure model (3.3) is similar to that of Lefèvre and Milhaud (1990), but the processes $X_i(t)$'s may be different.

A failure model described by (3.3) and (3.2) is rather general, and the following two special cases have better interpretation in some applications.

1. Similar to Singpurwalla and Youngren (1993), this special case of (3.3) describes that a system operates in a dynamic environment that needs not be

the same as the design environment. Assume that, given a multivariate shot-noise process $\mathbf{X} = \{\mathbf{X}(t), t \geq 0\}$ that is described by (3.2),

$$\lim_{u \downarrow 0} \frac{1}{u} P(t < T_i \leq t + u \mid T_i > t, \mathbf{X}) = \lambda_i(t) X_i(t), \qquad i = 1, 2, \ldots, n, \tag{3.4}$$

where $\lambda_i(t)$, $i = 1, \ldots, n$, can be viewed as the failure rate of component i in an ideal design environment.

2. We now consider a special case of the general multivariate process $\{\mathbf{X}(t), t \geq 0\}$ defined in (3.2). Let $h(t) = 1$ for $t \geq 0$ and zero otherwise, then (3.2) becomes

$$X_i(t) = \sum_{k=0}^{N(t)} A_k^i, \quad i = 1, \ldots, n, \tag{3.5}$$

where $N(t) = \max\{k : \tau_k \leq t\}$ is a counting process generated by the point process τ. This process is a multivariate cumulative damage process where shocks arrive at the system according to the point process τ. Therefore, (3.5) and (3.3) provide a cumulative shock model for the systems operating in dynamic environments.

Remark 3.1 In the second special case above, if we assume that A_k^i is a Bernoulli random variable for all k and i, then we obtain the process that was studied extensively in Li and Xu (2000, 2001). Using newly developed *majorization of weighted trees* (Xu and Li 2000), Li and Xu (2000, 2001) studied the dependence structures of shock models and queueing models with certain synchronized constraints. However, the distinction between the shock model scenario of Li and Xu (2001) and the scenario described above is the following. In Li and Xu (2001), the amount of some physically observable entity, such as amount of damage, is modeled by a multivariate process, whereas here an unobservable entity, the hazard rate, is so modeled. In the former, a component fails when the observable entity reaches a threshold, whereas here there is no parallel notion of a threshold for the hazard rate. ■

The dependence structures of a multivariate shot-noise process $\mathbf{X}(t)$ and the related systems described above are determined by the dependence nature of $\mathbf{A}_k$ over the various components and the autocorrelation structure of the point process τ. Intuitively, if the simultaneous damages $A_k^1, \ldots, A_k^n$ on the various components are spatially dependent in some sense, and the shock arrival epochs $\{\tau_0, \tau_1, \dots, \}$ are temporally dependent in some sense, then one would expect the lifelengths $T_1, \dots, T_n$ are dependent in some sense. We begin our discussion on dependence structures with the simpler models.

Lemma 3.2 Let $\mathbf{x}(t; \mathbf{a}, \mathbf{z}) = (x_1(t; \mathbf{a}, \mathbf{z}), \dots, x_n(t; \mathbf{a}, \mathbf{z}))$ be defined as, for $i = 1, \dots, n$,

$$x_i(t; \mathbf{a}, \mathbf{z}) = \int_0^t \lambda_i(u)(\sum_{k=0}^{\infty} a_k^i h_i(u - z_k))du,$$

where $\mathbf{a} = \{\mathbf{a}_k = (a_k^1, \dots, a_k^n), k \geq 0\}$ is a sequence of vectors of nonnegative real numbers, and $\mathbf{z} = \{z_k, k \geq 0\}$ is a sequence of nonnegative real numbers, and the attenuation function $h_i : \mathcal{R} \to \mathcal{R}$ is positive and decreasing for nonnegative arguments with $h_i(u) = 0, u < 0$. If $\lambda_i(t)$ is nonnegative and decreasing in t, $i = 1, \dots, n$, then $x_i(t; \mathbf{a}, \mathbf{z})$ is increasing in $\mathbf{a}$, and decreasing in $\mathbf{z}$ with respect to the usual product order.

Proof: From the monotone convergence theorem, we have, for $i = 1, \dots, n$,

$$x_i(t; \mathbf{a}, \mathbf{z}) = \int_0^t \lambda_i(u)(\sum_{k=0}^{\infty} a_k^i h_i(u - z_k))du = \sum_{k=0}^{\infty} a_k^i (\int_0^t \lambda_i(u) h_i(u - z_k)du). \quad (3.6)$$

Obviously, the function $x_i(t; \mathbf{a}, \mathbf{z})$ is increasing in $\mathbf{a}$ on the space $\mathcal{R}^\infty$ with respect to the usual product order. We next claim that these functions are decreasing in $\mathbf{z}$ with respect to the usual product order. Let $\mathbf{z} = \{z_k, k \geq 0\}$ and $\bar{\mathbf{z}} = \{\bar{z}_k, k \geq 0\}$ be two sequences such that $0 \leq z_k \leq \bar{z}_k$ for each k. Denote $d_k = \bar{z}_k - z_k$. Since $\lambda_i(t)$ is decreasing in t, we have, for $\bar{z}_k \leq t$, ,

$$\int_0^t \lambda_i(u) h_i(u - z_k)du = \int_{d_k}^{t+d_k} \lambda_i(v - d_k) h_i(v - \bar{z}_k)dv \geq \int_{d_k}^{t+d_k} \lambda_i(v) h_i(v - \bar{z}_k)dv \geq$$

$$\int_{d_k}^{t} \lambda_i(v) h_i(v - \bar{z}_k)dv \geq \int_{\bar{z}_k}^{t} \lambda_i(v) h_i(v - \bar{z}_k)dv = \int_0^t \lambda_i(u) h_i(u - \bar{z}_k)du, \quad k = 0, 1, \dots.$$

Therefore,

$$x_i(t; \mathbf{a}, \mathbf{z}) \geq x_i(t; \mathbf{a}, \bar{\mathbf{z}}).$$

Hence $x_i(t; \mathbf{a}, \mathbf{z})$ is decreasing in $\mathbf{z}$ with respect to the usual product order. ■

Theorem 3.3 Let $\mathbf{T} = (T_1, \ldots, T_n)$ be the lifelengths of components of a system operating in a multivariate shot-noise process environment as described in (3.2) and (3.4) where $\lambda_i(t)$ is decreasing in t, $i = 1, \ldots, n$. If $\mathbf{A}_k = (A_k^1, \ldots, A_k^n)$ is associated and the process $\tau = \{\tau_k, k \geq 0\}$ is associated in time, then $\mathbf{T}$ is associated.

Proof: First observe that the lifelength of component i is modeled by

$$T_i = \inf\{t \geq 0 : \Lambda_i(t) > S_i\}, \quad i = 1, \ldots, n,$$

where the cumulative failure rate $\Lambda_i(t) = x_i(t; \{\mathbf{A}_k\}, \tau) = \int_0^t \lambda_i(u) X_i(u) du$, and $S_1, \ldots, S_n$ are independent of $\Lambda = \{(\Lambda_1(t), \ldots, \Lambda_n(t)), t \geq 0\}$ and of each other and have the standard exponential distribution. Thus, for any increasing function ϕ,

$$E\phi(\mathbf{T} \mid \Lambda_1(t) = x_1(t), \ldots, \Lambda_n(t) = x_n(t)) \tag{3.7}$$

is a decreasing functional with respect to $(x_1(t), \ldots, x_n(t))$ on the space $[D_{\mathcal{R}}[0, \infty)]^n$, which is also a partially ordered Polish space with the product topology and the coordinate-wise partial order. Since $(\mathbf{T} \mid \Lambda_1(t) = x_1(t), \ldots, \Lambda_n(t) = x_n(t))$ is an increasing function with respect to $(S_1, \ldots, S_n)$, which are independent, then $(\mathbf{T} \mid \Lambda_1(t) = x_1(t), \ldots, \Lambda_n(t) = x_n(t))$ is associated from Lemma 2.3 (1). Now, from Lemma 2.3 (3), if the process Λ is associated (that is, the probability measure P_Λ on $[D_{\mathcal{R}}[0, \infty)]^n$ induced by the process Λ is associated), then $\mathbf{T}$ is associated. Thus we only need to show that Λ is associated.

Since the space $\mathcal{R}^\infty$ is normally ordered with respect to the usual product order in the sense of Lindqvist (1988), the time association property of the process τ implies the probability measure P_τ induced by τ is associated (Lindqvist 1988). Since $\mathbf{A}_k$ is associated for each k, and $\mathbf{A}_{k_1}, \mathbf{A}_{k_2}, \ldots, \mathbf{A}_{k_l}$ are independent for all $k_1 \leq k_2 \leq \ldots \leq k_l$, then, from Lemma 2.3 (2), $(\mathbf{A}_{k_1}, \mathbf{A}_{k_2}, \ldots, \mathbf{A}_{k_l})$ is associated. Hence, from Definition 2.2 (2), the process $\mathbf{A} = \{\mathbf{A}_k, k = 0, 1, \ldots\}$ is associated in time. From Lindqvist (1988) again, we obtain that the probability measure $P_{\mathbf{A}}$ induced by this discrete-time process $\{\mathbf{A}_k, k = 0, 1, \ldots\}$ is associated.

Since, for fixed $\mathbf{z}$, $\mathbf{x}(\cdot; \mathbf{a}, \mathbf{z})$ is increasing from $\mathcal{R}^\infty$ to $[D_{\mathcal{R}}[0, \infty)]^n$, then from Lemma 2.3 (1), for each fixed $\mathbf{z}$, the process $\mathbf{x}(t; \mathbf{A}, \mathbf{z})$ is associated. In addition, Lemma 3.2 implies that for any decreasing functional ϕ, $E\phi(\mathbf{x}(\cdot; \mathbf{A}, \mathbf{z}))$ is increasing in $\mathbf{z}$ with respect to the usual product order. Therefore, from lemma 2.3 (3), the process $\mathbf{x}(t; \mathbf{A}, \tau)$ is associated. Hence the process Λ, and thus P_Λ is associated. ■

To illustrate the result we just obtained, we discuss the following example.

Example 3.4 Consider the multivariate shot-noise process,

$$X_i(t) = \sum_{k=0}^{\infty} A_k^i h_i(t - \tau_k), \tag{3.8}$$

where $\{\mathbf{A}_k = (A_k^1, \ldots, A_k^n), k = 1, 2, \ldots\}$ is a sequence of independent and identically distributed Bernoulli random vectors, and $\tau = \{\tau_k, k \geq 0\}$ is a renewal process. Let $\mathbf{T} = (T_1, \ldots, T_n)$ be the lifelength vector of components of a system operating in a multivariate shot-noise environment as defined in (3.2) and (3.4). Assume that the function $\lambda_i(t)$ is a constant for $i = 1, \ldots, n$.

Since each τ_k can be written as a sum of some of independent and identically distributed random variables, τ is associated in time. Now we assume that $\mathbf{A}_k$ satisfies the FKG-inequality (see (2.4))

$$P(\mathbf{A}_k = \mathbf{a})P(\mathbf{A}_k = \mathbf{b}) \leq P(\mathbf{A}_k = \mathbf{a} \vee \mathbf{b})P(\mathbf{A}_k = \mathbf{a} \wedge \mathbf{b}) \tag{3.9}$$

for all $\mathbf{a} = (a_1, \ldots, a_n), \mathbf{b} = (b_1, \ldots, b_n) \in \{0, 1\}^n$, where $\mathbf{a} \vee \mathbf{b} = (\max\{a_1, b_1\}, \ldots, \max\{a_n, b_n\})$ and $\mathbf{a} \wedge \mathbf{b} = (\min\{a_1, b_1\}, \ldots, \min\{a_n, b_n\})$. Therefore, $\mathbf{A}_k$ is associated. Thus, from Theorem 3.3, we obtain that $\mathbf{T}$ is associated.

For example, consider a system with two components where (A_k^1, A_k^2) has the following distribution,

$$P((A_k^1, A_k^2) = (1, 1)) = P((A_k^1, A_k^2) = (0, 0)) = 1/8,$$

$$P((A_k^1, A_k^2) = (0, 1)) = 3/4, P((A_k^1, A_k^2) = (1, 0)) = 0.$$

It is easy to verify that (A_k^1, A_k^2) satisfies the FKG inequality, and so the corresponding lifelengths T_1, T_2 are associated. ■

Note that the approach used in Theorem 3.3 is to establish directly the association property of the cumulative hazard process Λ. An alternative approach is to establish the association property of the environmental process $\mathbf{X}(t)$, and then to obtain the association property of the lifelength vector $\mathbf{T}$ of a system operating in such random environment (Çinlar, Shaked and Shanthikumar 1989, Lefèvre and Milhaud 1990). Using such approach, we can study the dependence structure for a more general model (3.3), but a trade-off is that we have to impose a stronger condition on the point process τ. The following result is needed for employing this approach.

Proposition 3.5 Let $\mathbf{T} = (T_1, \ldots, T_n)$ be the lifelength vector of components of a system operating in a multivariate shot-noise environment as defined in (3.2) and (3.3). Assume that the functions $r_i(t, x)$, $i = 1, \ldots, n$, are all increasing (or decreasing) in x for each t. If the probability measure induced by $\mathbf{X} = \{\mathbf{X}(t), t \geq 0\}$ is associated, then $\mathbf{T}$ is associated.

Proof: Since $r_i(t, x)$ is increasing (decreasing) in x, then the cumulative hazard rate process of component i

$$\Lambda_i(t) = \int_0^t r_i(u, X_i(u))du$$

is an increasing (decreasing) function of the process $\mathbf{X} = \{\mathbf{X}(t), t \geq 0\}$ on the space $[D_{\mathcal{R}}[0, \infty)]^n$. Since the probability measure $P_{\mathbf{X}}$ on $[D_{\mathcal{R}}[0, \infty)]^n$ induced by the process $\mathbf{X}$ is associated, then, from Lemma 2.3 (1), the probability measure P_Λ induced by the joint cumulative hazard process $\Lambda = \{\Lambda(t) = (\Lambda_1(t), \ldots, \Lambda_n(t)), t \geq 0\}$ is associated. Now we employ a similar argument as in Theorem 3.3. Since $(\mathbf{T} \mid \Lambda(t) = \mathbf{x}(t))$ is associated and for any decreasing function f, $Ef(\mathbf{T} \mid \Lambda(t) = \mathbf{x}(t))$ is increasing with respect to $\mathbf{x}(t)$ in the space $[D_{\mathcal{R}}[0, \infty)]^n$ (see (3.7)), then from Lemma 2.3 (3), we obtain that $\mathbf{T}$ is associated. ■

Çinlar, Shaked and Shanthikumar (1989), and Lefèvre and Milhaud (1990) obtained the same conclusion as Proposition 3.5 under the (weaker) condition that the environmental process $\{\mathbf{X}(t), t \geq 0\}$ is associated in time. However, the current Proposition 3.5 is sufficient for development of our results. As a direct consequence of Proposition 3.5, we obtain the association property of the lifelengths of components of a system with the environmental process described in (3.5).

Theorem 3.6 Let $\mathbf{T} = (T_1, \ldots, T_n)$ be the lifelength vector of components of a system operating in a multivariate shot-noise environment as defined in (3.5) and

(3.3). Assume that the functions $r_i(t, x)$, $i = 1, \ldots, n$, are all increasing (decreasing) in x for each t. If $\mathbf{A}_k = (A_k^1, \ldots, A_k^n)$ is associated and the process τ is associated in time, then $\mathbf{T}$ is associated.

Proof: Let P_N be the probability measure induced by the counting process $N = \{N(t), t \geq 0\}$ and P_τ be the probability measure induced by the point process τ. Consider $\phi(\mathbf{z}) = \{n(t), t \geq 0\} \in D_{\mathcal{Z}_+}[0, \infty)$ where $\mathbf{z} = \{z_k, k \geq 0\} \in \mathcal{R}_+^\infty$ and

$$n(t) = \max\{k : z_k \leq t\}. \tag{3.10}$$

Clearly, ϕ is an decreasing mapping from $\mathcal{R}_+^\infty$ to $D_{\mathcal{Z}_+}[0, \infty)$, and $P_N = P_\tau \phi^{-1}$. Since τ is associated in time, then from Lindqvist (1988), P_τ is associated. Therefore, from Lemma 2.3 (1), we obtain that P_N is also associated.

Since $\mathbf{A}_k$ is associated, and $\{\mathbf{A}_k, k \geq 0\}$ is a sequence of independent random vectors, then, as we have shown in Theorem 3.3, the process $\{\mathbf{A}_k, k \geq 0\}$ is associated. Therefore, from Lemma 2.3 (1), the process $\{\sum_{k=0}^m \mathbf{A}_k, m \geq 0\}$ is associated. In addition, $\sum_{k=0}^m \mathbf{A}_k$ is increasing with respect to the coordinate-wise order, almost surely as m increases. Thus, from Lemma 2.3 (3), we obtain that $\mathbf{X} = \{\sum_{k=0}^{N(t)} \mathbf{A}_k, t \geq 0\}$ is associated. Hence, from Proposition 3.5, $\mathbf{T}$ is associated. ■

For example, as illustrated in Example 3.4, a renewal process τ is associated in time. Thus, if $\mathbf{A}_k$ satisfies the FKG-inequality (3.9), then the corresponding lifelengths $\mathbf{T}$ as defined in (3.5) and (3.3) is associated.

In order to analyze the dependence structure of a system operating in a general multivariate shot-noise environment as defined in (3.2) and (3.3), we need to discuss the association property of a point process with respect to thinning order (Burton and Franzosa 1990). As illustrated in the literature (see, for example, Shaked and Szekli 1995), a point process $\tau = \{\tau_k, k \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ can be viewed as a random measure. Each realization of $\{\tau_k, k \geq 0\}$ can be treated as a measure on $\mathcal{R}_+$,

$$N(\omega) = \sum_{k \geq 0} \delta_{\tau_k(\omega)}. \tag{3.11}$$

where δ_t denotes the atomic measure concentrated at t, that is, $\delta_t(B) = 1$ if $t \in B$ and zero otherwise, for all bounded Borel sets B of $\mathcal{R}_+$. Measures of this type are integer valued, and belong to the set $\mathcal{N}$ of all integer valued Radon (that is, finite on bounded sets) measures on $\mathcal{R}_+$, which is equipped with the vague topology to be a Polish space (Burton and Franzosa 1990). In other words, a point process τ can be viewed as a random element of $\mathcal{N}$ (random measure). As discussed in Shaked and Szekli (1995), such definition of a point process provides us a 'global' description of the process. Note that the space $\mathcal{N}$ is a Polish space equipped with the following partial order,

$$\mu \le \nu \text{ if } \mu(B) \le \nu(B), \tag{3.12}$$

for all bounded Borel sets B, where $\mu, \nu \in \mathcal{N}$.

Since $\mathcal{N}$ is naturally identified with the set of all finite or infinite configurations of points (including multiplicities) in $\mathcal{R}_+$ without limit points, the partial order in (3.12) is equivalent to the notion of thinning. Consider two increasing sequences $\mathbf{z} = \{z_k\}$ and $\bar{\mathbf{z}} = \{\bar{z}_k\}$ with $z_k \le z_{k+1}$ and $\bar{z}_k \le \bar{z}_{k+1}$. $\mathbf{z}$ is said to be a thinning of $\bar{\mathbf{z}}$, denoted as $\mathbf{z} \subset \bar{\mathbf{z}}$, if there exists an one-to-one mapping T from the index set of $\mathbf{z}$ to the index set of $\bar{\mathbf{z}}$ such that $z_k = \bar{z}_{T(k)}$. Thus, if $\mathbf{z} \subset \bar{\mathbf{z}}$, then $\bar{\mathbf{z}}$ is 'finer' than $\mathbf{z}$. It is easy to verify that the thinning $\subset$ is a closed partial order. Let

$$\mu_{\mathbf{z}} = \sum_{k\ge 0} \delta_{z_k}, \quad \mu_{\bar{\mathbf{z}}} = \sum_{k\ge 0} \delta_{\bar{z}_k}.$$

Clearly $\mu_{\mathbf{z}} \le \mu_{\bar{\mathbf{z}}}$ in the sense of (3.12) if and only if $\mathbf{z}$ is a thinning of $\bar{\mathbf{z}}$.

We say that a point process $\tau = \{\tau_k\}$ is associated with respect to thinning order if the probability measure P_τ on $\mathcal{N}$ induced by τ is associated with respect to (3.12) (Burton and Franzosa 1990). The following lemma is needed for characterizing the dependent structures of a multivariate shot-noise process (3.2) and the related system (3.3). Let $\mathbf{x}(t;\mathbf{a},\mathbf{z}) = (x_1(t;\mathbf{a},\mathbf{z}), x_2(t;\mathbf{a},\mathbf{z}), \ldots, x_n(t;\mathbf{a},\mathbf{z}))$ such that

$$x_i(t;\mathbf{a},\mathbf{z}) = \sum_{k=0}^{\infty} a_k^i h_i(t - z_k), \tag{3.13}$$

where $\mathbf{a} = \{\mathbf{a}_k = (a_k^1, \ldots, a_k^n), k \ge 0\}$, and $\mathbf{z} = \{z_k, k \ge 0\}$.

Lemma 3.7 If for each i the function h_i is positive and decreasing for nonnegative arguments and $h_i(u) = 0$ for $u < 0$, then for all increasing functionals g, $Eg(\mathbf{x}(\cdot;\mathbf{A},\mathbf{z}))$ is increasing in $\mathbf{z}$ with respect to the thinning order, where $\mathbf{A} =$

$\{\mathbf{A}_k = (A_k^1, \ldots, A_k^n), k = 0, 1, \ldots,\}$ are independent and identically distributed nonnegative random vectors.

Proof: Consider two sequences $\mathbf{z} = \{z_k\}$ and $\bar{\mathbf{z}} = \{\bar{z}_k\}$ with $z_k \le z_{k+1}$ and $\bar{z}_k \le \bar{z}_{k+1}$, such that $\mathbf{z} \subset \bar{\mathbf{z}}$. That is, $z_k = \bar{z}_{T(k)}$ for each k, and T is one-to-one. Without loss of generality, we assume that $\{\mathbf{A}_k\}$ are defined on the same probability space as τ. Let $\{\mathbf{B}_k = (B_k^1, \ldots, B_k^n), k = 0, 1, \ldots,\}$ be another sequence of independent and identically distributed random vectors defined on this probability space such that $\{\mathbf{B}_k\}$ and $\{\mathbf{A}_k\}$ are independent and $\mathbf{B}_k$ and $\mathbf{A}_k$ have the same distribution. Define

$$
\begin{aligned}
(\bar{A}_j^1, \ldots, \bar{A}_j^n) &= (A_k^1, \ldots, A_k^n), \text{ if } j = T(k) \\
&= (B_j^1, \ldots, B_j^n), \text{ otherwise.}
\end{aligned}
$$

Therefore, the random variable

$$
x_i(t; \{\bar{\mathbf{A}}_k\}, \bar{\mathbf{z}}) = \sum_{j=0}^{\infty} \bar{A}_j^i h_i(t - \bar{z}_j) = \sum_{k=0}^{\infty} A_k^i h_i(t - \bar{z}_{T(k)}) + \sum_{j \notin T(\mathcal{Z}_+)} B_j^i h_i(t - \bar{z}_j),
$$

where $\mathcal{Z}_+ = \{0, 1, 2, \ldots,\}$. Thus we obtain that $\mathbf{x}(t; \{\bar{\mathbf{A}}_k\}, \bar{\mathbf{z}})$ and $\mathbf{x}(t; \{\mathbf{A}_k\}, \bar{\mathbf{z}})$ have the same distribution and for all t,

$$
\mathbf{x}(t; \{\bar{\mathbf{A}}_k\}, \bar{\mathbf{z}}) \ge \mathbf{x}(t; \{\mathbf{A}_k\}, \mathbf{z}).
$$

Therefore, for any increasing functional g, $Eg(\mathbf{x}(\cdot; \mathbf{A}, \bar{\mathbf{z}})) = Eg(\mathbf{x}(\cdot; \{\bar{\mathbf{A}}_k\}, \bar{\mathbf{z}})) \ge Eg(\mathbf{x}(\cdot; \mathbf{A}, \mathbf{z}))$. ■

Theorem 3.8 Let $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))$ be a multivariate shot-noise process as defined in (3.2). If $\mathbf{A}_k = (A_k^1, \ldots, A_k^n)$ is associated and the process τ is associated with respect to thinning, then the probability measure $P_{\mathbf{X}}$ induced by $\mathbf{X} = \{\mathbf{X}(t), t \ge 0\}$ is associated.

Proof: Since $\mathbf{A}_k$ is associated for each k, then from the proof of Theorem 3.3, we obtain that the probability measure $P_{\mathbf{A}}$ induced by the discrete-time process $\{\mathbf{A}_k, k = 0, 1, \ldots\}$ is associated. Since the attenuation function h_i is nonnegative, then for every t, the function $x_i(t; \mathbf{a}, \mathbf{z})$ defined as in (3.13) is increasing in $\mathbf{a} = \{\mathbf{a}_k\}$

on the space $\mathcal{R}^\infty$ with respect to the usual product order. From Lemma 2.3 (1), the probability measure induced by the process $\{\mathbf{x}(t;\{\mathbf{A}_k\},\mathbf{z}), t \geq 0\}$ is associated.

From Lemma 3.7, for all increasing functional g, $Eg(\mathbf{x}(\cdot;\{\mathbf{A}_k\},\mathbf{z}))$ is increasing in $\mathbf{z}$ with respect to the thinning order. Thus, from Lemma 2.3 (3), we obtain that the process $\mathbf{X} = \{\mathbf{X}(t), t \geq 0\}$ is associated. ■

Using Proposition 3.5 and Theorem 3.8, we now obtain

Corollary 3.9 Let $\mathbf{T} = (T_1, \ldots, T_n)$ be the lifelength vector of components of a system operating in a multivariate shot-noise environment as defined in (3.2) and (3.3). Assume that the functions $r_i(t, x)$, $i = 1, \ldots, n$, are all increasing (or decreasing) in x for each t. If $\mathbf{A}_k = (A_k^1, \ldots, A_k^n)$ is associated and the process τ is associated with respect to thinning, then $\mathbf{T}$ is associated.

For example, from Burton and Franzosa (1990), a Poisson process is associated with respect to thinning order. Thus, if $\mathbf{A}_k$ satisfies the FKG-inequality (3.9), then the corresponding lifelengths $\mathbf{T}$ as defined in (3.2) and (3.3) is associated. Burton and Franzosa (1990) obtained some conditions on product densities of a point process under which the process is associated with respect to thinning order, but in general, it is difficult to verify whether a point process τ is associated with respect to thinning order. In contrast with Corollary 3.9, Theorems 3.3 and 3.6 are easier to use in applications.

4 Dependence Comparison of Systems Operating in Shot-Noise Process Environments

In this section, we discuss comparison of dependence structures of the component lifelength vectors of two systems as described in (3.3) operating in multivariate shot-noise process environments (3.2), and examine the impact of environmental shocks on the failure rates of component lifelengths.

Let $\mathcal{S}$ denote a system of n components as described in (3.3) with the multivariate shot-noise process $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))$ where

$$X_i(t) = \sum_{k=0}^{\infty} A_k^i h_i(t - \tau_k), i = 1, \ldots, n. \tag{4.1}$$

Let $\mathcal{S}'$ be another similar system of n components with the multivariate shot-noise process $\mathbf{Y}(t) = (Y_1(t), \ldots, Y_n(t))$ where

$$Y_i(t) = \sum_{k=0}^{\infty} B_k^i h_i(t - \tau_k'), i = 1, \ldots, n. \tag{4.2}$$

The following theorem describes a stochastic comparison about dependence structures of these two systems.

Theorem 4.1 Let $\mathbf{T} = (T_1, \ldots, T_n)$ and $\mathbf{T}' = (T_1', \ldots, T_n')$ be the component lifelengths of systems $\mathcal{S}$ and $\mathcal{S}'$, defined as in (3.3), with multivariate shot-noise processes $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ respectively. Assume that $r_i(t; x)$, $i = 1, \ldots, n$, are all increasing (decreasing) in x, and τ and τ' have the same distribution.

1. If $(A_k^1, \ldots, A_k^n) \leq_{lo} (B_k^1, \ldots, B_k^n)$ $((A_k^1, \ldots, A_k^n) \geq_{uo} (B_k^1, \ldots, B_k^n))$, then $\mathbf{T} \geq_{uo} \mathbf{T}'$.

2. If $(A_k^1, \ldots, A_k^n)$ is more lower (upper) orthant dependent than $(B_k^1, \ldots, B_k^n)$, then $\mathbf{T}$ is more upper orthant dependent than $\mathbf{T}'$.

Proof: We only prove the case of lower orthant comparison. The other case is similar.

(1) Let P_τ denote the probability measure on $\mathcal{R}_+^\infty$ induced by the process τ, and $P_{\mathbf{A}}$ $(P_{\mathbf{B}})$ denote the probability measure on $(\mathcal{R}^n)^\infty$ induced by $\mathbf{A} = \{\mathbf{A}_k, k \geq 0\}$ $(\mathbf{B} = \{\mathbf{B}_k, k \geq 0\})$. For any $\mathbf{z} = \{z_1, z_2, \ldots, \} \in \mathcal{R}_+^\infty$ with $z_1 < z_2 < \ldots$, denote

$$f(\mathbf{z}) = P(T_1 > t_1, \ldots, T_n > t_n \mid \tau = \mathbf{z}),$$

$$g(\mathbf{z}) = P(T_1' > t_1, \ldots, T_n' > t_n \mid \tau = \mathbf{z}).$$

Thus, we have

$$P(T_1 > t_1, \ldots, T_n > t_n) = \int f(\mathbf{z}) P_\tau(d\mathbf{z}),\ P(T_1' > t_1, \ldots, T_n' > t_n) = \int g(\mathbf{z}) P_\tau(d\mathbf{z}). \tag{4.3}$$

Since the lifelengths $T_1, \ldots, T_n$ are independent given the environmental process $\mathbf{X} = \{\mathbf{X}(t), t \geq 0\}$ (see (3.3)), we also have,

$$f(\mathbf{z}) = \int \prod_{j=1}^{n} e^{-\int_0^{t_j} r_j(s, \sum_{k=0}^{\infty} a_k^j h_j(s - z_k)) ds} P_{\mathbf{A}}(d\mathbf{a}), \tag{4.4}$$

$$g(\mathbf{z}) = \int \prod_{j=1}^{n} e^{-\int_0^{t_j} r_j(s, \sum_{k=0}^{\infty} b_k^j h_j(s-z_k))ds} P_{\mathbf{B}}(d\mathbf{b}), \tag{4.5}$$

where $\mathbf{a} = \{(a_k^1, \ldots, a_k^n), k \geq 0\}$, and $\mathbf{b} = \{(b_k^1, \ldots, b_k^n), k \geq 0\}$. To approximate (4.4) and (4.5), consider the following two integrals,

$$\int \prod_{j=1}^{n} e^{-\sum_{i=0}^{m_j} r_j(s_{ij}, \sum_{k=0}^{\infty} a_k^j h_j(s_{ij}-z_k))\Delta s_{ij}} P_{\mathbf{A}}(d\mathbf{a}),$$

$$\int \prod_{j=1}^{n} e^{-\sum_{i=0}^{m_j} r_j(s_{ij}, \sum_{k=0}^{\infty} b_k^j h_j(s_{ij}-z_k))\Delta s_{ij}} P_{\mathbf{B}}(d\mathbf{b}),$$

where $0 = s_{0j} \leq s_{1j} \leq \ldots \leq s_{m_j j} = t_j$, and $\Delta s_{ij} = s_{(i+1)j} - s_{ij}$ for $i = 0, \ldots, m_j - 1$ and $j = 1, \ldots, n$. Note that both integrands of above two integrals involve only finite number of $\mathbf{a}_k$'s or $\mathbf{b}_k$'s. Since $r_j(t; x)$ is increasing in x, then $\sum_{i=0}^{m_j} r_j(s_{ij}, \sum_{k=0}^{\infty} a_k^j h_j(s_{ij}-z_k))\Delta s_{ij}$ is increasing in a_k^j, and $\sum_{i=0}^{m_j} r_j(s_{ij}, \sum_{k=0}^{\infty} b_k^j h_j(s_{ij}-z_k))\Delta s_{ij}$ is increasing in b_k^j.

Since $\{\mathbf{A}_k, k \geq 0\}$ and $\{\mathbf{B}_k, k \geq 0\}$ are two sequences of independent and identically distributed random vectors, and $\mathbf{A}_k \leq_{lo} \mathbf{B}_k$ for any $k \geq 0$, then use Lemma 2.5 (3) repeatedly, we obtain,

$$\left(\sum_{i=0}^{m_1} r_1(s_{i1}, \sum_{k=0}^{\infty} A_k^1 h_1(s_{i1} - z_k))\Delta s_{i1}, \ldots, \sum_{i=0}^{m_n} r_n(s_{in}, \sum_{k=0}^{\infty} A_k^n h_n(s_{in} - z_k))\Delta s_{in}\right)$$

$$\leq_{lo} \left(\sum_{i=0}^{m_1} r_1(s_{i1}, \sum_{k=0}^{\infty} B_k^1 h_1(s_{i1} - z_k))\Delta s_{i1}, \ldots, \sum_{i=0}^{m_n} r_n(s_{in}, \sum_{k=0}^{\infty} B_k^n h_n(s_{in} - z_k))\Delta s_{in}\right).$$

Denote $x_j(t) = \sum_{k=0}^{\infty} a_k^j h_j(t - z_k)$ and $y_j(t) = \sum_{k=0}^{\infty} b_k^j h_j(t - z_k)$. Since e^{-x} is nonnegative and decreasing in x, we have, from Lemma 2.5 (1),

$$\int \prod_{j=1}^{n} e^{-\sum_{i=0}^{m_j} r_j(s_{ij}, x_j(s_{ij}))\Delta s_{ij}} P_{\mathbf{A}}(d\mathbf{a}) \geq \int \prod_{j=1}^{n} e^{-\sum_{i=0}^{m_j} r_j(s_{ij}, y_j(s_{ij}))\Delta s_{ij}} P_{\mathbf{B}}(d\mathbf{b}).$$

Now let $\max\{\Delta s_{ij}\} \to 0$, we have

$$\lim_{\max\{\Delta s_{ij}\}\to 0} \prod_{j=1}^{n} e^{-\sum_{i=0}^{m_j} r_j(s_{ij}, x_j(s_{ij}))\Delta s_{ij}} = \prod_{j=1}^{n} e^{-\int_0^{t_j} r_j(s, x_j(s))ds}, \tag{4.6}$$

$$\lim_{\max\{\Delta s_{ij}\}\to 0} \prod_{j=1}^{n} e^{-\sum_{i=0}^{m_j} r_j(s_{ij}, y_j(s_{ij}))\Delta s_{ij}} = \prod_{j=1}^{n} e^{-\int_0^{t_j} r_j(s, y_j(s))ds}. \tag{4.7}$$

Note that all the functions involved in (4.6) and (4.7) are bounded, then the bounded convergence theorem implies that for each fixed $\mathbf{z}$, as $\max\{\Delta s_{ij}\} \to 0$,

$$\int \prod_{j=1}^{n} e^{-\sum_{i=0}^{m_j} r_j(s_{ij}, x_j(s_{ij}))\Delta s_{ij}} P_{\mathbf{A}}(d\mathbf{a}) \to f(\mathbf{z}),$$

$$\int \prod_{j=1}^{n} e^{-\sum_{i=0}^{m_j} r_j(s_{ij}, y_j(s_{ij}))\Delta s_{ij}} P_{\mathbf{B}}(d\mathbf{b}) \to g(\mathbf{z}).$$

Therefore, $f(\mathbf{z}) \geq g(\mathbf{z})$ for any $\mathbf{z}$. Hence,

$$\int f(\mathbf{z}) P_\tau(d\mathbf{z}) \geq \int g(\mathbf{z}) P_\tau(d\mathbf{z}), \tag{4.8}$$

which says, from (4.3), that $\mathbf{T} \geq_{uo} \mathbf{T}'$.

(2) Under the current conditions, we have $\mathbf{T} \geq_{uo} \mathbf{T}'$. Also since τ and τ' have the same distribution, and $A_k^j =_{st} B_k^j$ for $j = 1, \ldots, n$, then T_j and T_j' have the same (marginal) distribution. Thus $\mathbf{T}$ is more upper orthant dependent than $\mathbf{T}'$. ■

Corollary 4.2 Let $\mathbf{T} = (T_1, \ldots, T_n)$ be the lifelength vector of components of a system operating in a multivariate shot-noise environment as defined in (3.2) and (3.3). Assume that the functions $r_i(t, x)$, $i = 1, \ldots, n$, are all increasing in x for each t, and $\mathbf{A}_k = (A_k^1, \ldots, A_k^n)$ is positively lower orthant dependent.

1. If $r_i(t, x) = \lambda_i(t)x$ where $\lambda_i(t)$ is decreasing in t for $i = 1, \ldots, n$, and the process τ is associated in time, then $\mathbf{T}$ is positively upper orthant dependent.

2. If $h_i(t) = 1$ for $t \geq 0$ and zero otherwise, $i = 1, \ldots, n$, and the process τ is associated in time, then $\mathbf{T}$ is positively upper orthant dependent.

3. If the process τ is associated with respect to thinning order, then $\mathbf{T}$ is positively upper orthant dependent.

Proof: Let $\{\bar{\mathbf{A}}_k, k \geq 0\}$ be a sequence of independent and identically distributed random vectors such that for each k, $\bar{\mathbf{A}}_k = (\bar{A}_k^1, \ldots, \bar{A}_k^n)$ is a vector of independent random variables and $\bar{A}_k^j =_{st} A_k^j$ for each $j = 1, \ldots, n$. Since $\mathbf{A}_k = (A_k^1, \ldots, A_k^n)$ is PLOD, then $\mathbf{A}_k$ is more lower orthant dependent than $\bar{\mathbf{A}}_k$ for each k. From Theorem 4.1 (2), we then have, $\mathbf{T}$ is more upper orthant dependent than $\bar{\mathbf{T}} = (\bar{T}_1, \ldots, \bar{T}_n)$

where $\bar{T}_i$ is the lifelength of component i of the system operating in an environment with the following multivariate shot-noise process, $i = 1, \ldots, n$,

$$\bar{X}_i(t) = \sum_{k=0}^{\infty} \bar{A}_k^i h_i(t - \tau_k).$$

If $\bar{\mathbf{T}}$ is associated then $\bar{\mathbf{T}}$ is PUOD (see (2.3)). Hence $\mathbf{T}$ is PUOD. We now argue that $\bar{\mathbf{T}}$ is associated under three different conditions.

1. If $r_i(t, x) = \lambda_i(t)x$ where $\lambda_i(t)$ is decreasing in t for $i = 1, \ldots, n$, and the process τ is associated in time, then from Theorem 3.3, $\bar{\mathbf{T}}$ is associated.

2. If $h_i(t) = 1$ for $t \geq 0$ and zero otherwise, $i = 1, \ldots, n$, and the process τ is associated in time, then from Theorem 3.6, $\bar{\mathbf{T}}$ is associated.

3. If the process τ is associated with respect to thinning order, then from Corollary 3.9, $\bar{\mathbf{T}}$ is associated. ■

To compare the dependence structures of two systems with different shock arrival processes, we need to employ some comparison methods of point processes. A point process $\tau = \{\tau_k, k \geq 0\}$ is said to be stochastically larger than a point process $\tau' = \{\tau_k', k \geq 0\}$ (denoted as $\tau \geq_{st} \tau'$) if $Ef(\{\tau_k, k \geq 0\}) \leq Ef(\{\tau_k', k \geq 0\})$ for all increasing functionals f on $\mathcal{R}_+^\infty$. A point process $\tau = \{\tau_k, k \geq 0\}$ is said to be larger than a point process $\tau' = \{\tau_k', k \geq 0\}$ with respect to thinning (denoted as $\tau \geq_{thinning} \tau'$) if $Ef(\tau) \geq Ef(\tau_k')$ for all functionals f defined on $\mathcal{N}$ that are increasing with respect to the ordering (3.12). It is known that $\tau \geq_{thinning} \tau'$ implies that $\tau \geq_{st} \tau'$. For the details on various notions of stochastic comparisons of point processes and their applications, the reader is referred to Shaked and Szekli (1995).

Theorem 4.3 Let $\mathbf{T} = (T_1, \ldots, T_n)$ and $\mathbf{T}' = (T_1', \ldots, T_n')$ be the component lifelengths of systems $\mathcal{S}$ and $\mathcal{S}'$, as defined in (3.3), with multivariate shot-noise processes $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ as described in (4.1) and (4.2) respectively. Assume that $r_i(t; x)$ is increasing in x, $i = 1, \ldots, n$, and $(A_k^1, \ldots, A_k^n) \leq_{lo} (B_k^1, \ldots, B_k^n)$ for each k.

1. If $r_i(t, x) = \lambda_i(t)x$ where $\lambda_i(t)$ is decreasing in t for $i = 1, \ldots, n$, and $\tau \leq_{st} \tau'$, then $\mathbf{T} \geq_{uo} \mathbf{T}'$.

2. If $h_i(t) = 1$ for $t \geq 0$ and zero otherwise, $i = 1, \dots, n$, and $\tau \leq_{st} \tau'$, then $\mathbf{T} \geq_{uo} \mathbf{T}'$.

3. If $\tau \leq_{thinning} \tau'$, then $\mathbf{T} \geq_{uo} \mathbf{T}'$.

Proof: Let $\bar{\mathbf{T}} = (\bar{T}_1, \dots, \bar{T}_n)$ be the component lifelengths of a system as defined in (3.3), with the following multivariate shot-noise process

$$Z_i(t) = \sum_{k=0}^{\infty} B_k^i h_i(t - \tau_k), \quad i = 1, \dots, n.$$

As in Theorem 4.1, denote, for fixed $t_1, \dots, t_n$,

$$f(\mathbf{z}) = P(T_1 > t_1, \dots, T_n > t_n \mid \tau = \mathbf{z}),$$

$$g(\mathbf{z}) = P(\bar{T}_1 > t_1, \dots, \bar{T}_n > t_n \mid \tau = \mathbf{z}),$$

$$\bar{g}(\mathbf{z}) = P(T'_1 > t_1, \dots, T'_n > t_n \mid \tau' = \mathbf{z}).$$

From the proof of Theorem 4.1 (see (4.8)), we have,

$$\int f(\mathbf{z}) P_\tau(d\mathbf{z}) \geq \int g(\mathbf{z}) P_\tau(d\mathbf{z}). \tag{4.9}$$

1. If $r_i(t, x) = \lambda_i(t)x$ where $\lambda_i(t)$ is decreasing in t for $i = 1, \dots, n$, then from Lemma 3.2, $x_i(t; \mathbf{a}, \mathbf{z})$ defined as in (3.6) is decreasing in $\mathbf{z}$ with respect to the usual product order. Thus, $g(\mathbf{z})$ is increasing in $\mathbf{z}$ with respect to the usual product order. Since $\tau \leq_{st} \tau'$ we have

$$\int g(\mathbf{z}) P_\tau(d\mathbf{z}) \geq \int g(\mathbf{z}) P_{\tau'}(d\mathbf{z}) = \int \bar{g}(\mathbf{z}) P_{\tau'}(d\mathbf{z}).$$

   This and (4.9) imply that $\mathbf{T} \geq_{uo} \mathbf{T}'$.

2. If $h_i(t) = 1$ for $t \geq 0$ and zero otherwise, $i = 1, \dots, n$, then from Theorem 3.6, $\phi(\mathbf{z}) = \{n(t), t \geq 0\}$, where $n(t)$ is given by (3.10), is decreasing in $\mathbf{z}$ with respect to the usual product order. Since $r_i(t; x)$ is increasing in x, $g(\mathbf{z})$ is increasing in $\mathbf{z}$ with respect to the usual product order. Using a similar argument as in (1), we obtain that $\mathbf{T} \geq_{uo} \mathbf{T}'$.

3. From Lemma 3.7, $x_i(t; \mathbf{a}, \mathbf{z})$ defined as in (3.13) is increasing in $\mathbf{z}$ with respect to the thinning order. Since $r_i(t; x)$ is increasing in x, $g(\mathbf{z})$ is decreasing in $\mathbf{z}$ with respect to the thinning order. Since $\tau \leq_{thinning} \tau'$ we have
$$\int g(\mathbf{z}) P_\tau(d\mathbf{z}) \geq \int g(\mathbf{z}) P_{\tau'}(d\mathbf{z}) = \int \bar{g}(\mathbf{z}) P_{\tau'}(d\mathbf{z}).$$
This and (4.9) imply that $\mathbf{T} \geq_{uo} \mathbf{T}'$. ■

5 Bounds

Using the results developed in Section 4, we can establish computable upper and lower bounds for the joint survival functions of the systems operating in certain dynamic environments. To simplify the exposition, we only consider the systems with two components to illustrate our results.

5.1 The Poisson Shock Arrival Process

We begin with the bounds on the joint survival functions of the systems (3.4) with $\lambda_i(t) = \lambda_i$ being a constant and the following bivariate shot-noise process,
$$X_i(t) = \sum_{k=0}^{\infty} A_k^i h_i(t - \tau_k), \quad i = 1, 2, \tag{5.1}$$
where $\tau = \{\tau_k, k \geq 0\}$ is a Poisson process with a known rate $m(t), t \geq 0$.

Theorem 5.1 Let (T_1, T_2) be the lifelength vector of a two component system operating in the bivariate shot-noise environment as defined in (5.1). If (A_k^1, A_k^2) is PLOD, then
$$P(T_1 > t_1, T_2 > t_2) \geq$$
$$\mathcal{L}_1^*[\lambda_1 H_1(t_1)] \mathcal{L}_2^*[\lambda_2 H_2(t_2)] \exp[-\sum_{i=1}^{2} M(t_i)] \cdot \exp\left[\sum_{i=1}^{2} \int_0^{t_i} \mathcal{L}_i^*(\lambda_i H_i(u)) m(t_i - u) du\right],$$
where $\mathcal{L}_i^*$ is the Laplace transform of the distribution of A_k^i, $H_i(t) = \int_0^t h_i(u) du$ and $M(t) = \int_0^t m(u) du$.

Proof: Since a Poisson process has independent increments, it is clearly associated in time. From Corollary 4.2, (T_1, T_2) is positively upper orthant dependent; that is, $P(T_1 > t_1, T_2 > t_2) \geq P(T_1 > t_1) P(T_2 > t_2)$. The inequality now follows from Theorem 3.1 of Singpurwalla and Youngren (1993). ■

Example 5.2 Let (A_k^1, A_k^2) be distributed as a Marshall-Olkin distribution (Marshall and Olkin 1967) with $E(A_k^1) = b_1, E(A_k^2) = b_2$. It is known that a Marshall-Olkin distribution is positively associated, and hence PLOD. Let $h_i(t) = \exp[-a_i t]$ for $a_i > 0, i = 1, 2$. Suppose that τ is a Poisson process with a constant rate m, and $\lambda_i = 1$ for $i = 1, 2$. Then from Theorem 5.1 and Singpurwalla (1995), we obtain that

$$P(T_1 > t_1, T_2 > t_2) \geq$$

$$\exp\left(-\sum_{i=1}^{2} \frac{m a_i b_i t_i}{1 + a_i b_i}\right) \cdot \left\{\prod_{i=1}^{2} \left(\frac{1 + a_i b_i - \exp(-a_i t_i)}{a_i b_i}\right)^{m b_i/(1+a_i b_i)}\right\}.$$

■

Consider a bivariate vector (A_k^1, A_k^2) such that $A_k^1 =_{st} A_k^2$. The distribution of (A_k^1, A_k^1) is known as the upper Fréchet bound (see, for example, Whitt 1976) and

$$(A_k^1, A_k^2) \geq_{lo} (A_k^1, A_k^1).$$

If (A_k^1, A_k^2) is PLOD, then Theorems 4.1 and 5.1, Theorem 3.3 of Singpurwalla and Youngren (1993) yield the following bounds.

Corollary 5.3 Let (T_1, T_2) be the lifelength vector of a two component system operating in the bivariate shot-noise environment as defined in (5.1). If $h_1(t) = h_2(t) = h(t)$ for all t, and (A_k^1, A_k^2) is PLOD with $A_k^1 =_{st} A_k^2$, then for all $0 \leq t_1 \leq t_2$,

$$\mathcal{L}^*[\sum_{i=1}^{2} \lambda_i H(t_i)] \cdot \exp\left[\int_0^{t_1} \mathcal{L}^*[\lambda_1 H(t_1 - u_1) + \lambda_2 H(t_2 - u_1)] m(u_1) du_1\right] \cdot$$

$$\exp\left[\int_{t_1}^{t_2} \mathcal{L}^*[\lambda_2 H(t_2 - u_2)] m(u_2) du_2 - M(t_2)\right] \geq$$

$$P(T_1 > t_1, T_2 > t_2) \geq$$

$$\mathcal{L}^*[\lambda_1 H(t_1)] \mathcal{L}^*[\lambda_2 H(t_2)] \exp[-\sum_{i=1}^{2} M(t_i)] \cdot \exp\left[\sum_{i=1}^{2} \int_0^{t_i} \mathcal{L}^*(\lambda_i H(u)) m(t_i - u) du\right],$$

where $\mathcal{L}^*$ is the Laplace transform of the distribution of A_k^i, $H(t) = \int_0^t h(u)du$ and $M(t) = \int_0^t m(u)du$.

5.2 The Renewal Shock Arrival Process

Using the bounds developed in Section 5.1, we can obtain some bounds for the joint survival functions of the systems (3.4) with $\lambda_i(t) = \lambda_i$ being a constant and the following bivariate shot-noise process,

$$X_i(t) = \sum_{k=0}^{\infty} A_k^i h_i(t - \tau_k), \quad i = 1, 2, \tag{5.2}$$

where $\tau = \{\tau_k, k \geq 0\}$ is a renewal process with NBU (NWU) interarrival times. A nonnegative random variable X is said to be new better (worse) than used (NBU, NWU) if

$$P(X > x + t) \leq (\geq)\ P(X > x)P(X > t),$$

for all x and t (see Barlow and Proschan 1981). It is known (see for example, Stoyan 1983) that if X is NBU (NWU) then

$$X \leq_{st} (\geq_{st}) \exp(\theta), \tag{5.3}$$

for some $\theta \leq (\geq)\ 1/E(X)$, where $\exp(\theta)$ is a random variable with the exponential distribution of mean $1/\theta$.

If the interarrival time of a renewal process τ is NBU (NWU), then $\tau \geq_{st} (\leq_{st}$ $)\ \tau'$ where τ' is a Poisson process with a rate $m \leq (\geq)\ 1/(E(\tau_{k+1} - \tau_k))$. Hence, employing Theorem 4.3, we can use the systems with (5.1) to bound the systems with (5.2). We illustrate this by discussing NWU case.

Theorem 5.4 Let (T_1, T_2) be the lifelength vector of a two component system operating in the bivariate shot-noise environment as defined in (5.2). If $\tau_{k+1} - \tau_k$ is NWU, then $(T_1, T_2) \geq_{uo} (T_1', T_2')$ where (T_1', T_2') is the lifelength vector of a two component system operating in the bivariate shot-noise environment as defined in (5.1) with a Poisson shock arrival process of rate $m \geq 1/(E(\tau_{k+1} - \tau_k))$.

Note that if $m = 1/(E(\tau_{k+1} - \tau_k))$ then $\tau_{k+1} - \tau_k$ has the exponential distribution and $\tau =_{st} \tau'$. Thus, (T_1, T_2) and (T_1', T_2') have the same distribution.

References

[1] Barlow, R. E. and F. Proschan (1981). *Statistical Theory of Reliability and Life Testing*, To Begin With, Silver Spring, MD.

[2] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] Burton, R. M. and M. M. Franzosa (1990). Positive dependence properties of point processes. *Annals of Probability*, **18**, 359-377.

[4] Çinlar, E. and S. Özekici (1987). Reliability of complex devices in random environments. *Probability in the Engineering and Informational Sciences* **1**, 97–115.

[5] Çinlar, E., M. Shaked and J. G. Shanthikumar (1989). On lifetimes influenced by a common environment. *Stochastic processes and Their Applications* **33**, 347–359.

[6] Cox, D. R. and V. Isham (1980). *Point Processes.* Chapman and Hall, London.

[7] Daley, D. J. (1971). The definition of a multi-dimensional generalization of shot noise. *J. Appl. Prob.* **8**, 128-135.

[8] Esary, J. D. and F. Proschan (1970). A reliability bound for systems of maintained, interdependent components. *J. Amer. Statist. Assoc.* **65**, 329-338.

[9] Esary, J. D., F. Proschan and D. W. Walkup (1967). Association of random variables, with applications. *Annals of Mathematical Statistics* **38**, 1466–1474.

[10] Fortuin, C. M., P. W. Kasteleyn and J. Ginibre (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* **22**, 89-103.

[11] Jogdeo, K. (1978). On a probability bound of Marshall and Olkin. *Ann. Statist.* **6**, 232-234.

[12] Kamae, T., U. Krengel and G. L. O'Brien (1977), Stochastic inequalities on partially ordered spaces, *Annals of Probability*, **5**, 899-912.

[13] Karlin, S. and Y. Rinott (1980). Classes of orderings os measures and related correlation inequalities, 1. Multivariate totally positive distributions. *Journal of Multivariate Analysis* **10**, 467-498.

[14] Kijima, M., H. Li and M. Shaked (2000). Stochastic processes in reliability. *Handbook of Statistics*, **19**, D. N. Shanbhag and C. R. Rao, eds. 471-510.

[15] Lefèvre, C. and X. Milhaud (1990). On the association of the lifelengths of components subjected to a stochastic environment. *Advances in Applied Probability* **22**, 961–964.

[16] Lemoine, A. J. and M. L. Wenocur (1986). A note on shot-noise and reliability modeling. *Operations Research* **34**, 320-323.

[17] Li, H. and S. H. Xu (2001). Stochastic bounds and dependence properties of survival times in a multicomponent shock model, *Journal of Multivariate Analysis*, **76**, 63-89.

[18] Li, H. and S. H. Xu (2000). On the dependence structure and bounds of correlated parallel queues and their applications to synchronized stochastic systems, *Journal of Applied Probability*, **37**, 1020-1043.

[19] Lindqvist, H. (1987). Monotone and associated Markov chains, with applications to reliability theory. *Journal of Applied Probability* **24**, 679-695.

[20] Lindqvist, H. (1988). Association of probability measures on partially ordered spaces. *Journal of Multivariate Analysis* **26**, 111-132.

[21] Marshall, A. W. and I. Olkin (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc.* **2**, 84-98.

[22] Shaked, M. and J. G. Shanthikumar (1994). *Stochastic Orders and Their Applications*, Academic Press.

[23] Shaked, M. and R. Szekli (1995). Comparison of replacement policies via point processes. *Adv. Appl. Prob.* **27**, 1079-1103.

[24] Singpurwalla, N. (1995). Survival in dynamic environments. *Statistical Sciences* **10**, 86-103.

[25] Singpurwalla, N. and M. Youngren (1991). Models for dependent life lengths induced by common environments. *Topics in Statistical Dependence* (H. W. Block, A. R. Sampson, and T. H. Savits, Eds.), 435-441. IMS, Hayward, CA

[26] Singpurwalla, N. and M. Youngren (1993). Multivariate distributions induced by dynamic environments. *Scandinavian Journal of Statistics* **20**, 251-261.

[27] Stone, C. (1963), Weak convergence of stochastic processes defined on semi-infinite time intervals, *Proc. Amer. Math. Soc.*, **14**, 694-697.

[28] Stoyan, D. (1983). *Comparison Methods for Queues and Other Stochastic Models*, Wiley & Sons.

[29] Szekli, R. (1995), *Stochastic Ordering and Dependence in Applied Probability*, Springer-Verlag.

[30] Tong, Y.L. (1980), *Probability Inequalities in Multivariate Distributions*, Academic Press, New York.

[31] Whitt, W. (1976). Bivariate distribution with given marginals. *Ann. Statist.* **4**, 1280-1289.

[32] Xu, S and H. Li (2000). Majorization of weighted trees: A new tool to study correlated systems. *Mathematics of Operations Research*, **35**, 298-323.