Set-Valued Analysis **9:** 159–168, 2001.

Convexity of Nonlinear Image of a Small Ball with Applications to Optimization

B. T. POLYAK
Institute for Control Science, Moscow, Russia. e-mail: boris@ipu.rssi.ru

(Received: 11 May 2000)

Abstract. Let $f\colon X \to Y$ be a nonlinear differentiable map, X, Y are Hilbert spaces, $B(a, r)$ is a ball in X with a center a and radius r. Suppose $f'(x)$ is Lipschitz in $B(a, r)$ with Lipschitz constant L and $f'(a)$ is a surjection: $f'(a)X = Y$; this implies the existence of $\nu > 0$ such that $||f'(a)^*y|| \geqslant \nu||y||$, $\forall y \in Y$. Then, if $\varepsilon < \min\{r, \nu/(2L)\}$, the image $F = f(B(a, \varepsilon))$ of the ball $B(a, \varepsilon)$ is convex. This result has numerous applications in optimization and control. First, duality theory holds for nonconvex mathematical programming problems with extra constraint $||x - a|| \leqslant \varepsilon$. Special effective algorithms for such optimization problems can be constructed as well. Second, the reachability set for 'small power control' is convex. This leads to various results in optimal control.

Mathematics Subject Classifications (2000): 52A05, 58E25, 90C46.

Key words: convexity, image, optimization, nonlinear transformation, duality, nonconvex programming, optimal control.

1. Introduction

Convexity plays a key role in optimization and control theory. If a mathematical programming problem is convex, then the duality theorem holds and effective numerical methods can be constructed [1]. However, convexity is an exception for general nonlinear optimization problems.

In the present paper, we describe a new class of problems, which are originally nonconvex, but yield convex techniques. The basic mathematical result is a new theorem, asserting convexity of a nonlinear image of a small ball in a Hilbert space (Section 2). The case of a quadratic map is considered as an example in Section 3. The duality theory for nonconvex mathematical programming problems, relying on the new convexity principle, 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. They are based on the convexity of the reachable set for nonlinear system with 'small power control'.

2. Basic Result

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

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

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

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

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

THEOREM 2.1. *If* (1), (2) *hold and* $\varepsilon < \min\{r, \nu/(2L)\}$, *then the image of a ball* $B(a,\varepsilon) = \{x \in X : ||x-a|| \leqslant \varepsilon\}$ *under the map* f *is convex, i.e.* $F = \{f(x) : x \in B(a,\varepsilon)\}$ *is a convex set in* Y.

We need the following results:

LEMMA 2.1. *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 2.2. *Suppose there exist* $L, \rho, \mu > 0$, *such that*

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

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

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

This Lemma coincides with Corollary 1, Theorem 1 of [2].

Proof of Theorem 2.1. Let x_1, x_2 be arbitrary points in $B(a,\varepsilon) \subset B(a,r)$, $y_i = f(x_i) \in F$, $i = 1,2$. 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

$$y_1 = f(x_0) + f'(x_0)(x_1 - x_0) + \epsilon_1, \qquad y_2 = f(x_0) + f'(x_0)(x_2 - x_0) + \epsilon_2,$$

where

$$||\epsilon_i|| \leqslant L||x_i - x_0||^2/2 = L||x_1 - x_2||^2/8, \quad i = 1,2$$

due to (1), see, e.g., [3, Theorem 3.2.12]. Hence

$$y_0 = f(x_0) + \epsilon_0, \quad \epsilon_0 = (\epsilon_1 + \epsilon_2)/2, ||\epsilon_0|| \leqslant L||x_1 - x_2||^2/8.$$

All conditions of Lemma 2.2 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 2.1,

$$||f(x_0) - y_0|| = ||\epsilon_0|| \leqslant L||x_1 - x_2||^2/8 = L\rho\varepsilon \leqslant \rho(\nu - L\varepsilon) = \rho\mu.$$

Moreover,

$$\begin{aligned} ||f'(x)^*y|| &\geqslant ||f'(a)^*y|| - ||(f'(x)^* - f'(a)^*)y|| \\ &\geqslant \nu||y|| - L||x - a||||y|| \geqslant (\nu - L\varepsilon)||y|| = \mu||y|| \quad \text{for } x \in B(x_0, \rho). \end{aligned}$$

Thus Lemma 2.2 provides the desired x^* and the proof of convexity of F is completed. □

Remark 1. We presented the proof, based on Lemma 2.2 (which has been proved in [2] by using a version of the Newton method). Another proof can be obtained by exploiting modern techniques, related to the Ljusternik theorem (see, e.g., [4, Theorem 2.7], [5]). However, the proofs of Ljusternik-like results are also based on the Newton method.

Remark 2. The idea of Theorem 2.1 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 cannot lose convexity for a nonlinear map f, which is close enough to its linearization. The same reasoning explains that the result cannot be extended to Banach spaces, where a ball is not strongly convex.

Remark 3. The result holds, if we replace the ball by any other strongly convex set (e.g. by a nondegenerate ellipsoid).

Remark 4. We can verify some additional properties of F: it is a strictly convex set; it has a nonempty interior which is generated by interior points of $B(a, \varepsilon)$; its boundary is the image of the sphere $||x - a|| = \varepsilon$. Indeed, we have proved that all points in $B(a, \varepsilon)$ are regular for the map f: $||f'(x)^*y|| \geqslant \mu||y||, \mu > 0$, for all y. A regular image of an open set is open thus $\{f(x) : ||x - a|| < \varepsilon\}$ is an open set and cannot contain points of ∂F.

Remark 5. The smoothness assumptions of Theorem 2.1 cannot be seriously relaxed. For instance, A. Ioffe constructed a counter-example with f continuously differentiable, but not in $C^{1,1}$. Then the result is false.

3. Example: Quadratic Transformation

In many cases, the conditions of Theorem 2.1 can be effectively checked and the radius ε of the ball can be estimated. One such example is a quadratic transformation.

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

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

Take $a = 0$, that is $B = \{x : ||x|| \leqslant \varepsilon\}$. Then $f_i'(x) = A_i x + a_i$ and (1) is satisfied on $\mathbf{R}^n$ with

$$L = \left(\sum_{i=1}^m ||A_i||^2\right)^{1/2},$$

where $||A_i||$ stands for the operator norm of matrices A_i. Consider the matrix A with columns a_i: $A = (a_1|a_2|\dots|a_m)$. Then $f'(0)^T y = Ay$, and if 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 2.1 states:

PROPOSITION 3.1. *If $\varepsilon < \nu/(2L)$, then the image of the ball B under the map f is convex: $F = \{f(x) : ||x|| \leqslant \varepsilon\}$ is a convex set in $\mathbf{R}^m$.*

This is in sharp contrast with the results on images of arbitrary balls under quadratic transformations, where the convexity can be validated [6] under some very restrictive assumptions.

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

$$f_1(x) = x_1x_2 - x_1,\ f_2(x) = x_1x_2 + x_2. \tag{4}$$

Then the estimates above guarantee that F is convex for $\varepsilon < \varepsilon^* = 1/(2\sqrt{2}) \approx 0.3536$. It can be directly proved for this case that F is convex for $\varepsilon \leqslant \varepsilon^*$ and loses convexity for $\varepsilon > \varepsilon^*$. Thus, the estimate provided by Proposition 3.1 is tight for this example. Figure 1 shows the images of the ε-discs $\{x \in \mathbf{R}^2 : ||x|| \leqslant \varepsilon\}$ under the mapping (4) for various values of ε.

4. Duality in Local Optimization Problems

Simultaneously, with the standard mathematical programming problem

$$\begin{aligned} &\min f_0(x), \quad x \in \mathbf{R}^n, \\ &f_i(x) \leqslant 0, \quad i = 1, \dots, l, \\ &f_i(x) = 0, \quad i = l+1, \dots, m, \end{aligned} \tag{5}$$

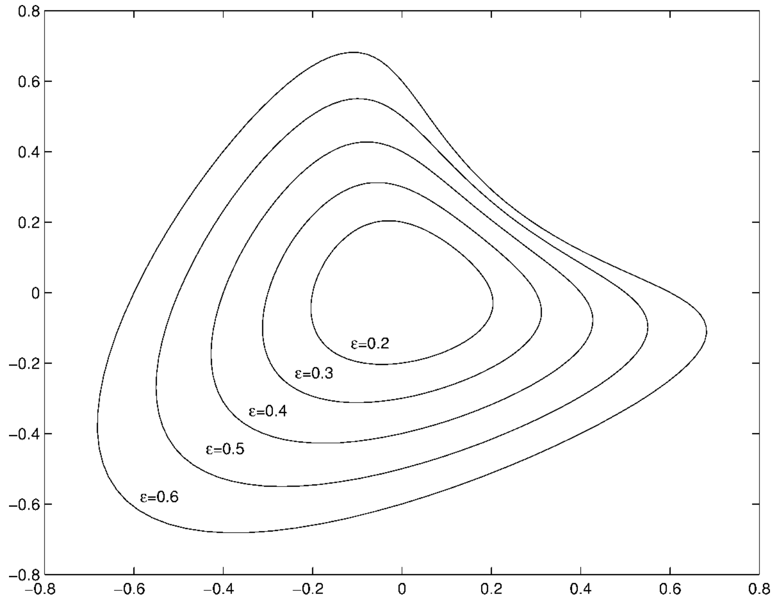

Figure 1. Images of ε-discs for various ε.

consider its local version with the extra constraint

$$\begin{aligned} &\min f_0(x), \quad x \in \mathbf{R}^n, \\ &f_i(x) \leqslant 0, \quad i = 1, \dots, l, \\ &f_i(x) = 0, \quad i = l+1, \dots, m, \\ &||x - a|| \leqslant \varepsilon. \end{aligned} \tag{6}$$

Suppose that the functions $f_i(x), i = 0, 1, \dots, m$ are from $C^{1,1}$ on $B(a, \varepsilon)$. Construct the Lagrange function

$$L(x, y) = \sum_{i=0}^{m} y_i f_i(x). \tag{7}$$

Denote $Y_+ = \{y \in \mathbf{R}^{m+1} : y_i \geqslant 0, i = 0, 1, \dots, l\}$. We assume that a is a feasible point in (5); moreover we can assume without loss of generality that all inequality constraints are active in a: $f_i(a) = 0,\ i = 1, \dots, m$, otherwise they play no role in (6) and can be rejected for ε small enough. Finally, we suppose that the gradients of $f_i(x), i = 0, 1, \dots, m$ at a are linearly independent, i.e. no $y^0 \neq 0$ exists 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 (5). In the presence of inequality constraints, this condition is more restrictive than the assumption 'a is not a Kuhn–Tucker point in problem (5)'. For instance, it implies $m < n$, i.e., the number of active constraints in a is less than the dimension.

THEOREM 4.1. *Under above assumptions, there exists* $\varepsilon^* > 0$ *such that a solution* x^* *of* (6) *with* $0 < \varepsilon < \varepsilon^*$ *exists, is unique, lies on the boundary of* $B(a, \varepsilon)$: $||x^* - a|| = \varepsilon$, *and the following inequality holds*

$$L(x, y^*) \geqslant L(x^*, y^*) \quad \forall x : ||x - a|| \leqslant \varepsilon \tag{8}$$

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

Proof. Problem (6) is equivalent to the optimization problem in the 'image space':

$$\begin{aligned} &\min f_0, \quad f \in F, \quad f_i \leqslant 0, \ i = 1, \ldots, l, \\ &F = \{f(x) : ||x - a|| \leqslant \varepsilon\}, \end{aligned} \tag{9}$$

where

$$f = (f_0, f_1, \ldots, f_m) \in \mathbf{R}^{m+1}, \qquad f(x) = (f_0(x), f_1(x), \ldots, f_m(x)).$$

The point a is a regular point for $f(x)$ because $f_i'(a)$ are linearly independent. Theorem 2.1 guarantees the convexity of F for ε small enough. Thus (9) is a convex problem and, for its solution $f^* = f(x^*)$, there exists a separating hyperplane $0 \neq y^* \in \mathbf{R}^{m+1}$, $(y^*, f) \geqslant 0 \, \forall f : f \in F$, $f_0 \geqslant f_0^*$, $f_i \leqslant 0$, $i = 1, \ldots, l$.

This condition is equivalent to (8). The other statements of the theorem follow from Remark 4 to Theorem 2.1. □

COROLLARY 4.1. *If a point* x^* *is a solution of* (6), *then there exists* $y^* \in Y_+$, $y^* \neq 0$, $y_i^* f_i(x^*) = 0$, $i = 1, \ldots, m$ *such that*

$$x^* = a - \varepsilon \frac{L_x(x^*, y^*)}{||L_x(x^*, y^*)||}. \tag{10}$$

Indeed, (10) is a necessary condition for x^* to be a minimum point of the function $L(x, y^*)$ on $B(a, \varepsilon)$. The factor ε in (10) is due to condition $||x^* - a|| = \varepsilon$.

COROLLARY 4.2. *Introduce*

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

if x^* *is a solution of* (6) *then there exists* $y^* \in Y_+$ *such that*

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

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

THEOREM 4.2. *Suppose that the following regularity condition holds: for any* $\varepsilon > 0, \sigma \in \mathbf{R}^m : \sigma_i = 1, i = 1, \ldots, l, |\sigma_i| = 1, i = l+1, \ldots, m$ *there exists* x_σ *such that*

$$\sigma_i f_i(x_\sigma) < 0, \quad i = 1, \ldots, m, \quad ||x_\sigma - a|| \leqslant \varepsilon. \tag{11}$$

Then, in Theorem 4.1 *we can take* $y_0^* = 1$ *and* (8) *is necessary and sufficient condition for optimality in* (6).

Proof. From (8) we get

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

Take $\sigma : \sigma_i = \operatorname{sign} y_i^*$ 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 (8) is obviously sufficient for optimality if $y_0^* = 1$. □

Regularity condition (11) can be replaced by other ones, e.g. $f_i'(a), i = l+1, \ldots, m$ are linearly independent and there exists $h \in \mathbf{R}^n : (f_i'(a), h) = 0, i = l+1, \ldots, m, (f_i'(a), h) < 0, i = 1, \ldots, l$.

Let's show how these results work for the case of quadratic functions. Consider (6) with $a = 0$ and

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

Suppose that $\alpha_i \leqslant 0, i = 1, \ldots, l, \alpha_i = 0, i = l+1, \ldots, m$ and the assumptions of Proposition 3.1 are satisfied (with obvious changes of notation). Then Theorem 4.1 can be applied,

$$L(x, y) = (1/2)(A(y)x, x) + (a(y), x) + \alpha(y),$$
$$A(y) = \sum_{i=0}^m y_i A_i, \qquad a(y) = \sum_{i=0}^m y_i a_i, \qquad \alpha(y) = \sum_{i=0}^m y_i \alpha_i.$$

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

$$\psi(y) = \min_{||x|| \leqslant \varepsilon} ((A(y)x, x) + 2(a(y), x) + \alpha(y)). \tag{12}$$

This problem is always tractable (even if $A(y)$ is not positive definite), and can be effectively solved [6]. 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 the 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:

$$\min_{||x-a|| \leqslant \varepsilon} f(x) \tag{13}$$

and the iterative method

$$x^{k+1} = x^k - \varepsilon \frac{f'(x^k)}{||f'(x^k)||}. \tag{14}$$

THEOREM 4.3. *Suppose that* $f\colon \mathbf{R}^n \to R^1$ *is* $C^{1,1}$ *on* $B(a, \varepsilon)$*:*

$$||f'(x) - f'(y)|| \leqslant L||x - y||, \quad x, y \in B(a, \varepsilon),\ \varepsilon < ||f'(a)||/(2L). \tag{15}$$

Then

(a) *The solution* x^* *of* (13) *exists and is unique,* $||x^* - a|| = \varepsilon$ *and the necessary and sufficient optimality condition holds:*

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

(b) *Method* (14) *converges with linear rate of convergence for any* $x^0 \in B(a, \varepsilon)$*:*

$$||x^k - x^*|| \leqslant q^k ||x^0 - x^*||, \quad q = \mathrm{O}(\varepsilon) = \frac{\varepsilon L}{||f'(a)|| - \varepsilon L} < 1. \tag{17}$$

Proof. The statement (a) follows from Theorem 4.1 and Corollary 4.1.

If we subtract (16) from (14) we get

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

For any $0 < \tau, x \in \mathbf{R}^n, ||x|| \geqslant \tau$ 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 $\tau = ||f'(a)|| - \varepsilon L$)

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

This is equivalent to the desired estimate (17). □

Note that (14) can be considered as the conditional gradient method [7] for solving (13) 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.

Convexity of the reachable set. A general nonlinear control system

$$\dot{x} = F(x, u, t), \quad x \in \mathbf{R}^n, \quad u \in \mathbf{R}^m, \quad 0 \leqslant t \leqslant T, \quad x(0) = c \tag{18}$$

with L_2-bounded control

$$u \in U = \left\{ u : \int_0^T ||u(t)||^2 \, \mathrm{d}t \leqslant \varepsilon \right\} \tag{19}$$

defines a reachable set

$$S = \{x(T) : x(t) \text{ is a solution of (18)}, u \in U\}. \tag{20}$$

Suppose that the linearized system

$$\dot{z} = F_x(x_0, 0, t)z + F_u(x_0, 0, t)u, \quad y(0) = 0 \tag{21}$$

is controllable [8]; 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 S is convex. Indeed, we can apply Theorem 2.1 with $X = L_2, Y = \mathbf{R}^n$, f: $u \to x(T)$. The controllability of (21) ensures regularity of this map at $u = 0$.

Sufficiency of the maximum principle. Consider the optimal control problem

$$\min \phi(x(T)), \tag{22}$$

where $x(t)$ is a solution of (18) subject to the constraint (19) and terminal time T is fixed and the function ϕ: $\mathbf{R}^n \to R^1$ is convex. Then this optimal control problem is equivalent to the finite-dimensional one $\min_{x\in S} \phi(x)$ which is convex under above conditions. Thus, the first-order necessary conditions for the extremum (which can be written in the form of maximum principle [8]) is also sufficient. Thus, we conclude that the maximum principle is the sufficient condition for optimality for (18), (19), (22).

Also from Theorem 4.1 we obtain that the solution is unique and it reduces (19) to equality.

Numerical methods. Iterative method (14) can be applied to solve the optimal control problem (18), (19), (22). It has the following form. At the kth iteration we have an approximation $u^k = u^k(t), 0 \leqslant t \leqslant T$ and calculate x^k as a solution of (18) 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 (14), where L_2 norm is used. Theorem 4.3 guarantees the convergence of this method to the optimal control.

Discrete maximum principle. It is well known that, in general, the Pontryagin maximum principle is not valid for discrete-time systems [9]. However, it can be validated for 'small power' control.

Let the states $x_k \in \mathbf{R}^n$ and controls $u_k \in \mathbf{R}^m$ be described by nonlinear difference equations

$$x_{k+1} = F(x_k, u_k, k), \ x_0 = c, \ k = 0, 1, \dots, N-1. \tag{23}$$

Our objective is

$$\min \phi(x(N)) \tag{24}$$

subject to the l_2-type constraint

$$\sum_{k=0}^{N-1} ||u_k||^2 \leqslant \varepsilon. \tag{25}$$

Then under the condition of controllability of the linearized system, we can prove (as was done above for the continuous-time case) that the reachable set is convex if ε is small enough. The standard technique allows us to obtain the maximum principle under this convexity assumption.

6. Conclusions

The general 'image convexity' principle is presented and some of its applications to optimization and control are described. It is possible that many more applications will arise in various fields of functional analysis and numerical analysis.

References

1. Rockafellar, R. T.: *Convex Analysis*, Princeton Univ. Press, Princeton, 1970.
2. Polyak, B.: Gradient methods for solving equations and inequalities, *USSR Comput. Math. Math. Phys.* **4**(6) (1964), 17–32.
3. Ortega, J. W. and Rheinboldt, W. C.: *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
4. Dmitruk, A. V., Miljutin, A. A. and Osmolovskii, N. M.: Ljusternik theorem and extremum theory, *Russian Math. Surveys* **55**(6) (1980), 11–46.
5. Ioffe, A. D.: On the local surjection property, *Nonlinear Anal.* **11**(5) (1987), 565–592.
6. Polyak, B. T.: Convexity of quadratic transformations and its use in control and optimization, *J. Optim. Theory Appl.* **99**(3) (1998), 553–583.
7. Bertsekas, D. P.: *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1998.
8. Lee, E. B. and Markus, L.: *Foundations of Optimal Control Theory*, Wiley, New York, 1970.
9. Jordan, B. K. and Polak, E.: Theory of a class of discrete optimal control systems, *J. Electr. Control* **17**(6) (1964).