*Dedicated to*
*Professor Czesław Olech*

**Control and Cybernetics**
vol. **34** (2005) No. 3

# On a nonlocal metric regularity of nonlinear operators

by
**A. V. Dmitruk**

Central Economics & Mathematics Institute, Russian Academy of Sciences
Nakhimovskii prospekt 47, Moscow 117418, Russian Federation
e-mail: dmitruk@member.ams.org

**Abstract:** We consider some versions and generalizations of the classical Lyusternik theorem on the covering property (metric regularity) of nonlinear mappings, study some related properties, and propose nonlocal theorems of the given type, which then are used in the proof of a relaxation theorem for a nonlinear control system with sliding modes and terminal equality constraints.

**Keywords:** nonlinear mapping, covering and metric regularity, Lyusternik iteration process, uniform covering, combined operator, sliding mode controls, weak-* convergence.

## 1. Introduction

Let $F : X \to Y$ be a mapping between Banach spaces $X$ and $Y$. Having $x_0 \in X$ and $y_0 = F(x_0)$, consider the level set

$$M = \{ x \in X \mid F(x) = y_0 \} = F^{-1}(y_0).$$

An important question is to estimate $dist\,(x, M)$ for all $x$ from a neighborhood of the point $x_0$. Historically, this question first arose in the theory of problems of conditional extremum:

$$\varphi(x) \to \min, \qquad \text{subject to} \quad F(x) = 0. \tag{1}$$

For such problems, in this general setting, the first order necessary condition for a local minimum (Lagrange multipliers rule) was proved by L.A. Lyusternik in his famous paper (Lyusternik, 1934). In that paper he actually obtained the following result:

THEOREM 1.1 (on the distance estimate to the level set) *Suppose that $F$ is strictly differentiable at $x_0$ and its derivative $F'(x_0)$ maps $X$ onto $Y$. Then there exists a constant $L$ such that in some neighborhood $\mathcal{O}(x_0)$ of the point $x_0$ the following estimate holds:*

$$dist\,(x, M) \le L\,||F(x) - y_0||, \qquad \forall\, x \in \mathcal{O}(x_0). \tag{2}$$

Recall that the mapping $F:\ X \to Y$ is strictly differentiable at $x_0$ if there exists a bounded linear operator $A:\ X \to Y$ such that $\forall\,\varepsilon > 0$ there exists a neighborhood $U(x_0)$ such that $\forall\, x', x'' \in U(x_0)$

$$||F(x'') - F(x') - A(x'' - x')|| \ \le\ \varepsilon\,||x'' - x'||.$$

The linear operator $A$ is called *the strict derivative* of the mapping $F$ at the point $x_0$ and is denoted by $A = F'(x_0)$. Clearly, if $F$ has at $x_0$ a strict derivative $F'(x_0)$, then it is also its Fréchet derivative. On the other hand, if $F$ has at $x_0$ a Fréchet derivative, then its strict differentiability is not guaranteed. However, if $F$ has a Fréchet derivative $F'(x)$ at each point in some neighborhood $\mathcal{O}(x_0)$, and this derivative is continuous at $x_0$ w.r.t. $x$ in the operator norm (i.e., $||F'(x) - F'(x_0)|| \to 0$ as $x \to x_0$), then one can easily show that $F'(x_0)$ is a strict derivative of $F$ at $x_0$.

Theorem 1.1 readily yields the following "theorem on the tangent subspace", the one often used in deriving necessary conditions for an extremum.

Recall that a vector $h \in X$ is tangent to a set $C \subset X$ at a point $x_0 \in C$ if $dist\,(x_0 + \varepsilon h, C) = o(\varepsilon)$ as $\varepsilon \to 0+$. The set of all such vectors (it is always a closed cone) is denoted by $T_{x_0} C$.

THEOREM 1.2 (Lyusternik) *Let $F'(x_0)$ be onto (the Lyusternik condition). Then $T_{x_0} M = ker\, F'(x_0)$, i.e., a vector $\bar{x}$ is tangent to the level set $M$ at $x_0$ iff $F'(x_0)\,\bar{x} = 0$.*

The inclusion $\subset$ here is obvious, and it holds even without the assumption about surjectivity of $F'(x_0)$. The reverse inclusion follows readily from the distance estimate (2): if $F'(x_0)\,\bar{x} = 0$, then $F(x_0 + \varepsilon\bar{x}) = F(x_0) + o(\varepsilon)$, and so

$$dist\,(x_0 + \varepsilon\bar{x}, M) \le L\,||F(x_0 + \varepsilon\bar{x}) - F(x_0)|| = o(\varepsilon).$$

## 2. Covering and metric regularity

Analyzing more thoroughly the proof in Lyusternik (1934), one can see that, under the conditions of Theorem 1.1, a stronger assertion actually holds.

THEOREM 2.1 (on the distance estimate to variable level sets) *Let the conditions of Theorem 1.1 be fulfilled. Then there exist a constant $L$ and neighborhoods $U(x_0)$ and $V(y_0)$ (of the points $x_0$ and $y_0$ respectively) such that* $\forall\, x \in U(x_0),\ \forall\, y \in V(y_0)$

$$dist\,(x, F^{-1}(y)) \le L\,||F(x) - y||. \tag{3}$$

This property of the mapping $F$ (noted by many researchers) has been recently named *metric regularity* with constant $L$. One can easily see that it is equivalent to the following property: there exist a number $a > 0$ and a neighborhood $\mathcal{O}(x_0)$ such that for any closed ball $B_r(x) \subset \mathcal{O}(x_0)$

$$F(B_r(x)) \supset B_{ar}(F(x)). \tag{4}$$

The last property is called *covering with (*or *openness at) linear rate a* in the neighborhood $\mathcal{O}(x_0)$.

The equivalence of these two properties for an arbitrary continuous $F$ means that, first, if $F$ covers in a neighborhood $\mathcal{O}(x_0)$ with a rate $a > 0$, then for some neighborhoods $U(x_0)$ and $V(y_0)$ it is metrically regular with the constant $L = 1/a$, and, second, if $F$ is metrically regular with a constant $L$ for some neighborhoods $U(x_0)$ and $V(y_0)$, then in some neighborhood $\mathcal{O}(x_0)$ it covers with any rate $a < 1/L$.

In the simplest case when $F: \mathbb{R} \to \mathbb{R}$ is a scalar function of one variable, both properties – covering and metric regularity in a neighborhood – mean that in this neighborhood $F'(x) \geq a, \quad F'(x) \geq 1/L, \quad$ and $a = 1/L$. A slight difference occurs in the general case because the $\mathit{dist}\,(x, F^{-1}(y))$ may be not attained.

The author's opinion is that among these two properties, the more convenient *for application*, i.e., for usage in concrete situations, is metric regularity (the distance estimate to the level sets), whereas the more convenient *to prove* is covering.

Different versions and generalizations of Theorem 2.1 (including those for metric and quasimetric spaces, for nonsmooth and set-valued mappings) were considered by many authors. We do not give here a survey of this; see e.g. papers Dmitruk, Milyutin and Osmolovskii (1980), Borwein and Zhuang (1988), Penot (1989), Ioffe (2000, 2001), Dontchev and Rockafellar (2004) providing a large number of references.

One of the most convenient and useful generalizations is the abstract version of Lyusternik theorem proposed by A.A. Milyutin. (Note that Milyutin himself said that he did not generalize the Lyusternik's theorem, but only put it in a proper formulation, purifying it from inessential details.) His formulation is as follows:

Let $X$ be a complete metric space, $Y$ be a vector space with a metric invariant w.r.t. translation (e.g., a normed space), $G$ be a set in $X$, and $T: X \to Y$ be a mapping. (We denote the metrics in $X$ and $Y$ by the same letter $d$, and the ball $B_r(x)$ is sometimes denoted by $B(x, r)$.)

Definition 2.1 *The mapping $T$ covers on $G$ with rate $a > 0$ if*

$$\forall\, B_r(x) \subset G \qquad T(B_r(x)) \supset B_{ar}(T(x)). \tag{5}$$

Now, let another mapping $S: X \to Y$ be also given.

DEFINITION 2.2 *The mapping $S$ contracts on $G$ with rate $b \geq 0$ if*

$$\forall B_r(x) \subset G \qquad S(B_r(x)) \subset B_{br}(S(x)). \tag{6}$$

(Obviously, any such mapping is continuous on $G$. On the other hand, any mapping $b$-Lipschitzian on $G$ contracts on $G$ with rate $b$.)

THEOREM 2.2 (Milyutin, see Dmitruk, Milyutin and Osmolovskii, 1980) *Let $T$ be continuous on $G$ and cover on $G$ with a rate $a > 0$, and let $S$ contract on $G$ with the rate $b < a$. Then their sum $F = T + S$ covers on $G$ with the rate $a - b > 0$. (The assumption of continuity of $T$ can be weakened to the closedness of its graph on $G$. In this case the graph of $F$ on $G$ is also closed.)*

*Proof.* The proof is so important and at the same time transparent, that it worth to be given here completely. Take any ball $B(x_0, \rho) \subset G$. We must show that

$$F(B(x_0, \rho)) \;\supset\; B(F(x_0), (a-b)\rho).$$

Without loss of generality, assume that $a = 1$ and $b < 1$. Denote for brevity $y_0 = F(x_0)$, $\quad r = (1-b)\rho$. Take any $\hat{y} \in B(y_0, r)$. We have to show that $\exists\, \hat{x} \in B(x_0, \rho)$ such that $F(\hat{x}) = \hat{y}$.

The point $\hat{x}$ will be obtained as the limit of a sequence $\{x_n\}$, which will be generated now by a special iteration process.

At the beginning, we have the following situation:

$$T(x_0) + S(x_0) = y_0\ , \tag{7}$$

and we need to obtain $T(\hat{x}) + S(\hat{x}) = \hat{y}$. Rewrite equation (7) in the form $T(x_0) = y_0 - S(x_0)$ and use the 1-covering of mapping $T$. Since

$$d\,(\hat{y} - S(x_0),\, y_0 - S(x_0)) \;=\; d\,(\hat{y},\, y_0) \;\leq r,$$

and $B(x_0, r) \subset G$, there exists $x_1 \in B(x_0, r)$, such that $T(x_1) = \hat{y} - S(x_0)$, i.e.,

$$T(x_1) + S(x_0) = \hat{y}\ . \tag{8}$$

Now, replace here $S(x_0)$ by $S(x_1)$. Since the mapping $S$ is $b$-contracting on the ball $B(x_0, r)$, we have $d(S(x_1), S(x_0)) \leq br$, and so

$$T(x_1) + S(x_1) = y_1\ , \tag{9}$$

where $d(\hat{y}, y_1) \leq br$.

So, we moved from equation (7) for a "base" point $x_0$ to equation (9) for a new "base" point $x_1$, where

$$d(x_0, x_1) \leq r, \qquad d(\hat{y}, y_1) \leq br.$$

Consider now equation (9) and try to replace $y_1$ by $\hat{y}$. Since

$$r + br < r\,(1 + b + b^2 + \ldots) = r\,\frac{1}{1-b} = \rho,$$

the ball $B(x_1, br)$ is contained in the ball $B(x_0, \rho)$; hence the 1–covering of $T$ and $b$-contracting of $S$ hold on $B(x_1, br)$. Then, by analogy with the preceding step, there exists $x_2 \in B(x_1, r)$, such that

$$T(x_2) + S(x_2) = y_2\,, \qquad d(\hat{y}, y_2) \le b^2 r,$$

and so on. Continuing this process infinitely, we obtain a sequence of points $x_n,\ y_n$ such that

$$F(x_n) = T(x_n) + S(x_n) = y_n\,, \tag{10}$$

$$d(x_{n-1}, x_n) \le b^{n-1} r, \qquad d(\hat{y}, y_n) \le b^n r. \tag{11}$$

Moreover, we have

$$\begin{aligned} d(x_0, x_n) + b^n r \ &\le\ d(x_0, x_1) + d(x_1, x_2) + \ldots + d(x_{n-1}, x_n) + b^n r \ \le \\ &\le\ r + br + \ldots + b^{n-1} r + b^n r \ <\ r\,\frac{1}{1-b} = \rho, \end{aligned} \tag{12}$$

whence the ball $B(x_n, b^n r)$ is contained in the initial ball $B(x_0, \rho)$, which makes the next step possible.

Consider the obtained sequence $\{x_n\}$. The first inequality in (11) implies that it is fundamental (i.e., a Cauchy sequence), and since $X$ is complete, this sequence has a limit $\hat{x}$. By (12) we get $d(x_0, \hat{x}) \le \rho$, i.e., $\hat{x} \in B(x_0, \rho)$. The second inequality in (11) implies that $y_n \to \hat{y}$, and then, from (10) and continuity of $F$ on the initial ball (or from the closedness of its graph) we get $F(\hat{x}) = \hat{y}$, which is exactly what was required. ■

The iteration process in this proof is much similar to that in the Newton method: the role of derivative $F'$, involved in the Newton method, is played in our case by the mapping $T$, and the role of the small nonlinear residual is played by the mapping $S$. The covering property of the mapping $T$ allows us to "solve" equation (8) with respect to $x_1$, while the small additional term $S$ knocks us off the desired goal $\hat{y}$. This abstract Newton-like method is called *the Lyusternik iteration process*. Note however, that this process does not completely coincide with the Newton method, because the mapping $T$ is not one-to-one in general, and therefore, equation (8) is not solved uniquely, in contrast with the Newton method. So, the Lyusternik process is more general than the Newton method. For example, the Lyusternik iteration process is actually used in the standard proof of the Banach open mapping theorem, whereas the Newton method cannot be used there.

Note by the way the following simple fact.

Lemma 2.1 *If a space $X$ is complete, $T : X \to Y$ is continuous and covers on $G \subset X$ with a rate $a > 0$, and $G$ has nonempty interior, then $Y$ is complete, too.*

(Thus, in Theorem 2.2 the space $Y$ is also complete.)

*Proof.* By definition, the set $G$ contains a ball $B_\varepsilon(x_0)$ of radius $\varepsilon > 0$. Let $Tx_0 = y_0$. Since the metric in $Y$ is invariant w.r.t. translation, it is enough to show that the ball $B_{a\varepsilon/2}(y_0)$ is complete as a metric space in the metric of $Y$.

Without loss of generality, assume that $a = 1$. Take any fundamental sequence $y_n \in B_{\varepsilon/2}(y_0)$. It suffices to show that it has a limit point $\hat{y} \in B_{\varepsilon/2}(y_0)$ (then all the sequence $y_n \to \hat{y}$ as well). Passing to a subsequence, we can assume that

$$\sum_{n=1}^{\infty} d(y_n, y_{n+1}) < \varepsilon/2\,.$$

Let us show that this sequence has a limit in $B_{\varepsilon/2}(y_0)$.

Denote $r_n = d(y_n, y_{n+1})$. Since $y_1 \in B(y_0, \varepsilon/2)$, and $T$ covers with rate 1, there exists $x_1 \in B(x_0, \varepsilon/2)$ such that $Tx_1 = y_1$.

Further, since $d(x_0, x_1) + r_1 < \varepsilon$, then $B(x_1, r_1) \subset B_\varepsilon(x_0)$, and by the 1-covering of $T$ (on $G$) there exists $x_2 \in B(x_1, r_1)$ such that $Tx_2 = y_2$. Continuing this process to infinity, we obtain a sequence of points $x_n \in B(x_{n-1}, r_{n-1})$ such that $Tx_n = y_n$.

Moreover, we have $d(x_n, x_{n+1}) \le r_n$, therefore

$$\begin{aligned} d(x_0, x_{n+1}) \;&\le\; d(x_0, x_1) + d(x_1, x_2) + \ldots + d(x_n, x_{n+1}) \;\le \\ &\le\; \frac{\varepsilon}{2} + r_1 + \ldots + r_n < \;\frac{\varepsilon}{2} + \frac{\varepsilon}{2} \;= \varepsilon, \end{aligned}$$

and hence the sequence $x_n$ is fundamental in the ball $B_\varepsilon(x_0)$. Since $X$ is complete, $x_n$ converges to a point $\hat{x} \in B_\varepsilon(x_0)$. But then, from the continuity of $T$ we get

$$y_n = Tx_n \;\to\; \hat{y} = T\hat{x},$$

and since all $y_n$ lie in the ball $B(y_0, \varepsilon/2)$, their limit $\hat{y}$ lies in the same ball. Thus, $y_n \to \hat{y} \in B(y_0, \varepsilon/2)$, Q.E.D. ■

The classical Lyusternik theorem (Theorem 2.1) easily follows from Theorem 2.2. Represent $F$ in the form

$$F(x) = \; F(x_0) + F'(x_0)(x - x_0) \; + S(x),$$

and set $T(x) = F(x_0) + F'(x_0)(x - x_0)$. Since $F'(x_0)$ is onto, by the Banach open mapping theorem $\exists\, a > 0$ such that $F'(x_0)\, B_1(0) \supset B_a(0)$, which implies that the affine operator $T$ covers with rate $a$ on the whole space $X$. The strict differentiability of $F$ at $x_0$ exactly means that $\forall\, \varepsilon > 0$ the residual $S$ is $\varepsilon$-Lipschitzian in some neighborhood $\mathcal{O}(x_0)$. Choosing $\varepsilon < a$, we get by Theorem 2.2 that $F$ covers in $\mathcal{O}(x_0)$ with rate $a - \varepsilon > 0$, Q.E.D. ■

The following two features in Theorem 2.2 are important:

a) the covering rate of the resulting mapping $F$ is given explicitly: $a - b$.

b) the covering of $F$ is obtained on the same set $G$, not on a smaller set.

These features allow for obtaining the covering and distance estimates not just in a small neighborhood of a given point (i.e., locally), but on a "large" set $G$ (nonlocally), and moreover, for obtaining these properties not for a given single mapping $F$, but uniformly for a family of mappings (and hence, to pass to perturbation stability of these properties). We will turn to this issue a bit later. Now we give yet another generalization of the Lyusternik's theorem (close in the spirit to the recent notion of porosity).

Let $X$ be a complete metric space, $Y$ a metric space, $F : X \to Y$, and $G \subset X$.

We say that a set $A \subset Y$ is an $\varepsilon$-net for a set $B \subset Y$, if $\forall\, b \in B \;\; \exists\, a \in A$ such that $d(a, b) \le \varepsilon$.

THEOREM 2.3 (Dmitruk, see Dmitruk, Milyutin and Osmolovskii, 1980) *Let there exist numbers $a > b \ge 0$ such that for any ball $B_\rho(x) \subset G$ the mapping $F$ is continuous on this ball (or just has a closed graph), and for $r = (1 - b/a)\rho$ the set $F(B_r(x))$ is a $br$-net for the ball $B_{ar}(F(x))$. Then $F$ covers on $G$ with rate $a - b$.*

Note that here we do not a priori have any covering mapping $T$ and perturbing additive $S$; all the time we have only one mapping $F$. Theorem 2.3 asserts that its "almost covering" guarantees its real covering. Note also that Theorem 2.2 follows from Theorem 2.3 (one should apply the last theorem to the mapping $F = T + S$). [1]

*Proof.* The proof is similar to that of Theorem 2.2 with some alterations. As before, set $a = 1$, $\;r = (1-b)\rho$, take any $\hat{y} \in B(y_0, r)$ and try to find $\hat{x} \in B(x_0, \rho)$ such that $F(\hat{x}) = \hat{y}$. Again, the point $\hat{x}$ will be obtained as the limit of a sequence, generated now by the following iteration process.

At the beginning, we now have the following situation:

$$F(x_0) = y_0\,, \qquad d(y_0\,, \hat{y}) \le r. \tag{13}$$

Since $F(B_r(x_0))$ is a $br$-net for the ball $B_{ar}(y_0))$, there exists $x_1 \in B(x_0, r)$ such

[1] The author, together with A.A. Milyutin and N.P. Osmolovskii, presented Theorems 2.2, 2.3 and other results of Dmitruk, Milyutin and Osmolovskii (1980) to L.A. Lyusternik during one of his last visits to Mechanical-Mathematical Department of Moscow State University in 1980.

that

$$F(x_1) = y_1\,, \qquad d(y_1\,, \hat{y}) \le br. \tag{14}$$

So, we moved from equation (13) for a "base" point $x_0$ to equation (14) for a new "base" point $x_1$. Since

$$d(x_0\,, x_1) + br < \; r\,(1 + b + b^2 + \ldots) \;=\; r\,\frac{1}{1-b} \;= \rho,$$

the ball $B(x_1, br)$ is contained in the initial ball $B(x_0, \rho)$; hence we can iterate the procedure: the image $F(B(x_1\,, br))$ is a $b^2r$-net for the ball $B(y_1\,, br))$, and then there exists $x_2 \in B(x_1\,, br)$, such that

$$F(x_2) = y_2\,, \qquad d(y_2\,, \hat{y}) \le b^2 r, \tag{15}$$

and so on. The remaining part of the proof repeats the proof of Theorem 2.2. ■

A first simple and well known fact about uniform covering of a family of mappings (or, in other words, the simplest fact about perturbation stability of covering) is given by the following

LEMMA 2.2 *If a linear bounded operator $F_0 : X \to Y$ between Banach spaces $X$ and $Y$ is onto, then there exist constants $\varepsilon > 0$ and $c > 0$ such that any linear operator $F$ with $||F - F_0|| < \varepsilon$ covers (on the whole space) with rate $c$.*

*Proof.* Since $F_0$ is onto, it covers with a rate $a > 0$, and then, for any $\varepsilon < a$ and any linear $F$ with $||F - F_0|| < \varepsilon$ we obtain by Theorem 2.2 that $F = F_0 + (F - F_0)$ covers with rate $a - \varepsilon$. Taking $\varepsilon = a/2$, we get $c = a/2$. ■

As a practically useful corollary, we obtain the following

THEOREM 2.4 *Let $X$ and $Y$ be Banach spaces, and $F_\alpha : X \to Y$ be linear bounded operators, indexed by $\alpha$ from a topological space $\mathcal{A}$. Suppose that for some $\alpha_0 \in \mathcal{A}$ the operator $F_{\alpha_0}$ is onto, and $||F_\alpha - F_{\alpha_0}|| \to 0$ as $\alpha \to \alpha_0$. Then there exist a neighborhood $\mathcal{O}(\alpha_0)$ and a constant $c > 0$, such that $\forall\, \alpha \in \mathcal{O}(\alpha_0)$ the operator $F_\alpha$ covers with rate $c$.*

*Proof.* The proof follows from Lemma 2.2 and the fact that $\forall\, \varepsilon > 0$ there exists a neighborhood $\mathcal{O}(\alpha_0)$ such that $||F_\alpha - F_{\alpha_0}|| < \varepsilon$ for all $\alpha \in \mathcal{O}(\alpha_0)$. ■

All the above is rather well known and given here just for completeness of exposition. Let us now present a connection between the nonlocal covering (i.e., covering on a given set $G$) and the distance estimate to the level set of the mapping. We confine ourselves to the fixed (zero) level as the most essential case.

THEOREM 2.5 *Let $X$ be a complete metric space, $Y$ a normed space, and a mapping $F : X \to Y$ cover with rate $a > 0$ on a set $G \subset X$. Suppose that the set $M = \{ x \in G \mid F(x) = 0 \}$ is nonempty.*

*Let a bounded set $\Omega \subset G$ and a number $\delta > 0$ be such that $\mathcal{O}_\delta(\Omega) \subset G$. Then there exists a constant $L$ such that $\forall\, x \in \Omega$ the following estimate holds:*

$$\mathit{dist}\,(x, M) \;\leq\; L\,||F(x)||. \tag{16}$$

*Proof.* Take any point $x \in G$. From the $a$-covering of $F$ it follows that, if the ball

$$B(x,\, ||F(x)||/a) \subset G, \tag{17}$$

then its image contains the ball $B(F(x),\, ||F(x)||)$, which in turn obviously contains zero, and therefore $\exists\, x' \in B(x,\, ||F(x)||/a)$ such that $F(x') = 0$. Since this $x' \in M$, then

$$\mathit{dist}\,(x, M) \leq d(x, x') \leq \frac{1}{a}\,||F(x)||. \tag{18}$$

Consider now the set $\Omega$ with $\mathcal{O}_\delta(\Omega) \subset G$. Take any point $x \in \Omega$. Then $B_\delta(x) \subset G$. If for this point $||F(x)|| < \delta a$, then we have $B(x,\, ||F(x)||/a) \subset B_\delta(x) \subset G$, so inclusion (17) and hence estimate (18) hold, i.e., the required estimate (16) holds with $L = 1/a$.

If, on the contrary, $||F(x)|| \geq \delta a$, then, since $M$ is nonempty, we take an arbitrary point $x_0 \in M$, and since $\Omega$ is bounded, $\Omega \subset B(x_0, R)$ for some radius $R$, and hence we obtain a trivial estimate

$$\mathit{dist}\,(x, M) \;\leq\; d(x, x_0) \;\leq\; R \;=\; \frac{R}{\delta a}\,\delta a \;\leq\; \frac{R}{\delta a}\,||F(x)||,$$

i.e., in this case the estimate (16) holds with $L = R/(\delta a)$. It remains now to set $L = \max\{\, 1/a,\; R/(\delta a)\,\}$. ■

Now, a natural question arises: how to obtain a nonlocal covering? Let us pass to this issue.

## 3. From the local to a nonlocal covering

The following assertion allows one to pass from the local covering (i.e., covering in a neighborhood of a point) to the covering on a "macro" set.

LEMMA 3.1 *Let be given a complete metric space $X$, a normed space $Y$, a mapping $F : X \to Y$, and an open set $G \subset X$. Let $a > 0$ be such that $\forall\, x \in G$ the following property ($a$-covering at the point $x$) holds: $\forall\, \varepsilon > 0 \quad \exists\, \delta \in (0, \varepsilon)$ such that $F(B_\delta(x)) \supset B_{a\delta}(F(x))$. Then $F$ covers with this rate $a$ on the whole set $G$.*

*Proof.* Consider any ball $B_r(x_0) \subset G$, and set $F(x_0) = y_0$. We must show that $F(B_r(x_0)) \supset B_{ar}(y_0)$.

Take an arbitrary $y_1 \in B_{ar}(y_0)$. We have to show that $\exists\, x_1 \in B_r(x_0)$, for which $F(x_1) = y_1$. Consider the segment $I = [y_0,\, y_1]$ with the parameterization

$$y_t = \; y_0 + t(y_1 - y_0), \qquad 0 \le t \le 1.$$

By this, the segment $I$ is equipped with a linear order. In the product $B_r(x_0) \times I$ define the set $Q$ consisting of all pairs $(x, y)$ for which $F(x) = y$ and $\rho(x_0, x) \le \frac{1}{a}\,\rho(y_0, y)$ (the corresponding points $y$ can be called "properly covered"). On the set $Q$ define the following partial order relation:

$$(x', y') \preceq (x'', y''), \;\text{ if } y' \le y'' \text{ and } \rho(x', x'') \le \frac{1}{a}\,\rho(y', y'').$$

(Its transitivity is obvious.) Since the segment $I$ is compact, the space $X$ is complete, the ball $B_r(x_0)$ is closed, and the mapping $F$ is continuous, then any increasing chain $\{(x_\alpha, y_\alpha)\}$ w.r.t. this order has an upper bound. (Passing to a countable cofinal subchain $\{(x_{\alpha_n}, y_{\alpha_n})\}$, we obtain a fundamental sequence $y_{\alpha_n}$, whence the sequence $x_{\alpha_n}$ is fundamental too, and so we can take their limits.) Then, by the Zorn lemma, $Q$ contains a maximal element $(\hat{x}, \hat{y})$. We claim that $\hat{y} = y_1$.

Indeed, by the assumption of the lemma, for the point $\hat{x}$ the $a$-covering of $F$ holds: there exist arbitrarily small $\delta > 0$, such that the image of $B_\delta(\hat{x})$ contains the ball $B_{a\delta}(\hat{y})$. If we suppose $\hat{y} < y_1$, then one can take such a small $\delta > 0$ that a bit "farther" point $y' = \hat{y} + a\delta(y_1 - y_0)/||y_1 - y_0||$ still belongs to the segment $I$. Then one obtains an $x' \in B_\delta(\hat{x})$ such that $F(x') = y'$. Since $(\hat{x}, \hat{y}) \in Q$ and $\rho(x', \hat{x}) \le \frac{1}{a}\,\rho(y', \hat{y})$, the pair $(x', y') \in Q$. Moreover, this pair is strictly greater than the pair $(\hat{x}, \hat{y})$, which contradicts the maximality of the last one.

Thus, the point $(\hat{x}, \hat{y}) \in Q$ is such that $\hat{y} = y_1$, and then, by the definition of $Q$, for $x_1 = \hat{x}$ we have $F(x_1) = y_1$ and $\rho(x_0, x_1) \le \frac{1}{a}\,\rho(y_0, y_1)$, Q.E.D. ■

Replacing the assumption of $a$-covering at each point $x$ in this lemma by a stronger assumption of local $a$-covering, we get the following assertion, simpler in formulation and quite enough for practical use.

THEOREM 3.1 *Suppose $\exists\, a > 0$, such that any point $x \in G$ has a neighborhood $\mathcal{O}(x)$, in which $F$ covers with rate $a$. Then $F$ covers with this rate $a$ on the whole set $G$.*

In the case when $X$ is a Banach space, and the mapping $F$ is differentiable, there is a simple and efficient Lyusternik condition guaranteeing the local covering. Therefore, we get the following result on "nonlocal" covering.

LEMMA 3.2 *Let $X$, $Y$ be Banach spaces, and $F:\; X \to Y$ be strictly differentiable on a set $\Omega \subset X$. Let $a > 0$ be such that $\forall\, x \in \Omega$ the linear operator $F'(x)$ covers with rate $a$. (We say in this case that Lyusternik condition holds* *uniformly on* $\Omega$.*) Then* $\forall\, a' < a$ *there exists an open set* $G \supset \Omega$ *on which the nonlinear operator* $F$ *covers with rate* $a'$.

*Proof.* Take an arbitrary $a' < a$. Let $x \in \Omega$. By Theorem 2.1 there exists a neighborhood $\mathcal{O}(x)$, on which $F$ covers with rate $a'$. Define a set $G = \bigcup_{x\in\Omega} \mathcal{O}(x)$. It is an open set, each point $x$ of which has a neighborhood $V(x)$, on which $F$ covers with rate $a'$. Hence, by Theorem 3.1 $F$ covers with this rate $a'$ on the whole set $G$. ∎

As a simple corollary of this lemma, we get the following

THEOREM 3.2 *Let* $F:\ X \to Y$ *be strictly differentiable on an open set* $G \subset X$. *Let* $a > 0$ *be such that* $\forall\, x \in G$ *the linear operator* $F'(x)$ *covers with rate* $a$ *Then* $\forall\, a' < a$ *the nonlinear operator* $F$ *covers on* $G$ *with rate* $a'$.

Note that Theorems 3.1 and 3.2 have a nonlocal character.

Now, the above question can be stated as follows: how a "broad" set $G$ (bigger than just a neighborhood) with a uniform covering can appear? The simplest case is provided by the following "strengthened version" of the classical Lyusternik theorem (Theorem 2.1).

Let, as before, be given Banach spaces $X$, $Y$, and a mapping $F:\ X \to Y$.

THEOREM 3.3 ( Dmitruk, Milyutin and Osmolovskii, 1980) *Let* $X$ *be equipped with another topology* $\tau$, *weaker (in the nonstrict sense) than its norm topology. Suppose that in some* $\tau$*-neighborhood* $\mathcal{O}_\tau(x_0)$ *of a point* $x_0$ *the mapping* $F$ *is strictly differentiable, and its derivative* $F'(x)$ *is* $\tau$*-continuous at* $x_0$, *i.e.,* $||F'(x) - F'(x_0)|| \to 0$ *as* $x \to x_0$ *w.r.t.* $\tau$. *Suppose also that* $F'(x_0)$ *maps onto. Then in some* $\tau$*-neighborhood* $\mathcal{V}(x_0)$ *the mapping* $F$ *covers with some rate* $a > 0$.

*Proof.* Since $F'(x_0)$ is onto, it covers with some $a_0 > 0$. For any $\varepsilon > 0$, in some $\mathcal{V}_\tau(x_0)$ we have $||F'(x) - F'(x_0)|| < \varepsilon$, hence the linear operator $F'(x) = F'(x_0) + (F'(x) - F'(x_0))$ covers with rate $a_0 - \varepsilon$. Then, by Theorem 3.2 the nonlinear $F$ covers on this $\mathcal{V}_\tau(x_0)$ with rate $a_0 - 2\varepsilon > 0$. ∎

Note that the existence of a "broad" set $G = \mathcal{V}_\tau(x_0)$ in this theorem is proved, not assumed! In fact, its existence is hidden in the assumption about $\tau$-continuity of $F'(x)$, but this assumption seems more natural than the direct existence of a required $\mathcal{V}_\tau(x_0)$, and is more easy to verify; it can really hold in some cases.

Now we point out a class of operators having $\tau$-continuous derivative with respect to a weaker topology than the standard norm topology.

EXAMPLE 3.1 Consider a mapping $F:\ L_\infty \times L_\infty[0,T] \longrightarrow L_1[0,T]$ of the form

$$(\alpha,\, u) \;\longmapsto\; \alpha(t)\, f(t, u(t)), \tag{19}$$

where for simplicity $f$ and $f_u$ are continuous in $(t, u)$. Its Fréchet derivative maps as follows:

$$F'(\alpha, u)(\bar\alpha, \bar u) = \;\bar\alpha\, f(t, u) + \alpha\, f_u(t, u)\bar u,$$

and so

$$||F'(\alpha,u) - F'(\alpha_0,u_0)|| = \\ = \sup_{||\bar\alpha||_\infty \le 1} ||(f(t,u) - f(t,u_0))\,\bar\alpha||_1 \ + \sup_{||\bar\alpha||_\infty \le 1} ||(\alpha\, f_u(t,u) - \alpha_0\, f_u(t,u_0))\,\bar u||_1\,. \tag{20}$$

Define the topology $\tau$ generated by the norm $||\alpha||_1 + ||u||_\infty$ in the space $L_\infty \times L_\infty$, which obviously is weaker than the standard norm of this space. Let us show that if $||\alpha - \alpha_0||_1 + ||u - u_0||_\infty \to 0$, then the right hand side in (20) tends to zero. For the first summand it is obvious, and for the second one it follows from the estimate

$$||(\alpha f_u(t,u) - \alpha_0 f_u(t,u_0))\,\bar u||_1 \ \le \\ \le \ ||(\alpha f_u(t,u) - \alpha f_u(t,u_0))\,\bar u||_1 + ||(\alpha f_u(t,u_0) - \alpha_0 f_u(t,u_0))\,\bar u||_1\ ,$$

and therefore, supremum of this expression over $||\bar u||_\infty \le 1$ is estimated as

$$\le \ ||\alpha||_1 \cdot ||f_u(t,u) - f_u(t,u_0)||_\infty + ||\alpha - \alpha_0||_1 \cdot ||f_u(t,u_0)||_\infty \to 0.$$

Thus, $F'(\alpha,u)$ is continuous w.r.t. the norm $||\alpha||_1 + ||u||_\infty$. If one adds to this mapping $F$ any linear operator $A$ such that $A + F'(x_0)$ is onto, then one gets an operator satisfying the conditions of Theorem 3.3.

One can note that the key point in this example is that the family of linear operators

$$P_\alpha :\ \bar u \longmapsto \alpha(t)\, f_u(t,u_0(t))\, \bar u(t) \tag{21}$$

continuously depends on $\alpha$ w.r.t. $||\alpha||_1$.

Let us try to weaken more the topology for the mapping (19). Namely, let $\alpha$ be taken from a bounded set $\mathcal{A} \subset L_\infty$, and instead of the topology $||\alpha||_1$ consider now the weak-* topology for $\alpha$ (corresponding to the convergence on elements of $L_1$); denote it by $\sigma^*$. Obviously $\sigma^*$ is weaker than $||\alpha||_1$. Thus, in the space $L_\infty \times L_\infty$ we now have a topology $\tilde\tau$, which is the product of $\sigma^*$-topology for $\alpha$ and $||u||_\infty$ for $u$. However, $F'$ is not continuous in this weakened topology, because the above family (21) is not continuous with respect to weak-* convergence of $\alpha$. Let us check it.

Example 3.2 Consider the family of linear operators

$$P_\alpha :\ L_\infty \longrightarrow L_1\,, \qquad \bar u(t) \longmapsto \alpha(t)\, B(t) \bar u(t),$$

where the functional parameter $\alpha$ is taken from a bounded set $\mathcal{A} \subset L_\infty$, and $B(t)$ is a given function from $L_\infty[0,T]$ not identically zero. Assume for simplicity that $\alpha_0(t) \equiv 0 \in int\,\mathcal{A}$, and so $P_{\alpha_0} \equiv 0$. Let $\alpha \in \mathcal{A}$ and $\alpha \xrightarrow{\sigma^*} 0$. Then

$$||P_\alpha|| = \sup_{||\bar u||_\infty \le 1} \int_0^T |\alpha B u|\, dt \ = \int_0^T |\alpha B|\, dt\ ,$$

and this value does not tend to zero, in general. For example, if one takes $|\alpha(t)| = c = \text{const} > 0$, then one obtains

$$||P_\alpha|| = c\int_0^T |\,B(t)\,|\,dt \ = \text{const} > 0.$$

This effect can be seen even in a simpler situation.

EXAMPLE 3.3 For the same $\alpha \in \mathcal{A}$ and $B(t)$ consider the family of linear functionals

$$\varphi_\alpha :\ L_\infty \longrightarrow \mathbb{R}\,, \qquad \bar{u} \longmapsto \int_0^T \alpha\, B\bar{u}\, dt\,. \tag{22}$$

If $\alpha \xrightarrow{\sigma*} 0$ and $|\alpha(t)| = c = \text{const} > 0$, then

$$||\varphi_\alpha|| = \sup_{||\bar{u}||_\infty \le 1} |\int_0^T \alpha B u\, dt| = \int_0^T |\alpha B|\, dt = c\int_0^T |B|\, dt > 0,$$

i.e., $||\varphi_\alpha||$ does not tend to zero.

However, in spite of the lack of continuity of $\varphi_\alpha$ in the operator norm as $\alpha \xrightarrow{\sigma*} \alpha_0$, the uniform covering of $\varphi_\alpha$ for $\alpha$ in some $\sigma^*$-neighborhood $\mathcal{O}(\alpha_0)$ holds, if the functional $\varphi_{\alpha_0}$ covers (i.e., if it is not zero). We prove this simple fact in the following abstract setting.

Let $\mathcal{A}$ be a topological space, and $\forall\, \alpha \in \mathcal{A}$ be given a linear operator $P_\alpha :$ $X \to \mathbb{R}^q$ from a Banach space to a finite-dimensional space.

THEOREM 3.4 *Let for some $\alpha_0 \in \mathcal{A}$ the operator $P_{\alpha_0}$ map onto, and $\forall\, x \in X$*

$$P_\alpha x \to P_{\alpha_0} x \qquad \text{as} \qquad \alpha \to \alpha_0\ . \tag{23}$$

*Then there exist a neighborhood $\mathcal{O}(\alpha_0)$ and a constant $c > 0$, such that $\forall\, \alpha \in \mathcal{O}(\alpha_0)$ the operator $P_\alpha$ covers with rate $c$.*

*Proof.* Since $P_{\alpha_0}$ is onto, there exists a subspace $L \subset X$ of dimension $q$, such that $P_{\alpha_0} L = \mathbb{R}^q$, and moreover, the restricted operator $\widetilde{P}_{\alpha_0} :\ L \to \mathbb{R}^q$ covers with a rate $c > 0$. From (23) it follows that then the restricted operators $\widetilde{P}_\alpha :\ L \to \mathbb{R}^q$ converge to $\widetilde{P}_{\alpha_0}$ in the operator norm, as $\alpha \to \alpha_0$. Then, by Theorem 2.2 $\forall\, c' < c\ \ \exists\, \mathcal{O}(\alpha_0)$ (or by Theorem 2.4 $\exists\, c' > 0$ and $\mathcal{O}(\alpha_0)$) such that $\forall\, \alpha \in \mathcal{O}(\alpha_0)$ the operator $\widetilde{P}_\alpha$ covers with rate $c'$. The more so, then $P_\alpha$ covers with this rate $c'$. ∎

Thus, if an operator depending on a parameter $\alpha$ maps into a finite-dimensional space, then, to obtain its uniform covering for $\alpha$ from some neighborhood $\mathcal{O}(\alpha_0)$ one does not need to require its continuity in the operator norm, it is sufficient to have its convergence for each element of the space $X$.

In particular, functionals (22) with $\alpha \in L_\infty[0,T]$ obviously satisfy condition (23) w.r.t. $\sigma^*$-convergence.

## 4. Covering of a combined operator

Consider now the following important case of a combined operator. Let $D: W \to Y$ and $P: W \to Z$ be linear operators between Banach spaces $W$ and $Y$, $Z$. Define a combined operator

$$G = (D,\, P):\; W \to Y \times Z, \qquad w \mapsto (Dw,\, Pw).$$

The following lemma is useful for such operators.

LEMMA 4.1 *Suppose that $D$ covers with a rate $a > 0$, the restriction of $P$ to $L = ker\, D$ covers with a rate $b > 0$, and $||P|| \le \mu$. Then the combined operator $G = (D, P)$ covers with rate*

$$c = c(a, b, \mu) = \left( \max \left\{ \frac{1}{a} \left( 1 + \frac{\mu}{b} \right),\; \frac{1}{b} \right\} \right)^{-1}. \tag{24}$$

*The reverse assertion holds in the following form: If $G$ covers with a rate $c > 0$, then $D$ and $P|_L$ cover with at least the same rate $c$, independently of $||P||$.*

In proving this it is more convenient to deal with the inverse constants. If a linear operator covers with rate $a > 0$, we will say that it is regular with constant $A = 1/a$. (Regularity with constant $k$, or $k$-regularity, means here that any element in the image space with norm 1 has a preimage with norm $\le k$. This is almost the same as the usual metric regularity.) Then the lemma asserts that if $D$ is regular with constant $A$, the restriction $P|_L$ is regular with constant $B$, and $||P|| \le \mu$, then $G$ is regular with constant

$$C = max\, \{ A\,(1 + B\mu),\; B \}\,. \tag{25}$$

*Proof.* Let us take any element $(y, z) \in Y \times Z$, $\;||y|| + ||z|| \le 1$, and show that it has a preimage with a suitable norm.

By the $A$-regularity of operator $D$, there is $w' \in W$ such that $Dw' = y$, $||w'|| \le A||y||$. Moreover, $Pw' = z'$, where $||z'|| \le ||P|| \cdot ||w'|| \le \mu A\, ||y||$.

However, we must obtain the equality $Pw = z$. Set $\bar z = z - z'$. By the $B$-regularity of $P$ on $ker\, D$, there is $\bar w \in ker\, D$ such that $P\bar w = \bar z$, $\;||\bar w|| \le B\, ||\bar z||$. Then $w = w' + \bar w$ satisfies both the required equalities:

$$Dw = Dw' + D\bar w = y + 0 = y,$$
$$Pw = Pw' + P\bar w = z' + \bar z = z.$$

It remains to estimate $||w||$. Since

$$||\bar w|| \le B\, ||\bar z|| \le B\, (||z|| + ||z'||) \le B\, (||z|| + \mu A\, ||y||),$$

we get

$$||w|| \le ||w'|| + ||\bar w|| \le A\, ||y|| + B\, (||z|| + \mu A\, ||y||) = A\, (1 + B\mu) ||y|| + B\, ||z||.$$

The maximum of the obtained expression over the set $||y|| + ||z|| \leq 1$ is equal to the value (25), Q.E.D. ■

Again, what is important here is that the constant $c$ depends only on the constants $a$, $b$, $\mu$, but not on the operators $P$, $D$ themselves. This makes it possible to apply this lemma for obtaining a uniform covering for a family of operators.

LEMMA 4.2 *If a family of linear operators $D_\alpha : W \to Y$ covers with a common rate $a > 0$, and a family of operators $P_\alpha : W \to Z$ is uniformly bounded, $||P_\alpha|| \leq \mu$, and $\forall\, \alpha$ the restriction of $P_\alpha$ to $L_\alpha = ker\, D_\alpha$ covers with a common rate $b > 0$, then the combined operator*

$$G_\alpha = (D_\alpha,\, P_\alpha): \quad W \to Y \times Z, \qquad w \mapsto (D_\alpha w,\ P_\alpha w). \tag{26}$$

*covers with the common rate (24).*

Consider now the following question. Let be given linear operators $P_\alpha : W \to Z$, where $\alpha$ runs through a topological space $\mathcal{A}$, and let $L_\alpha$ be a family of subspaces in $W$. How to obtain the uniform covering of a family of restricted operators $P_\alpha : L_\alpha \to Z$? Assume that a point $\alpha_0 \in \mathcal{A}$ is given.

LEMMA 4.3 *Let $H$ be a Banach space, on which are given linear operators $\Phi_\alpha : H \to W$, so that we have $H \xrightarrow{\Phi_\alpha} L_\alpha \xrightarrow{P_\alpha} Z$, with the following properties:*

*a)* $||\Phi_\alpha|| \leq R \qquad$ *for some $R$;*

*b)* $\forall\, \alpha \in \mathcal{A}, \qquad \Phi_\alpha H \subset L_\alpha$*;*

*c)* $P_{\alpha_0}\Phi_{\alpha_0}$ *maps $H$ onto $Z$;*

*d)* $||P_\alpha \Phi_\alpha - P_{\alpha_0}\Phi_{\alpha_0}|| \to 0 \qquad$ *as* $\alpha \to \alpha_0$.

*Then there exist a neighborhood $\mathcal{O}(\alpha_0)$ and a constant $b > 0$, such that $\forall\, \alpha \in \mathcal{O}(\alpha_0)$ the operator $P_\alpha$ covers on $L_\alpha$ with rate $b$.*

It is easy to see that when a) holds and $||P_\alpha - P_{\alpha_0}|| \to 0$, the condition d) is equivalent to the condition

d') $||P_{\alpha_0}(\Phi_\alpha - \Phi_{\alpha_0})|| \to 0 \qquad$ as $\alpha \to \alpha_0$.

*Proof.* From conditions c) and d) by Theorem 2.4 the existence of a neighborhood $\mathcal{O}(\alpha_0)$ follows, in which all $P_\alpha\Phi_\alpha$ cover with a common rate $c > 0$, i.e., $P_\alpha\Phi_\alpha B_1 \supset B_c$. In view of a), $\Phi_\alpha B_1 \subset B_R \cap L_\alpha$, therefore $P_\alpha(B_R \cap L_\alpha) \supset B_c$, so $P_\alpha$ covers with rate $c/R$. ■

In the case when the space $Z$ is finite-dimensional, the conditions of Lemma 4.3 can be simplified. As before, let be given operators $P_\alpha : W \to Z$ and subspaces $L_\alpha \subset W$.

LEMMA 4.4 *Suppose that $Z = \mathbb{R}^q$,*

*e)* $||P_\alpha - P_{\alpha_0}|| \to 0 \qquad$ *as $\alpha \to \alpha_0$;*

*f)* $P_{\alpha_0}$ *maps* $L_{\alpha_0}$ *onto* $Z$ ; *and, moreover,*

*g)* $\forall\, z \in Z$ *there exists a mapping* $\alpha \mapsto w_\alpha \in L_\alpha$ *such that* $P_{\alpha_0} w_{\alpha_0} = z$, $||w_\alpha|| \le$ *const*, *and*

$$P_\alpha w_\alpha \to P_{\alpha_0} w_{\alpha_0} \qquad as \qquad \alpha \to \alpha_0\ . \tag{27}$$

*Then the assertion of Lemma 4.3 holds.*

Assumption g) can be replaced by the following stronger assumption:

g') $\forall\, w_{\alpha_0} \in L_{\alpha_0}$ there exists a mapping $\alpha \mapsto w_\alpha \in L_\alpha$ such that $||w_\alpha|| \le \text{const}$, and $P_\alpha w_\alpha \to P_{\alpha_0} w_{\alpha_0}$ as $\alpha \to \alpha_0$.

*Proof.* Let us construct a mapping $\Phi_\alpha$ satisfying properties a)–d). In view of e) we replace d) by d'), and so, the operators $P_\alpha : W \to Z$ can be assumed not depending on $\alpha$.

For the given $\alpha_0$ choose a finite set $w_{\alpha_0,1}, \ldots, w_{\alpha_0,q} \in L_{\alpha_0}$, such that their images

$$\{\, P\, w_{\alpha_0, j} \;=\; z_{\alpha_0, j} \,\} \qquad \text{form a basis in } \mathbb{R}^q. \tag{28}$$

This is possible in view of assumption f).

By condition g), $\forall\, j$ there exists a mapping $\alpha \mapsto w_{\alpha,j} \in L_\alpha$ such that

$$||w_{\alpha,j}|| \le \text{const}, \qquad \text{and} \tag{29}$$

$$P\, w_{\alpha,j} \to P\, w_{\alpha_0,j} \qquad \text{as } \alpha \to \alpha_0\,. \tag{30}$$

Take the space $H = \mathbb{R}^q$ with elements $h = (h_1, \ldots, h_q)$, and $\forall\, \alpha \in \mathcal{A}$ define the mapping

$$\Phi_\alpha :\, H \to W, \qquad \Phi_\alpha(h) = \sum_j h_j\, w_{\alpha,j}\ .$$

Now, let us check conditions a)–d) of Lemma 4.3. In view of (29), we have $||\Phi_a|| \le$ *const* too, so condition a) is fulfilled. Condition b) is fulfilled by construction. Due to (28), condition c) is fulfilled too. It remains to check condition d').

Since the operator $P\Phi_a$ maps from $H = \mathbb{R}^q$, it suffices to check d') on any basis vector $e_j = (0, \ldots, 0, 1, 0, \ldots, 0) \in H$, i.e., we must check that $\forall\, j$

$$(P\Phi_a - P\Phi_{\alpha_0})\, e_j \to 0.$$

But this means that $P\, w_{\alpha,j} - P\, w_{\alpha_0,j} \to 0$, which is really true due to (30), Q.E.D. ■

Lemmas 4.2– 4.4, yield the following final results. Let, as before, $\forall\, \alpha \in \mathcal{A}$ be given an operator $G_\alpha$ of the form (26).

THEOREM 4.1 *Suppose that all operators* $D_\alpha : W \to Y$, $\alpha \in \mathcal{A}$, *cover with a common rate* $a > 0$, *all* $P_\alpha : W \to Z$ *are uniformly bounded:* $||P_\alpha|| \le \mu$, *there is an auxiliary Banach space* $H$, *on which there are given auxiliary operators* $\Phi_\alpha : H \to W$ *with the properties:* $Im\, \Phi_\alpha \subset ker\, D_\alpha$, $P_{\alpha_0} \Phi_{\alpha_0}$ *is onto, and* $||P_\alpha \Phi_\alpha - P_{\alpha_0} \Phi_{\alpha_0}|| \to 0$ *as* $\alpha \to \alpha_0$. *Then there exist a neighborhood* $\mathcal{O}(\alpha_0)$ *and a constant* $c > 0$, *such that* $\forall\, \alpha \in \mathcal{O}(\alpha_0)$ *the combined operator* $G_\alpha$ *covers with rate* $c$.

In case of $Z = \mathbb{R}^q$, we come to the following

THEOREM 4.2 *Suppose that all operators* $D_\alpha : W \to Y$, $\alpha \in \mathcal{A}$, *cover with a common rate* $a > 0$,

*e)* $||P_\alpha - P_{\alpha_0}|| \to 0$ *as* $\alpha \to \alpha_0$*;*

*f)* $P_{\alpha_0}$ *maps* $ker\, D_{\alpha_0}$ *onto* $Z$*; and, moreover,*

*g)* $\forall\, z \in Z$ *there exists a bounded mapping* $\alpha \mapsto w_\alpha \in ker\, D_\alpha$ *such that* $P_\alpha w_\alpha \to P_{\alpha_0} w_{\alpha_0} = z$ *as* $\alpha \to \alpha_0$.
*Then the assertion of Theorem 4.1 holds.*

## 5. Application

Theorem 4.2 can be applied to the following specific operator, arising in the study of control systems.

Let $W = AC^n \times L^s_\infty[0,T]$, $Y = L^n_1[0,T]$, $Z = \mathbb{R}^q$, $\mathcal{A}$ be a bounded set in $L^N_\infty[0,T]$ equipped with the weak-* topology, and $\forall\, \alpha \in \mathcal{A}$ let be given a linear operator (here we write the parameter $\alpha$ in the brackets)

$$G[\alpha] : \; AC^n \times L^s_\infty \to L^n_1 \times \mathbb{R}^q, \qquad \bar{w} = (\bar{x}, \bar{u}) \mapsto (\bar{y},\, \bar{z}),$$

$$\bar{y} = \; D[\alpha](\bar{x}, \bar{u}) = \; \dot{\bar{x}} - (\sum_1^N \alpha^i\, \Gamma^i)\, \bar{x} - (\sum_1^N \alpha^i \Lambda^i)\, \bar{u}, \tag{31}$$

$$\bar{z} = \; P(\bar{x}, \bar{u}) = \; K_0\, \bar{x}(0) \; + \; K_T\, \bar{x}(T), \tag{32}$$

where $dim\, \bar{x} = n$, $dim\, \bar{u} = s$, and $\Gamma^i(t)$, $\Lambda^i(t)$ are measurable bounded matrices of corresponding dimensions, $K_0$, $K_T$ are $n \times n$-matrices. The operator $P$ does not depend on the functional parameter $\alpha$.

THEOREM 5.1 *Suppose that* $G[\alpha_0]$ *covers. Then there exist a weak-* neighborhood* $\mathcal{V}(\alpha_0)$ *and a constant* $c > 0$, *such that* $\forall\, \alpha \in \mathcal{V}(\alpha_0) \cap \mathcal{A}$ *the operator* $G[\alpha]$ *covers with rate* $c$.

*Proof.* Since here $Z = \mathbb{R}^q$, it suffices to check the assumptions of Theorem 4.2. Let us first show that our family of operators $D[\alpha] : W \to Y$ covers with a common rate $a > 0$. Indeed, for any $\alpha \in \mathcal{A}$ and any $\bar{y} \in L_1$ consider the solution $\bar{x}(t)$ to equation (31) with $\bar{u}(t) = 0$ and $\bar{x}(0) = 0$. Then $||\bar{x}||_\infty \le \mu_0\, ||\bar{y}||_1$, where $\mu_0$ depends only on $||\alpha||_1$ (and, naturally, on $||\gamma^i||_\infty$, $||\Lambda^i||_\infty$), and hence, from the same equation (31) we obtain the estimate $||\dot{\bar{x}}||_1 \le \mu_1\, ||\bar{y}||_1$, which implies $||\bar{x}||_{AC} \le \mu\, ||\bar{y}||_1$, and therefore $D[\alpha]$ covers with rate $a = 1/\mu > 0$.

Now, let us check the other assumptions. Assumption e) is fulfilled trivially, since our $P$ does not depend on $\alpha$. Next, since $G[\alpha_0]$ covers, then $P : ker\, D[\alpha_0] \to Z$ covers with some $a_0 > 0$, so f) is fulfilled too.

Finally, let us check assumption g). Take an arbitrary $\bar{w}_{\alpha_0} = (\bar{x}_{\alpha_0}, \bar{u}_{\alpha_0}) \in ker\, D[\alpha_0]$. For any $\alpha \in \mathcal{A}$ construct $\bar{w}_\alpha = (\bar{x}_\alpha, \bar{u}_\alpha) \in ker\, D[\alpha]$ as follows. Set $\bar{u}_\alpha(t) = \bar{u}_{\alpha_0}(t)$, and let $\bar{x}_\alpha(t)$ be the solution to equation

$$\dot{\bar{x}} - (\sum \alpha^i \Gamma^i)\, \bar{x} - (\sum \alpha^i \Lambda^i)\, \bar{u}_{\alpha_0} = 0$$

with the initial condition $\bar{x}(0) = \bar{x}_{\alpha_0}(0)$. As is well known, since $\alpha$ comes linearly in this equation, we have

$$||\bar{x}_a - \bar{x}_{\alpha_0}||_\infty \to 0 \qquad \text{as } \alpha \xrightarrow{\sigma *} \alpha_0\,.$$

Then, in particular, $\bar{x}_a(T) \to \bar{x}_{\alpha_0}(T)$, and hence $P\,\bar{w}_a \to P\,\bar{w}_{\alpha_0}$.

Thus, all the assumptions of Theorem 4.2 are fulfilled, and so, the required $\mathcal{V}(\alpha_0)$ and $c > 0$ do exist. ■

**Remarks.** 1) In Dmitruk (2002) and Milyutin, Dmitruk and Osmolovskii (2004) this theorem was proved by using the specificity of the operator $G[\alpha]$. 2) The assertion of Theorem 5.1 remains valid if we replace the set $\mathcal{A} \subset L_\infty$ by a broader set $\mathcal{A} \subset L_1$, and replace the weak-* topology in $L_\infty$ (with respect to $L_1$) by a weaker one — the weak topology in $L_1$ (with respect to $L_\infty$ only). However, in the applications we met only the case $\alpha \in L_\infty$, but have not yet met the more general case $\alpha \in L_1$.

Theorem 5.1 allows to obtain a nonlocal covering of the following nonlinear operator. Let now $W = AC^n \times L^s_\infty \times L^N_\infty$ with elements $(x, u, \alpha)$, where $\alpha = (\alpha^1, \dots, \alpha^N)$. As before, let $Y = L^n_1$, $Z = \mathbb{R}^q$. Consider the operator

$$F:\ W \to Y \times Z, \qquad F(x,u,\alpha) = (y,z), \tag{33}$$
$$\dot{x} - f^0(x,u,t) - \sum_{i=1}^N \alpha^i(t)\, f^i(x,u,t) = y, \qquad K(x(0), x(T)) = z,$$

where all $f^i$ and $K$ are smooth functions of their arguments. Obviously, $F$ has the Fréchet derivative

$$F'(x,u,\alpha) = G[x,u,\alpha]:\ W \longrightarrow Y \times Z$$

acting as follows: $(\bar{x}, \bar{u}, \bar{\alpha}) \mapsto (\bar{y}, \bar{z})$, where

$$\left.\begin{aligned} \dot{\bar{x}} - \big(f^0_x(x,u,t) + \textstyle\sum \alpha^i f'_x(x,u,t)\big)\, \bar{x} - \\ - \big(f^0_u(x,u,t) + \textstyle\sum \alpha^i f^i_u(x,u,t))\big)\, \bar{u}^i - \textstyle\sum \bar{\alpha}^i f^i(x,u,t) \;=\; \bar{y} \in L_1\,, \\ K'_{x(0)}(x_0, x_T)\, \bar{x}(0) \;+\; K'_{x(T)}(x_0, x_T)\, \bar{x}(T) \;=\; \bar{z} \in \mathbb{R}^q. \end{aligned}\right\} \tag{34}$$

This derivative continuously depends on $(x, u, \alpha)$. Moreover, it is Lipschitz continuous with respect to variations of $(x, u)$ in the norm $||x||_\infty + ||u||_\infty$ uniformly over $\alpha$ from any bounded set $\mathcal{A} \subset L_\infty^N$. The following theorem holds for the family of operators $G[x, u, \alpha]$.

Theorem 5.2 *Suppose that a triple $w_0 = (x_0, u_0, \alpha_0)$ with $\alpha_0 \in \mathcal{A}$ is such that the linear operator $G[w_0]$ is onto. Then there exist a weak-* neighborhood $\mathcal{V}(\alpha_0)$ and numbers $c > 0$, $\varepsilon > 0$ such that for any triple $(x, u, \alpha) \in W$ satisfying the conditions*

$$||x - x_0||_\infty < \varepsilon, \qquad ||u - u_0||_\infty < \varepsilon, \qquad \alpha \in \mathcal{V}(\alpha_0) \cap \mathcal{A}, \tag{35}$$

*the linear operator $G[x, u, \alpha]$ covers with rate $c$.*

*Proof.* First, we fix $(x_0\,, u_0)$ and consider the family of operators $\widetilde{G}[\alpha] = G[x_0, u_0, \alpha]$, $\alpha \in \mathcal{A}$. Let us show that this family satisfies the conditions of Theorem 5.1. The arguments $\bar{\alpha}^i$ of $\widetilde{G}$ should be regarded here as additional control variations (and could be denoted as $\bar{u}^{s+i}$). Then the operator $\widetilde{G}[\alpha]$ has the above form (31), (32). Since $\widetilde{G}[\alpha_0]$ is onto, all the conditions of Theorem 5.1 are fulfilled. By this theorem, there exists a weak-* neighborhood $\mathcal{V}(\alpha_0)$ and a constant $c > 0$, such that $\forall\, \alpha \in \mathcal{V}(\alpha_0) \cap \mathcal{A}$ the operator $\widetilde{G}[\alpha] = G[x_0, u_0, \alpha]$ covers with rate $c$. But then, since the functions $f^i$, $f_x^i$, $f_u^i$ and $K'$ are continuous in $(x, u)$, and the set $\mathcal{A} \subset L_\infty$ is bounded, then $\forall\, \delta > 0$ we have $||G[x, u, \alpha] - G[x_0, u_0, \alpha]|| < \delta$ uniformly for all $\alpha \in \mathcal{A}$, provided that $(x(t), u(t))$ are uniformly close enough to $(x_0(t), u_0(t))$. Therefore, by Theorem 2.2 $\forall\, c' < c \;\; \exists\, \varepsilon > 0$ such that for any triple $(x, u, \alpha)$ satisfying (35) the operator $G[x, u, \alpha]$ covers with rate $c'$. ■

From Theorems 5.2 and 3.2 we readily get the following

Theorem 5.3 *Suppose that $\mathcal{A}$ is bounded and open, and $F'[w_0]$ is onto. Then there exist a weak-* neighborhood $\mathcal{V}(\alpha_0)$ and numbers $c > 0$, $\varepsilon > 0$ such that the nonlinear operator $F$ covers with rate $c$ on the set (35).*

Finally, this theorem and Theorem 2.5 yield the following nonlocal distance estimate to the level set $\;M = \{\, w \mid \; F(w) = 0\,\}$.

Theorem 5.4 *Let $\mathcal{A} \subset L_\infty^N$ be bounded, $\alpha_0 \in \mathcal{A}$, and the triple $w_0 = (x_0, u_0, \alpha_0) \in M$ be such that the derivative $F'[w_0]$ is onto. Then there exists a weak-* neighborhood $\mathcal{V}(\alpha_0)$ and numbers $\varepsilon > 0$, $L$ such that for any triple $w = (x, u, \alpha) \in W$ satisfying conditions (35), we have*

$$dist\,(w,\, M) \;\le\; L\, ||F(w)||\,. \tag{36}$$

*Proof.* Let $\widetilde{\mathcal{A}}$ be the 1-neighborhood of $\mathcal{A}$ in $L_\infty^N$. By Theorem 5.3 $\exists\, c,\, \varepsilon > 0$ and a weak-* neighborhood of zero $\mathcal{V}(0) \subset L_\infty^N$ such that $F$ covers with rate $c$

on the set

$$||x - x_0||_\infty < 2\varepsilon, \qquad ||u - u_0||_\infty < 2\varepsilon, \qquad \alpha \in (\alpha_0 + 2\mathcal{V}(0)) \cap \widetilde{\mathcal{A}}.$$

Denote this set by $\mathcal{G}$. Reduce, if necessary, $\varepsilon > 0$ so that $\varepsilon < 1$ and $B_\varepsilon(0) \subset \mathcal{V}(0)$, and define the set $\Omega$ by conditions (35) with $\mathcal{V}(\alpha_0) = \alpha_0 + \mathcal{V}(0)$. Then the $\varepsilon$-neighborhood of $\Omega$ is contained in $\mathcal{G}$, and hence, by Theorem 2.5, for some $L$ the estimate (36) holds on $\Omega$. ■

## 6. A relaxation theorem

The obtained nonlocal estimate is essentially used in the proof of the following relaxation (or approximation) theorem.

Consider the following control system on a fixed time interval $[0, T]$:

$$\dot{x} = f(x, u, t), \qquad K(x(0), x(T)) = 0, \tag{37}$$

where $x \in AC^n[0, T]$, $u \in L_\infty^r[0, T]$, $dim\, K = q$, and the functions $f$, $K$ are assumed smooth.

Along with system (37), consider also an extended (relaxed) system, obtained by the convexification of its velocity set:

$$\dot{x} = \sum_{i=0}^{N} \alpha^i(t)\, f(x, u^i, t), \qquad K(x(0), x(T)) = 0, \tag{38}$$

where $i = 0, 1, \ldots, N$, all $u^i \in L_\infty^r$, all $\alpha^i(t) \geq 0$, and $\sum \alpha^i(t) = 1$. So, the vector function $\bar{\alpha}(t) = (\alpha^0, \ldots, \alpha^N) \in L_\infty^{N+1}$ takes its values in the simplex

$$A = \{\, \bar{\alpha} \in \mathbb{R}^{N+1} \mid \forall i \quad \alpha^i \geq 0, \quad \sum_{i=0}^{N} \alpha^i = 1 \,\}.$$

Denote, as usual, by $ex\, A$ the set of vertices of $A$; so, $\bar{\alpha} \in ex\, A$ means that $\bar{\alpha}$ is a basis vector $e^i$ for some $i$.

Obviously, the set of solutions to (38) is wider than that to (37) in the sense that any solution $x(t), u(t)$ to (37) can be also considered as a solution to (38) for the extended collection of controls $u^0(t) = u(t)$, arbitrary $u^1(t), \ldots, u^N(t)$, and weight coefficients $\bar{\alpha}(t) = (1, 0, \ldots, 0)$. However, the reverse inclusion is not so obvious, and generally, it is not true.

So, the question is, when the passage to the extended system is valid, i.e., when a given trajectory of system (38) can be approximated by trajectories of the initial system (37)?

The first idea is, given a solution to system (38), fix the controls $u^0(t)$, $u^1(t), \ldots, u^N(t)$, and consider a sequence of weight coefficients $\bar{\alpha}_m(t) \in ex\, A$, such that $\bar{\alpha}_m(t) \xrightarrow{\sigma *} \bar{\alpha}(t)$ (which means that $\alpha^i_m(t) \xrightarrow{\sigma *} \alpha^i(t) \;\; \forall\, i = 0, 1, \ldots, N$). Then, setting $x_m(0) = x(0)$, we get $x_m(t) \longrightarrow x(t)$ uniformly on $[0,T]$. If the function $K$ does not depend on $x(T)$ (i.e., $x(T)$ is free), then, setting $u_m(t) = \sum_{i=0}^N \alpha^i_m(t)\, u^i_m(t)$, we get a sequence of pairs $(x_m, u_m)$ satisfying system (38). (In engineering applications such a sequence is called a sliding mode regime.) This is essentially the approach proposed a long time ago by N.N. Bogolyubov (see Ioffe and Tikhomirov, 1979), L.C. Young (1969) and E.J. McShane (1967); after the paper of R.V. Gamkrelidze (1962) it became a standard tool in control theory and was exploited by many authors (to mention just a few of them, see e.g. Warga, 1972; Olech, 1976; Balder, 1984; Artstein, 1989; Rosenblueth and Vinter 1991; Roubicek, 1997; Tolstonogov, 2000).

But in our case, when $K$ depends on both endpoints of the trajectory, we have an obstacle: generally, $K(x_m(0), x_m(T)) \neq 0$, so the pair $(x_m, u_m)$ does not satisfy system (38)!

In order to overcome this obstacle, we proceed as follows. First, taking into account that $\alpha^0 = 1 - \sum_{i=1}^N \alpha^i$, it is convenient to rewrite the differential equation in (38) in terms of "independent" coefficients $\alpha^1, \ldots, \alpha^N$:

$$\dot{x} = \; f(x, u^0, t) \; + \; \sum_{i=1}^{n} \alpha^i \; \bigl( f(x, u^i, t) - f(x, u^0, t) \bigr) \, .$$

Then we define the Banach space $W = AC^n \times (L^r_\infty)^{N+1} \times L^N_\infty$ with elements $(x, u, \alpha)$:

$$x \in AC^n, \qquad u = (u^0, u^1, \ldots, u^N) \in (L^r_\infty)^{N+1}, \qquad \alpha = (\alpha^1, \ldots, \alpha^N) \in L^N_\infty,$$

define the Banach spaces $Y = L^n_1$, $Z = \mathbb{R}^q$, and consider the operator

$$F: \; W \to Y \times Z, \qquad F(x, u, \alpha) = (y, z),$$

where

$$\left. \begin{array}{c} \dot{x} - f(x, u^0, t) - \displaystyle\sum_{i=1}^{n} \alpha^i \; \bigl( f(x, u^i, t) - f(x, u^0, t) \bigr) = y, \\ K\,(x(0),\, x(T)) \; = \; z, \end{array} \right\} \tag{39}$$

so, the solution to system (38) is the zero set of operator $F$. This operator has the above form (33) with $s = r(N+1)$ and $f^i(x, u, t) = f(x, u^i, t) - f(x, u^0, t)$. The set $\mathcal{A}$ now consists of all $\alpha = (\alpha^1, \ldots, \alpha^N) \in L^N_\infty$ satisfying a.e. on $[0,T]$ the conditions:

$$\alpha^i(t) \geq 0 \quad \forall\, i, \qquad\qquad \sum_{i=1}^N \alpha^i(t) \leq 1.$$

(Note that if we leave all $\alpha^i$, $\; i = 0, 1, \ldots, N$, then we should extend the operator $F$ by adding the third component: $\sum_{i=0}^N \alpha^i(t) - 1 = y^0$.)

In the space $W$, along with its natural topology, generated by the norm $||w|| = ||x||_{AC} + ||u||_\infty + ||\alpha||_\infty$, we also define the $(C, L_\infty, \sigma^*)$-topology, which is the product of $C$-norm topology for $x$, $L_\infty$-norm topology for $u$, and weak-* topology for $\alpha$.

As before, we consider the level set $M = \{\, w \mid\ F(w) = 0\,\}$.

THEOREM 6.1 *Let a triple $(x_0, u_0, \alpha_0) \in M$ satisfy the following two conditions:*

*a) $F'(x_0, u_0, \alpha_0)$ is onto (the Lyusternik condition);*

*b) $\alpha_0 \in int\,\mathcal{A}$, i.e., $\forall\, i \quad \alpha_0^i(t) \geq const > 0$,*
*and $1 - \sum_{i=1}^N \alpha_0^i(t) \geq const > 0$ a.e. on $[0,T]$.*

*Then, in any $(C, L_\infty, \sigma^*)$-neighborhood of the triple $(x_0, u_0, \alpha_0)$ there exists a triple $(x, u, \alpha)$ still belonging to $M$ and such that $\forall\, i$ the function $\alpha^i(t)$ takes only two values: 0 or 1.*

In other words: there exists a sequence $(x_m, u_m, \alpha_m)$ such that

$$||x_m - x_0||_C \to 0, \qquad ||u_m - u_0||_\infty \to 0, \qquad (\alpha_m - \alpha_0) \xrightarrow{\sigma*} 0,$$

and $\forall\, m$ all $\alpha_m^i(t) =$ 0 or 1 almost everywhere.

Assumption a) means in other words that the linearization of system (39), which has the form (34), is controllable in the sense that each pair $\bar{y}, \bar{z}$ in its right hand side is attainable, or equivalently, that attainable are $\bar{y} = 0$ and each $\bar{z} \in \mathbb{R}^q$.

This theorem enables one to approximate a trajectory of the relaxed system (38) involving sliding mode controls by trajectories of the initial system (37) involving just ordinary controls, in the following sense: constructing, as shown above, the controls

$$\tilde{u}_m(t) = \sum_{i=0}^N \alpha_m^i(t)\, u_m^i(t), \qquad \text{where} \qquad \alpha_m^0(t) = 1 - \sum_{i=1}^N \alpha_m^i(t)$$

(so, $\bar{\alpha}_m(t) = (\bar{\alpha}_m^0(t),\ \ldots,\ \bar{\alpha}_m^N(t)) \in\ ex\,A$), we obtain a sequence $(x_m(t), \tilde{u}_m(t))$ satisfying system (37).

The proof is based on a specific iteration process of corrections which allow for satisfying the equality $K = 0$ (see Dmitruk, 1976, Sec. 5-7; Milyutin, Dmitruk and Osmolovskii, 2004, Sec. 5.4.)

Theorem 6.1 can be used in a proof of the Maximum Principle for the general optimal control problem with state and regular mixed constraints by passing to a relaxed system with sliding mode controls. The author learned this idea from A.Ja. Dubovitskii and A.A. Milyutin, and realized it in Dmitruk (1993) and Milyutin, Dmitruk and Osmolovskii (2004). A similar theorem was proved by S.V. Chukanov (1990) for control systems with integral equations; he also applied it to obtain the corresponding Maximum Principle.

Note in conclusion that this paper should be considered just as one among few steps in the study and usage of the nonlocal Lyusternik estimates. To our opinion, it would be interesting to find a possibility of having such estimates for other classes of operators, e.g., for PDE operators.

**Acknowledgments**

This work was supported by the Russian Foundation for Basic Research, project 04-01-00482, and by the Government Program for Leading Scientific Schools, project NSh-304.2003.1. The author thanks anonymous referees for useful remarks and suggestions.

## References

Artstein, Z. (1989) Rapid oscillations, chattering systems, and relaxed controls. *SIAM J. on Control and Opt.* **27** (5), 940–948.

Balder, E.J. (1984) A general denseness result for relaxed control theory. *Bull. Aust. Math. Soc.* **30**, 463–475.

Borwein, J.M. and Zhuang, D.M. (1988) Verifiable necessary and sufficient conditions for openness and regularity of set-valued and single-valued maps. *J. Math. Anal. Appl.* **134** (2), 441–459.

Bulgakov, A.I. and Vasilyev, V.V. (2002) On the theory of functional-differential inclusions of neutral type. *Georgian Math. J.* **9** (1), 33–52.

Chukanov, S.V. (1990) Maximum principle for optimal control problems with integral equations. In: A.A. Milyutin, ed., *Necessary Condition in Optimal Control.* Nauka, Ch. 6.

Dmitruk, A.V. (1976) The justification of the sliding mode method to optimal control problems with mixed constraints. *Functional Analysis and its Appl.* **10**, 197–201.

Dmitruk, A.V. (1993) Maximum principle for a general optimal control problem with state and regular mixed constraints. *Computational Mathematics and Modeling* **4** (4), 364–377.

Dmitruk, A.V. (2002) A nonlocal Lyusternik estimate and its application to control systems with sliding modes. In: A.B. Kurzhanski and A.L. Fradkov, eds., *Nonlinear Control Systems 2001*, **2**, 1061–1064, Elsevier.

Dmitruk, A.V., Milyutin, A.A. and Osmolovskii, N.P. (1980) Luysternik's theorem and the theory of extrema. *Russian Math. Surveys* **35** (6), 11–51.

Dontchev, A.L. and Rockafellar, R.T. (2004) Regularity and conditioning of solution mappings in variational analysis. *Set-Valued Analysis* **12** (1), 79–109.

Gamkrelidze, R.V. (1962) Optimal sliding states. *Soviet Math. Dokl.* **3**, 559–562.

Ioffe, A.D. (2000) Metric regularity and subdiffertential calculus. *Russian Math. Surveys* **55** (3), 501–558.

Ioffe, A.D. (2001) On perturbation stability of metric regularity. *Set-Valued Analysis* **9** (1-2), 101–109.

Ioffe, A.D. and Tikhomirov, V.M. (1974) *Theory of Extremal Problems*. M., Nauka, 1974; English translation: Amsterdam, North-Holland, 1979.

Lyusternik, L.A. (1934) On the conditional extrema of functionals. *Mat. Sbornik* **41**, 390–401 (in Russian).

McShane, E.J. (1967) Relaxed controls and variational problems. *SIAM J. on Control* **5**, 438–485.

Milyutin, A.A., Dmitruk, A.V. and Osmolovskii, N.P. (2004) *The Maximum Principle in Optimal Control*. Mech.-Math. Faculty of Moscow State University, (in Russian).

Olech, C. (1976) Existence theory in optimal control. In: *Control theory and topics in functional analysis*, I, Vienna, 291–328.

Penot, J.-P. (1989) Metric regularity, openness and Lipschitzian behavior of multifunctions. *Nonlinear Analysis, TMA* **13** (6), 629–643.

Rosenblueth, J.F. and Vinter, R.B. (1991) Relaxation procedures for time delay systems. *J. Math. Anal. Appl.* **162** (2), 542–563.

Roubicek, T. (1997) *Relaxation in Optimization Theory and Variational Calculus*. de Gruyter, Berlin.

Tolstonogov, A.A. (2000) *Differential Inclusions in a Banach Space*. Dordrecht, Kluwer Academic Publishers.

Warga, J. (1972) *Optimal Control of Differential and Functional Equations*. New York, Academic Press.

Young, L.C. (1969) *Lectures on the Calculus of Variations and Optimal Control Theory*. Saunders, Philadelphia.