BULLETIN OF THE BRAZILIAN MATHEMATICAL SOCIETY

Bull Braz Math Soc, New Series 34(1), 59-75

The convexity principle and its applications

B. T. Polyak

— Dedicated to IMPA on the occasion of its 50^{th} *anniversary*

Abstract. Recently [1, 2] the new convexity principle has been validated. It states that a nonlinear image of a small ball in a Hilbert space is convex, provided that the map is $C^{1,1}$ and the center of the ball is a regular point of the map. This result has numerous applications in linear algebra, optimization and control.

Keywords: Convexity, image, nonlinear transformation, linear algebra, pseudospectrum, duality, nonconvex programming, optimal control.

Mathematical subject classification: 47H, 52A05, 15A, 49K15, 90C46.

1 Introduction

Convexity plays a key role in functional analysis, optimization and control theory. For instance, if a mathematical programming problem is convex, then necessary optimality conditions coincide with sufficient ones, duality theorems hold and effective numerical methods can be constructed [3]. However, convex problems are just a small island in the ocean of nonconvex ones.

In the present paper we describe the technique, which is useful for establishing convexity. It is based on the recent result [1, 2], asserting convexity of a nonlinear image of a small ball in a Hilbert space. This result is addressed in Section 2, while all other Sections deal with its applications to various fields of mathematics. We start with linear algebra and prove convexity of the spectrum of a family of perturbed matrices, zero sets of perturbed polynomials and value sets of some determinants (Section 3). The duality theory for a class of nonconvex mathematical programming problems (local programming)is developed in Section 4. Special numerical methods for solving such problems are also provided. Various applications to control problems are described in Section 5.

Received 11 September 2002.

They are based on the convexity of the reachable set for nonlinear system with "small energy control".

2 The convexity principle

Let X, Y be two Hilbert spaces, let $f : X \to Y$ be a nonlinear map with Lipschitz derivative on a ball $B(a, r) = \{x \in X : ||x - a|| \leq r\}$, thus

$$||f'(x) - f'(z)|| \leq L||x - z|| \quad \forall x, z \in B(a, r). \tag{1}$$

Suppose that a is a regular point of f, i.e. the linear operator $f'(a)$ maps X onto Y; then there exists $\nu > 0$ such that

$$||f'(a)^* y|| \geq \nu||y|| \quad \forall y \in Y. \tag{2}$$

For instance, for X, Y finite dimensional: $X = \mathbf{R}^n, Y = \mathbf{R}^m$, this condition holds if $\operatorname{rank} f'(a) = m$; for this case $\nu = \sigma_1(f'(a))$ — the least singular value of $f'(a)$.

Theorem 1. *If (1), (2) hold and $\varepsilon < \min\{r, \nu/(2L)\}$, then the image of a ball $B(a, \varepsilon) = \{x \in X : ||x - a|| \leq \varepsilon\}$ under the map f is convex, i.e. $F = \{f(x) : x \in B(a, \varepsilon)\}$ is a convex set in Y. Moreover, the set is strictly convex and its boundary is generated by boundary points of the ball: $\partial F \subset f(\partial B(a, \varepsilon))$.*

We need the following results:

Lemma 1. *A ball in a Hilbert space is strongly convex: if $x_1, x_2 \in B(a, \varepsilon)$, $x_0 = (x_1 + x_2)/2$, then $B(x_0, \rho) \subset B(a, \varepsilon)$ for $\rho = ||x_1 - x_2||^2/(8\varepsilon)$.*

This result is well known and follows immediately from the parallelogram equality.

Lemma 2. *Suppose there exist $L, \rho, \mu > 0$, such that*

$$||f'(x) - f'(z)|| \leq L||x - z|| \quad \forall x, z \in B(x_0, \rho)$$
$$||f'(x)^* y|| \geq \mu||y|| \quad \forall y \in Y, \forall x \in B(x_0, \rho)$$
$$||f(x_0) - y_0|| \leq \rho\mu,$$

then the equation $f(x) = y_0$ has a solution $x^* \in B(x_0, \rho)$ and

$$||x^* - x_0|| \leq \frac{||f(x_0) - y_0||}{\mu}.$$

This Lemma coincides with Corollary 1, Theorem 1 of [4].

Proof of Theorem 1. Let x_1, x_2 be arbitrary points in $B(a,\varepsilon) \subset B(a,r)$, $y_i = f(x_i) \in F, \quad i = 1,2$. Denote $x_0 = (x_1 + x_2)/2$, $y_0 = (y_1 + y_2)/2$. To prove convexity of F it suffices to find $x^* \in B(a,\varepsilon)$ such that $f(x^*) = y_0$. We have $y_1 = f(x_0) + f'(x_0)(x_1 - x_0) + \epsilon_1$, $y_2 = f(x_0) + f'(x_0)(x_2 - x_0) + \epsilon_2$, where $||\epsilon_i|| \le L||x_i - x_0||^2/2 = L||x_1 - x_2||^2/8, i = 1,2$ due to (1), see e.g. [5, Theorem 3.2.12]. Hence $y_0 = f(x_0) + \epsilon_0$, $\epsilon_0 = (\epsilon_1 + \epsilon_2)/2$, $||\epsilon_0|| \le L||x_1 - x_2||^2/8$. All conditions of Lemma 2 are satisfied for $\mu = \nu - L\varepsilon > 0$, $\rho = ||x_1 - x_2||^2/(8\varepsilon)$, because (1),(2) hold, $B(x_0,\rho) \subset B(a,\varepsilon)$ due to Lemma 1, $||f(x_0) - y_0|| = ||\epsilon_0|| \le L||x_1 - x_2||^2/8 = L\rho\varepsilon \le \rho(\nu - L\varepsilon) = \rho\mu$. Moreover, $||f'(x)^* y|| \ge ||f'(a)^* y|| - ||(f'(x)^* - f'(a)^*)y|| \ge \nu||y|| - L||x - a||||y|| \ge (\nu - L\varepsilon)||y|| = \mu||y||$ for $x \in B(x_0,\rho)$. Thus Lemma 2 provides the desired x^* and the proof of convexity of F is completed.

From the above reasoning it follows that for $x_1 \ne x_2$ the equation $f(x) = y$ has a solution for y close enough to y_0; this validates strict convexity of F. Finally, if x_0 is an interior point of $B(a,\varepsilon)$ then there exists $B(x_0,\rho) \subset B(a,\varepsilon)$, $\rho > 0$ such that Lemma 2 holds. Hence for any y close enough to $f(x_0)$ the equation $f(x) = y$ has a solution in $B(a,\varepsilon)$. This means that the image of interior points of the ball $B(a,\varepsilon)$ lies in the interior of F, that is ∂F is generated by the boundary of $B(a,\varepsilon)$. □

Remarks. 1. We presented the proof, based on Lemma 2 (which has been derived in [4] by use of a version of Newton method). Another proofs can be obtained by modern techniques, related to Ljusternik theorem (see e.g. [7, Theorem 2.7], [8]). However, the proofs of the Ljusternik-like results are also based on the Newton method. On the other hand an attempt to use the fundamental theorem by Graves on solvability of nonlinear equations [6] instead of Lemma 2 fails, because the theorem does not provide explicit bounds for solutions which are required in the proof.

2. The idea of Theorem 1 is very simple. The ball $B(a,\varepsilon)$ is strongly convex, thus its image under linear map $f'(a)$ is strongly convex as well. But it can not loose convexity for a nonlinear map f, which is close enough to its linearization. The same reasoning explains that the result can not be extended to an arbitrary Banach spaces, where a ball is not strongly convex. However the extension of the principle to uniformly convex Banach spaces (such as L_p, $1 < p < \infty$) is an open problem.

3. The result holds, if we replace the ball by any other strongly convex set (e.g. by a nondegenerate ellipsoid). For particular case $f : \mathbf{R}^n \to \mathbf{R}^n$ Theorem 1 has

been extended in [9] for strictly convex (not necessarily strongly convex) sets.

4. Smoothness assumptions of Theorem 1 can not be seriously relaxed. For instance, A.Ioffe constructed a counterexample with f continuously differentiable but not in $C^{1,1}$. Then the result is false.

In many cases the conditions of Theorem 1 can be effectively checked, and the radius ε of the ball can be estimated. One of such examples is a quadratic transformation.

Example. Let $x \in \mathbf{R}^n$ and $f(x) = (f_1(x), ...f_m(x))^T$ where $f_i(x)$ are quadratic functions:

$$f_i(x) = (1/2)(A_i x, x) + (a_i, x), \quad A_i = A_i^T \in \mathbf{R}^{n\times n}, \\ a_i \in \mathbf{R}^n, \quad i = 1, ...m. \tag{3}$$

Take $a = 0$, that is $B = \{x : ||x|| \le \varepsilon\}$. Then $f_i'(x) = A_i x + a_i$ and (1) is satisfied on $\mathbf{R}^n$ with $L = (\sum_{i=1}^m ||A_i||^2)^{1/2}$ where $||A_i||$ stands for the spectral norm of matrices A_i. Consider the matrix A with columns a_i: $A = (a_1|a_2|...|a_m)$. Then $f'(0)^T y = Ay$, and if $\mathrm{rank} A = m$, then (2) holds with $\nu = \sigma_1(A)$ — the minimal singular value of A, that is $\nu = (\min \lambda_1(A^T A))^{1/2}$, where λ_1 is the minimal eigenvalue of the corresponding matrix. Hence, Theorem 1 implies:

Proposition 1. *If $\varepsilon < \nu/(2L)$, then the image of the ball B under the map (3) is convex:*

$$F = \{f(x) : ||x|| \le \varepsilon\}$$

is a convex set in $\mathbf{R}^m$.

This is in a sharp contrast with the results on images of arbitrary balls under quadratic transformations, where the convexity can be validated [10] just under very restrictive assumptions.

For instance, let $n = m = 2$ and

$$f_1(x) = x_1 x_2 - x_1, \; f_2(x) = x_1 x_2 + x_2. \tag{4}$$

Then the estimates above guarantee that F is convex for $\varepsilon < \varepsilon^* = 1/(2\sqrt{2}) \approx 0.3536$. It can be proved directly for this case that F is convex for $\varepsilon \le \varepsilon^*$ and looses convexity for $\varepsilon > \varepsilon^*$. Thus the estimate provided by Proposition 1 is tight for this example.

Figure 1 shows the images of the ε-discs $\{x \in \mathbf{R}^2 : ||x|| \le \varepsilon\}$ under the mapping (4) for various values of ε.

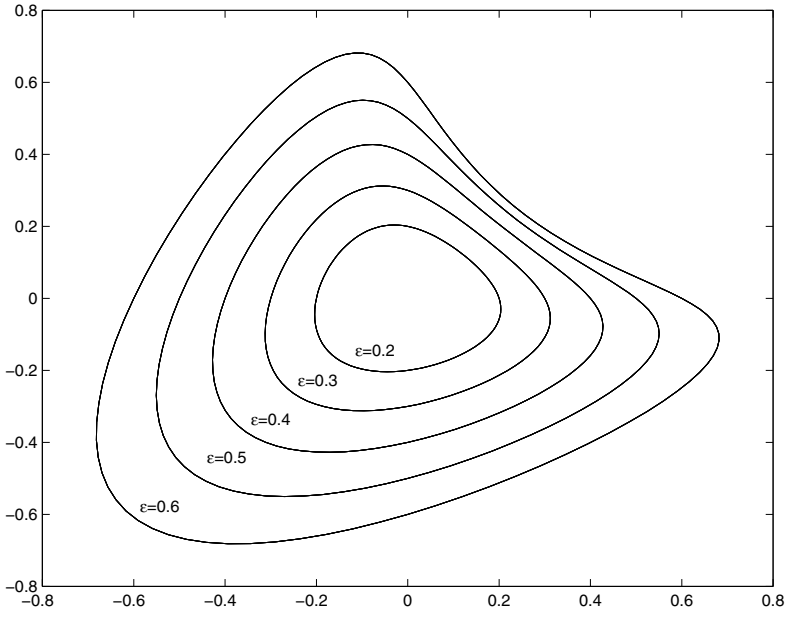

Figure 1: Images of ε-discs for various ε.

3 Applications to linear algebra

In this section we consider three applications of the convexity principle to linear algebra: perturbation of spectrum of real matrices, zero set of perturbed polynomials and the set arising in so-called μ-analysis.

a. Pseudospectrum. The set of all eigenvalues of a family of perturbed matrices is called *pseudospectrum*. More rigorously, the pseudospectrum of a nominal matrix $A \in \mathbf{R}^{n\times n}$ is

$$\Lambda_\varepsilon(A) = \{\lambda \in \mathbf{C} : \exists \Delta \in \mathbf{R}^{n\times n}, ||\Delta||_F \leq \varepsilon, \lambda \text{ is an eigenvalue of } A + \Delta.\} \quad (5)$$

Usually [11, 12, 13] matrices A, Δ are assumed to be complex, while the matrix norm is the spectral one $||\Delta|| = (\max \lambda(\Delta^*\Delta))^{1/2}$. For this case the closed-form characterization of pseudospectrum is available. The real case is much more difficult, and neither effective description of $\Lambda_\varepsilon(A)$ nor its qualitative behavior for small ε are known. The result below provides such information for Frobenius norm of matrix perturbations: $||\Delta||_F = (\sum \delta_{ij}^2)^{1/2}$, where $\delta_{ij}, i, j = 1, \ldots, n$ are entries of Δ.

Theorem 2. *If A has all distinct eigenvalues, then its pseudospectrum is the union of n nonintersecting convex sets on the complex plane provided that ε is small enough.*

Proof. Consider a map $f : \Delta \to \lambda$ where $\Delta \in X$, X is the space of $n \times n$ real matrices equipped with the scalar product $\langle A, B\rangle = \mathrm{Trace} A^T B$ and corresponding norm $||\Delta||_F = \langle \Delta, \Delta\rangle^{1/2}$ and $\lambda \in \mathbf{C}$ is one fixed eigenvalue of $A + \Delta$; the space $\mathbf{C}$ can be identified with $Y = \mathbf{R}^2$. All eigenvalues of A are distinct hence the same holds for $A + \Delta$ with $||\Delta||_F$ small enough. Thus f is well defined. Due to standard results of perturbation theory [14, Chapter 2] the map f is twice differentiable if all eigenvalues of A are distinct and explicit formulae for second derivatives confirm that they are bounded for $||\Delta||_F$ small enough, so $f \in C^{1,1}$.

Now, the derivative of f is given by [14]

$$g = f'(0)\Delta = \frac{y^T \Delta x}{y^T x}$$

where x, y are left and right eigenvectors of A, corresponding to the eigenvalue λ:

$$Ax = \lambda x, \quad A^T y = \lambda y, \quad y^T x \neq 0.$$

If λ is real, then the eigenvalue remains real under small perturbations of A (a simple real eigenvalue can not become complex), so the pseudospectrum is an interval and hence it is convex in this case. If λ is complex: $\lambda = \alpha + i\beta$, $\beta \neq 0$ then we shall prove that g runs the entire complex plane when Δ runs X. First note that if $x = u + iv$, $y = s + it$ then $u \neq 0$, $v \neq 0$ and u, v are linearly independent (and similar is true for s, t). Indeed, $Au = \alpha u - \beta v$, $Av = \alpha v + \beta u$, and $u = 0$ implies (due to the first equality and $\beta \neq 0$) that $v = 0$; this contradicts the assumption $x \neq 0$. Simultaneously $v = 0$ leads to the contradiction. If $u = \gamma v$, $\gamma \neq 0$ then from above equations $(\alpha - \beta\gamma)x = (\alpha + \beta/\gamma)x$, that is $\gamma^2 = -1$; this is impossible. The properties of u, v, s, t ensure the existence of vectors $u^\perp, v^\perp, s^\perp, t^\perp \in \mathbf{R}^n$ such that

$$u^T u^\perp = 0, u^T v^\perp \neq 0, v^T v^\perp = 0, v^T u^\perp \neq 0,$$

$$s^T s^\perp = 0, s^T t^\perp \neq 0, t^T t^\perp = 0, t^T s\perp \neq 0.$$

Now take $\Delta = \mu v^\perp (s^\perp)^T + \nu u^\perp (t^\perp)^T$ with some real μ, ν. Then after simple calculations one gets

$$y^T \Delta x = \mu (u^T v^\perp)(s^T t^\perp) + i\nu (v^T u^\perp)(t^T s\perp) = \mu c_1 + i\nu c_2,$$

where $c_1, c_2 \neq 0$. For arbitrary $\mu, \nu \in \mathbf{R}^1$ this expression runs the entire $\mathbf{C}$. We conclude that $f'(0)\Delta$ maps X onto Y. This means that we are in the framework of Theorem 1 and each eigenvalue of A diffuses into a convex set in $\mathbf{C}$. □

Note that if Frobenius norm is replaced with another matrix norm, the space of matrices looses its Hilbert structure and we can not apply Theorem 1. However it is not clear if there exists an analog of Theorem 2 for some other norms.

The assumption on simplicity of all eigenvalues is significant. For instance if $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ then the pseudospectrum of A reminds a cross centered at (1,0) and it is nonconvex for arbitrary small ε.

b. Zero set for polynomials. Consider a family of polynomials with real coefficients

$$P(x, a) = a_1 + a_2 x + \cdots + a_n x^{n-1} + x^n, \quad a \in \mathbf{R}^n, ||a - a^0|| \leq \varepsilon, \qquad (6)$$

where a^0 are the coefficients of the nominal polynomial and $||.||$ stands for Euclidean norm. The set of all zeros of such polynomials is called a *zero set*:

$$Z_\varepsilon = \{x \in \mathbf{C} : \exists\, a, ||a - a^0|| \leq \varepsilon, P(x, a) = 0.\} \qquad (7)$$

A formula for computation of the zero set for arbitrary ε is known [15], however it does not provide the qualitative description of the set. The result below yields the construction of the zero set for small ε and distinct roots of the nominal polynomial.

Theorem 3. *If $P(x, a^0)$ has all distinct zeros, then the zero set of family (6) is the union of n nonintersecting convex sets on the complex plane provided that ε is small enough.*

Proof. The proof follows the same lines as for Theorem 2, and we focus on the key points only. Introduce $f : a \to \lambda$ where $\lambda \in \mathbf{C}$ is a fixed simple zero of $P(x, a)$; this function is well defined in the neighborhood of a^0, $\lambda_0 = f(a^0)$. This function is in $C^{1,1}$ and the derivative is given by $f'(a^0)\Delta = -q^T\Delta / P'_x(\lambda_0, a^0)$, where $\Delta = a - a^0 \in \mathbf{R}^n$, $q = (1, \lambda_0, \lambda_0^2, \ldots, \lambda_0^{n-1})^T \in \mathbf{R}^n$ and P'_x denotes differentiation with respect to x, while $P'_x(\lambda_0, a^0) \neq 0$ because λ_0 is a simple zero of $P(x, a_0)$. If λ_0 is real, the corresponding component of the zero set is an interval (real simple zero remains real under small perturbations of the coefficients) and hence convex. If λ_0 is complex ($\Im\lambda_0 \neq 0$) then it is not hard

to prove that vectors $\Re q, \Im q$ are linearly independent and thus $f'(a^0)\Delta$ maps $\mathbf{R}^n$ onto $\mathbf{C}$. The use of Theorem 1 terminates the proof. □

For multiple roots the zero set can be the union of nonconvex sets for arbitrary small ε, as simple examples demonstrate. The same situation holds if we replace Euclidean norm with ∞-norm. For instance, the zero set for the second order polynomial family $P(x, a) = a_1 + a_2 x + x^2, |a_1 - 1| \leq \varepsilon, |a_2| \leq \varepsilon$ is the union of two distorted rectangles which are nonconvex. The case of p-norms, $1 < p < \infty$ remains uninvestigated.

c. A problem in μ-analysis. So called μ-analysis (or structured singular value problem) is an important tool in modern control theory [16]. For a given matrix $M \in \mathbf{C}^{n\times n}$ finding $\mu(M)$ requires to estimate $r_{\max}$ — the largest r which preserves $\det(I + M\Delta)$ nonvanishing for all matrix perturbations Δ with norm less then r. It is assumed that Δ has a specified block structure with each block real or complex and specified norm of each block. We address a particular case of the problem — real perturbations with one-block structure and Frobenius norm (sometimes such version is called spherical μ). Define *the value set* of the determinant $\det(I + M\Delta)$ when Δ runs the ε-ball:

$$D_\varepsilon = \{\det(I + M\Delta) \in \mathbf{C} : \Delta \in \mathbf{R}^{n\times n}, ||\Delta||_F \leq \varepsilon\}. \tag{8}$$

We say that the matrix M is essentially complex if $M \neq zW$, $z \in \mathbf{C}$, $W \in \mathbf{R}^{n\times n}$. In other words, for $M = U + iV, U, V \in \mathbf{R}^{n\times n}$, M is essentially complex if U, V are linearly independent or equivalently $\langle U, U\rangle\langle V, V\rangle \neq \langle U, V\rangle^2$.

Theorem 4. *If M is essentially complex and ε is small enough, then the set D_ε is convex.*

Proof. The equality

$$\det(I+M\Delta) = 1+\mathrm{Trace}(M\Delta)+o(||\Delta||) = 1+\langle U^T, \Delta\rangle+i\langle V^T, \Delta\rangle+o(||\Delta||)$$

verifies that the map $\Delta \to \det(I + M\Delta)$ is differentiable; it can be also proved to be $C^{1,1}$. Due to essential complexity of M the image of $\langle U^T, \Delta\rangle+i\langle V^T, \Delta\rangle$, $\Delta \in \mathbf{R}^{n\times n}$ is the entire $\mathbf{C}$. Thus Theorem 1 is applicable. □

Figure 2 presents the set D_ε for

$$M = \begin{pmatrix} 1+i & i \\ 0 & 1+i \end{pmatrix}, \quad \varepsilon = 0.8.$$

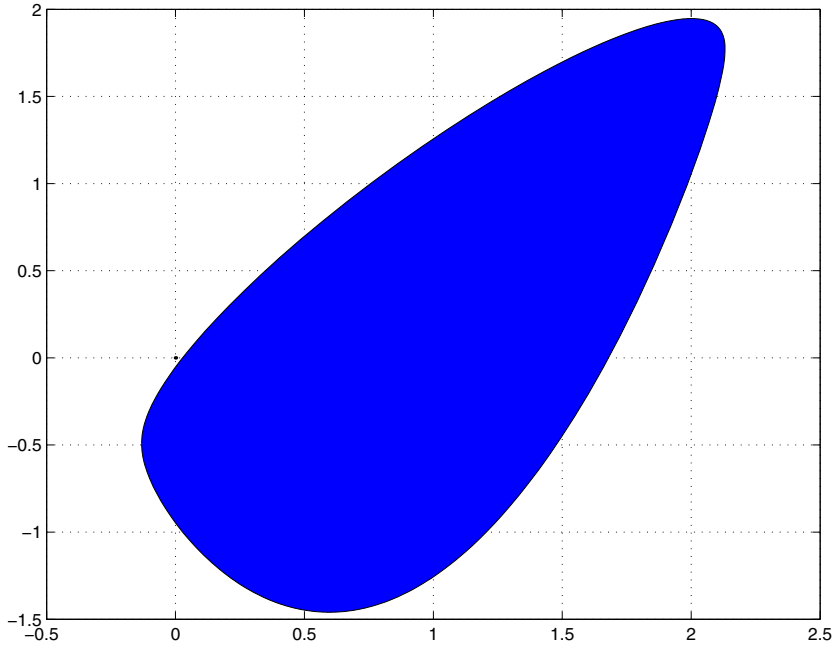

Figure 2: Image of $\det(I + M\Delta)$, $||\Delta||_F \leq 0.8$.

The set has been constructed via the algorithm of Section 4. To find a boundary point $g(\theta)$ of the set having a normal $c = (c_1, c_2)^T$ with $c_1 = \cos\theta, c_2 = \sin\theta, \theta \in [0, 2\pi]$ being one-dimensional parameter, one should minimize the function

$$f(X) = c_1 \Re \det(I + MX) + c_2 \Im \det(I + MX)$$

subject to $X \in \mathbf{R}^{n\times n}$, $||X||_F \leq \varepsilon$. The derivative of f reads

$$f'(X) = c_1(uU - vV) + c_2(vU + uV),$$

$$\det(I + MX) = u + iv, ((I + MX)^{-1}M)^T = U + iV.$$

Thus we can apply method (16) of Section 4 for this minimization problem; its solution $X(\theta)$ provides the desired point $g(\theta) = \det(I + MX(\theta)) \in \mathbf{C}$. One can see that the origin does not belong to the set $D_{0.8}$, but is very close to it. We conclude that $0.8 < r_{\max}$ and $r_{\max} - 0.8$ is small. Thus in this case we can estimate $r_{\max}$ by using the above technique for construction of D_ε. However this is not the universal tool — the set D_ε can loose convexity for $\varepsilon < r_{\max}$. For instance five matrices M from [17, Table 1] have been checked; for three of them the set D_ε becomes nonconvex with some $\varepsilon < r_{\max}$.

4 Local programming

Simultaneously with the standard mathematical programming problem

$$\begin{aligned} \min f_0(x), \quad & x \in \mathbf{R}^n \\ f_i(x) \le 0, \quad & i = 1, ..., l \\ f_i(x) = 0, \quad & i = l+1, ..., m \end{aligned} \tag{9}$$

consider its version with the extra constraint

$$\begin{aligned} \min f_0(x), \quad & x \in \mathbf{R}^n \\ f_i(x) \le 0, \quad & i = 1, ..., l \\ f_i(x) = 0, \quad & i = l+1, ..., m \\ & ||x - a|| \le \varepsilon. \end{aligned} \tag{10}$$

which we call *local programming problem*. Suppose that the functions $f_i(x)$, $i = 0, 1, ...m$ are from $C^{1,1}$ on $B(a, \varepsilon)$. Construct the Lagrange function

$$L(x, y) = \sum_{i=0}^{m} y_i f_i(x). \tag{11}$$

Denote $Y_+ = \{y \in \mathbf{R}^{m+1} : y_i \ge 0, i = 0, 1, ..., l\}$. We assume, that a is a feasible point in (9), moreover we can assume without loss of generality that all inequality constraints are active in a:

$$f_i(a) = 0, \quad i = 1, ..., m,$$

otherwise they play no role in (10) and can be rejected for ε small enough. Finally we suppose that the gradients of $f_i(x)$, $i = 0, 1, ..., m$ at a are linearly independent, i.e. there exists no $y^0 \ne 0$ such that $L_x(a, y^0) = 0$. If there are no inequality constraints, this condition means that a is not a stationary point in (9). In the presence of inequality constraints this condition is more restrictive than the assumption "a is not a Kuhn-Tucker point in problem (9)". For instance, it implies $m < n$, that is the number of active constraints in a is less than the dimension.

Theorem 5. *Under above assumptions there exists $\varepsilon^* > 0$ such that a solution x^* of (10) with $0 < \varepsilon < \varepsilon^*$ exists, is unique, lies on the boundary of $B(a, \varepsilon)$: $||x^* - a|| = \varepsilon$ and the following inequality holds*

$$L(x, y^*) \ge L(x^*, y^*) \quad \forall x : ||x - a|| \le \varepsilon \tag{12}$$

for some $y^* \in Y_+$, $y^* \ne 0$, $y_i^* f_i(x^*) = 0, i = 1, ..., l$.

Proof. The problem (10) is equivalent to the optimization problem in the "image space":

$$\min f_0, \quad f \in F, \quad f_i \leq 0, i = 1, ..., l, \tag{13}$$

where $f = (f_0, f_1, ..., f_m) \in \mathbf{R}^{m+1}$, $f(x) = (f_0(x), f_1(x), ..., f_m(x))$, $F = \{f(x) : ||x - a|| \leq \varepsilon\}$. The point a is a regular point for $f(x)$ because $f_i'(a)$ are linearly independent. Theorem 1 guarantees the convexity of F for ε small enough. Thus (13) is a convex problem and for its solution $f^* = f(x^*)$ there exists a separating hyperplane: $0 \neq y^* \in \mathbf{R}^{m+1}$, $(y^*, f) \geq 0 \quad \forall f : f \in F$, $f_0 \geq f_0^*$, $f_i \leq 0, i = 1, ..., l$. This condition is equivalent to (12). Strict convexity of F implies that the solution of (13) is unique and lies on the boundary of F which (as Theorem 1 claims) is the image of the points with $||x - a|| = \varepsilon$, thus $||x^* - a|| = \varepsilon$. □

Note that for

$$\psi(y) = \min_{||x-a|| \leq \varepsilon} L(x, y);$$

the result can be formulated as follows: if x^* is a solution of (10) then there exists $y^* \in Y_+$ such that

$$L(x^*, y^*) = \max_{y \in Y_+} \psi(y).$$

This is the dual formulation of the problem.

Under some Slater-like condition we can ensure $y_0^* \neq 0$, that is y_0^* can be taken equal to one.

Theorem 6. *Suppose that the following regularity condition holds: for any* $\varepsilon > 0, \sigma \in \mathbf{R}^m : \sigma_i = 1, i = 1, ..., l, |\sigma_i| = 1, i = l + 1, ..., m$ *there exists* x_σ *such that*

$$\sigma_i f_i(x_\sigma) < 0, \quad i = 1, ..., m, \quad ||x_\sigma - a|| \leq \varepsilon. \tag{14}$$

Then in Theorem 5 we can take $y_0^* = 1$ *and (12) is necessary and sufficient condition for optimality in (10).*

Proof. From (12) we get

$$y_0^*(f_0(x) - f_0(x^*)) + \sum_{i=1}^{m} y_i^* f_i(x) \geq 0 \quad \forall ||x - a|| \leq \varepsilon.$$

Take $\sigma : \sigma_i = \text{sign} y_i^*$ and the corresponding x_σ. Then for $y_0^* = 0$ we have $\sum_{i=1}^m y_i^* f_i(x_\sigma) < 0$ (because $y^* \neq 0$), which contradicts the inequality above for $x = x_\sigma$. Thus $y_0^* > 0$, of course we can scale y^* to make $y_0^* = 1$. Condition (12) is obviously sufficient for optimality if $y_0^* = 1$. □

Regularity condition (14) can be replaced by other ones, e.g.: $f_i'(a), i = l+1, ..., m$ are linearly independent and there exists $h \in \mathbf{R}^n : (f_i'(a), h) = 0, i = l+1, ..., m, (f_i'(a), h) < 0, i = 1, ..., l$.

Let us show how these results work for the case of quadratic functions. Consider (10) with $a = 0$ and

$$f_i(x) = (1/2)(A_i x, x) + (a_i, x) + \alpha_i, \quad i = 0, 1, ..., m.$$

Suppose that $\alpha_i \leq 0, i = 1, ..., l, \alpha_i = 0, i = l+1, ..., m$ and the assumptions of Proposition 1 are satisfied (with obvious changes of notation). Then Theorem 5 can be applied,

$$L(x, y) = (1/2)(A(y)x, x) + (a(y), x) + \alpha(y),$$

$$A(y) = \sum_{i=0}^m y_i A_i, \quad a(y) = \sum_{i=0}^m y_i a_i, \quad \alpha(y) = \sum_{i=0}^m y_i \alpha_i.$$

Then $\psi(y)$ can be found as the solution of the problem

$$\psi(y) = \min_{||x|| \leq \varepsilon} ((A(y)x, x) + 2(a(y), x) + \alpha(y)).$$

This problem is always tractable (even if $A(y)$ is not positive definite), and can be effectively solved [10]. Thus we can calculate $\psi(y)$, it is not hard to calculate $\partial_y \psi(y)$ as well. Hence we can apply the subgradient method for maximization of $\psi(y)$ on Y_+.

In more general case, when $f_i(x)$ are nonquadratic functions, minimization of $L(x, y)$ on a ball can be performed by use of the special iterative method. Consider the simplest optimization problem:

$$\min_{||x-a|| \leq \varepsilon} f(x) \tag{15}$$

and the iterative method

$$x^{k+1} = a - \varepsilon \frac{f'(x^k)}{||f'(x^k)||}. \tag{16}$$

Theorem 7. *Suppose that $f : \mathbf{R}^n \to \mathbf{R}^1$ is $C^{1,1}$ on $B(a, \varepsilon)$:*

$$||f'(x) - f'(y)|| \leq L||x - y||, \quad x, y \in B(a, \varepsilon)$$

and $f'(a) \neq 0$ while $\varepsilon < ||f'(a)||/(2L)$. Then

a) The solution x^* of (15) exists and is unique, $||x^* - a|| = \varepsilon$ and the necessary and sufficient optimality condition holds:

$$x^* = a - \varepsilon \frac{f'(x^*)}{||f'(x^*)||}. \tag{17}$$

b) Method (16) converges with linear rate of convergence for any $x^0 \in B(a, \varepsilon)$:

$$||x^k - x^*|| \leq q^k ||x^0 - x^*||, \quad q = O(\varepsilon) = \frac{\varepsilon L}{||f'(a)|| - \varepsilon L} < 1. \tag{18}$$

Proof. The statement a) follows from Theorem 3; (17) is the necessary condition of x^* to be the minimum point in (15).

If we subtract (17) from (16) we get

$$x^{k+1} - x^* = \varepsilon \left(\frac{f'(x^k)}{||f'(x^k)||} - \frac{f'(x^*)}{||f'(x^*)||} \right).$$

For any $0 < \tau, x \in \mathbf{R}^n$, $||x|| \geq \tau$ the vector $\tau x/||x||$ is a projection of x on the ball $B(0, \tau)$. Projection is a nonexpanding map, so we can proceed (with $\tau = ||f'(a)|| - \varepsilon L$)

$$||x^{k+1} - x^*|| \leq (\varepsilon/\tau)||f'(x^k) - f'(x^*)|| \leq q||x^k - x^*||.$$

This is equivalent to the desired estimate (18). □

Note that (16) can be considered as the conditional gradient method [18] for solving (15) with the special stepsize rule. However, its structure is rather peculiar: each new step is performed from the point a, not x^k.

5 Control applications

We consider very briefly (with no technical details) some control applications of the "image convexity" principle.

a. Convexity of the reachable set. A general nonlinear control system

$$\dot{x} = F(x, u, t), x \in \mathbf{R}^n, u \in \mathbf{R}^m, 0 \leq t \leq T, x(0) = c \tag{19}$$

with L_2-bounded control

$$u \in U = \{u : \int_0^T ||u(t)||^2 dt \leq \varepsilon\} \tag{20}$$

defines a reachable set

$$R_\varepsilon = \{x(T) : x(t) \text{ is a solution of (19)}, u \in U\}. \tag{21}$$

Suppose that the linearized system

$$\dot{z} = F_x(x_0, 0, t)z + F_u(x_0, 0, t)u, \quad z(0) = 0 \tag{22}$$

is controllable [19]; here x_0 is the solution of the nominal system

$$\dot{x_0} = F(x_0, 0, t), \quad x_0(0) = c.$$

Then (under some technical assumptions to guarantee the smoothness of the map $f : u \to x(T)$) we can conclude, that for ε small enough the reachable set R_ε is convex. Indeed, we can apply Theorem 1 with $X = L_2$, $Y = \mathbf{R}^n$, $f : u \to x(T)$. The controllability of (22) ensures regularity of this map at $u = 0$.

b. Sufficiency of the local maximum principle. Consider the optimal control problem

$$\min \phi(x(T)) \tag{23}$$

where $x(t)$ is a solution of (19) subject to the constraint (20) and terminal time T is fixed and the function $\phi : \mathbf{R}^n \to \mathbf{R}^1$ is convex. Then this optimal control problem is equivalent to finite-dimensional one:

$$\min_{x \in R_\varepsilon} \phi(x)$$

which is convex under above conditions. Thus the first-order necessary conditions for the extremum (which can be written in the form of local maximum principle [19]) is also sufficient. Thus we conclude that the local maximum principle is the sufficient condition for optimality for (19), (20), (23).

Also from Theorem 5 we obtain that the solution is unique and it reduces (20) to equality.

c. Numerical methods. Iterative method (16) can be applied to solve the optimal control problem (19), (20), (23). It has the following form. At k-th iteration we have an approximation $u^k = u^k(t), 0 \le t \le T$ and calculate x^k as a solution of (19) with $u = u^k$. Then the gradient of the objective function can be found as

$$f'(u^k) = -F_u^T(x^k, u^k, t)\psi^k(t),$$

where ψ^k is a solution of the adjoint system

$$\dot{\psi} = -F_x^T(x^k, u^k, t)\psi, \quad \psi(T) = -\phi'(x^k(T)).$$

Then the updated control is found by (16), where L_2 norm is used. Theorem 7 guarantees convergence of this method to the optimal control.

d. Discrete-time case. Let the states $x_k \in \mathbf{R}^n$ and controls $u_k \in \mathbf{R}^m$ be described by nonlinear difference equations

$$x_{k+1} = F(x_k, u_k, k), x_0 = c, k = 0, 1, ..., N.$$

Our objective is

$$\min \phi(x(N))$$

subject to l_2-type constraint

$$\sum_{k=0}^{N-1} ||u_k||^2 \le \varepsilon.$$

Then under condition of controllability of the linearized system we can prove (as it was done above for the continuous-time case) that the reachable set is convex if ε is small enough. The standard technique allows to obtain the first-order optimality condition which is necessary and sufficient.

6 Conclusions

The new "image convexity" principle is a promising tool for analysis of various problems. In this paper we have presented some of its application to linear algebra, optimization and control. Probably, much more applications can arise in other fields, including functional analysis and numerical analysis.

References

[1] B. Polyak, Convexity of nonlinear image of a small ball with applications to optimization, *Set-Valued Analysis*, **9**(1/2) (2001), 159–168.

[2] B. Polyak, Local programming, *Comp. Math. and Math. Phys.,* **41**(9) (2001), 1259–1266.

[3] R. T. Rockafellar, *Convex Analysis,* Princeton University Press, Princeton, (1970).

[4] B. Polyak, Gradient methods for solving equations and inequalities, *USSR Comput. Math. and Math. Phys.*, **4**(6) (1964), 17–32.

[5] J. W. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables,* Academic Press, New York – London, (1970).

[6] L. M. Graves, Some mapping theorems, *Duke Math. J.*, **17** (1950), 111–114.

[7] A. V. Dmitruk, A. A. Miljutin and N. M. Osmolovskii, Ljusternik theorem and extremum theory, *Russian Math. Surveys*, **55**(6) (1980), 11–46.

[8] A. D. Ioffe, On the local surjection property, *Nonlinear Analysis: Theory, Methods and Appl.*, **11**(5) (1987), 565–592.

[9] S. V. Emelyanov, S. K. Korovin and N. A. Bobylev, On convexity of images of convex sets under smooth transformations, *Doklady RAN,* **385**(3) (2002), 302–304.

[10] B. T. Polyak, Convexity of quadratic transformations and its use in control and optimization, *Journ. Optim. Th. and Appl.,* **99**(3) (1998), 553–583.

[11] Pseudospectra gateway: `www.comlab.ox.ac.uk/pseudospectra`.

[12] L. N. Trefethen, Pseudospectra of matrices, in: *Numerical Analysis,* D. F. Griffith and G. A. Watson eds., Harlow, Longman, (1992), 234–266.

[13] L. N. Trefethen, Pseudospectra of linear operators, *SIAM Review,* **30** (1997), 383–406.

[14] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, (1965).

[15] B. R. Barmish and R. Tempo, On the spectral set for a family of polynomials, *IEEE Trans. Autom. Contr.,* **36** (1991), 111–115.

[16] K. Zhou, J. C. Doyle and K. Glover, *Robust and Optimal Control*, Prentice-Hall, Upper Saddle River, NJ, (1996).

[17] L. Qiu, B. Bernhardson, A. Rantzer, E. J. Davison, P. M. Young and J. C. Doyle, A formula for computation of the real stability radius, *Automatica,* **31**(6) (1995), 879–890.

[18] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, (1998).

[19] E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*, John Wiley, New York, (1970).

B. T. Polyak
Institute for Control Science, Moscow
RUSSIA
E-mail: boris@ipu.rssi.ru