Revised version, Dec. 198

PROXIMITY-SPACE METHODS IN OPTIMIZATION WITH CONSTRAINTS

Tomáš Roubíček

Institute of Information Theory and Automation, Czechoslovak Academy of Sciences, Pod vodárenskou věží 4, 182 08 Praha 8, Czechoslovakia.

0. Introduction

The theory presented below is an attempt to overcome by proper definitions some problems in optimization theory such as existence or stability of the (classical) solutions. This problems are felt rather artificial from the viewpoint of practically oriented engineers, economists, etc. Indeed, in practice it is entirely sufficient to find an "almost" feasible solution (say up to a small $\varepsilon > 0$) which is, at the same time, "almost" optimal (again up to $\varepsilon > 0$). Then the question whether there is an exact minimizer is not so important. As to possible unstability of a solution, it means simply that, roughly speaking, the solution varies considerably when data are changed only a little, or also that around the solution there are points enough at which the mappings in question have "nearly" the same values as at the solution, hence the "quality" of the solution may not be much worsened when it moves. Therefore the unstability of the solution is to be understood in fact as a good property, which is due to the general character of optimization problems as inverse ones.

Our approach will reflect these engineering feelings and, at the same time, will be exact to satisfy mathematicians. It should be emphasized, however, that this approach has a limited application: e.g. if we are to solve an equation having a potential via minimizing this potential, we must look for accurate minimizers only, and in such case the classical existence and stability requirements are sensible indeed.

1. Topological background

We briefly outline some topics from the proximity-space theory which was proposed axiomatically by Efremovič [1] and which becomes now quite usual tool in general topology - for a survey (also historical) we refer to [2].

The pair $(X,\gg)$ will be called a proximity space if $\gg$ is a binary relation on 2^X satisfying the following six axioms:

i) $X \gg X$,

ii) $A \gg B \Rightarrow A \supset B$,

iii) $A_1 \supset A_2 \gg B_2 \supset B_1 \Rightarrow A_1 \gg B_1$,

iv) $A_1 \gg B,\ A_2 \gg B \Rightarrow A_1 \cap A_2 \gg B$,

v) $A \gg B \Rightarrow (X-B) \gg (X-A)$,

vi) $A_1 \gg A_2 \Rightarrow \exists B:\ A_1 \gg B \gg A_2$.

If $A \gg B$, we say that A is a proximal neighbourhood of B. Particularly, this relation induces a topology on X when B is taking as a singleton. The proximity δ is a binary relation on 2^X defined by $A\bar{\delta}B$ iff $(Y-B)\gg A$ ($\bar{\delta}$ means the negation of δ). Thanks to v), both δ and $\bar{\delta}$ are symmetric. If $A\delta B$ (or $A\bar{\delta}B$), we say that A and B are near to (or far from) each other. A natural example is the proximity induced by a metric d: $A\gg B$ iff A contains some ε-neighbourhood of B, i.e. $\exists\, \varepsilon > 0\ \forall x\in B:\ d(x,y)\leq\varepsilon \Rightarrow y\in A$.

Let us recall briefly some definitions more: A uniformity $\mathcal{U}$ on X is a filter on $X\times X$ such that $\forall U\in\mathcal{U}:\ \Delta\subset U,\ U^{-1}\in\mathcal{U}$, and $\exists\, V\in\mathcal{U}:\ V\circ V\subset U$, where $\Delta=\{(x_1,x_2)\mid x_1=x_2\}$, $U^{-1}=\{(x_1,x_2)\mid (x_2,x_1)\in U\}$, and $U\circ V=\{(x_1,x_2)\mid \exists x_3:\ (x_1,x_3)\in U,\ (x_3,x_2)\in V\}$. A filter base $\mathcal{B}$ on X is a subset of 2^X such that $\mathcal{B}\neq\emptyset$, $\emptyset\notin\mathcal{B}$, $A_1,A_2\in\mathcal{B}\Rightarrow A_1\cap A_2\in\mathcal{B}$. If, in addition, $A_1\supset A_2\in\mathcal{B}\Rightarrow A_1\in\mathcal{B}$, then $\mathcal{B}$ is called a filter on X. The sets $U(B)=\{x\in X\mid \exists x_1\in B:\ (x,x_1)\in U\}$ with some $U\in\mathcal{U}$ are called uniform neighbourhoods of B. The uniformity $\mathcal{U}$ is precompact if $\forall U\in\mathcal{U}\ \exists$ a finite subset M of X: $X=U(M)$. A uniformity is precompact if and only if the completion with respect to this uniformity is compact.

There exists just one precompact uniformity on X such that $A\gg B$ iff A is a uniform neighbourhood of B. The completion $\bar{X}$ of X with respect to this uniformity is thus compact and is called the Smirnov compactification of the proximity space $(X,\gg)$; see [7]. If X and Y are two uniform spaces, then every uniformly continuous mapping $X\to Y$ can be extended continuously to a mapping $\bar{X}\to\bar{Y}$ ($\bar{X},\bar{Y}$ are the corresponding Smirnov compactifications of X,Y regarding to the proximities induced by the respective uniformities). This extension is unique and, if X is a metric space, then even only uniformly continuous mappings can be thus extended.

The significant role of the proximity structure in optimization has been underlined in [3]. A full characterization of the Smirnov compactification by means of the possibility to extend (in a stable way) an explicitely constrained optimization problem has been stated in [6]: Let us consider a family of optimization problems ($p \in P$ is a parameter):

$$f(x,p) \to \inf \mid x \in G(p),$$

where $f: X \times P \to \bar{R} = R \cup \{+\infty, -\infty\}$ and $G: P \to 2^X$ (in other words, G is a multivalued mapping from P to X). Let P be a topological space, X be endowed by a uniformity $\mathcal{U}$, $f(\cdot,p): X \to \bar{R}$ be uniformly continuous for every $p \in P$, the family of functions $\{f(x,\cdot): P \to \bar{R}\}_{x \in X}$ be equicontinuous (we employ the compactness of $\bar{R}$), and G be upper and lower Hausdorff semicontinuous with respect to $\mathcal{U}$. Consider another uniformity $\mathcal{V}$ on X and the extended problem

$$f^{\mathcal{V}}(x,p) \to \inf \mid x \in G^{\mathcal{V}}(p),$$

where $f^{\mathcal{V}}: X^{\mathcal{V}} \times P \to \bar{R}$ is defined by $f^{\mathcal{V}}(x,p) = \liminf_{\tilde{x} \to x, \tilde{x} \in X} f(\tilde{x},p)$, $X^{\mathcal{V}}$ is the completion of X with respect to $\mathcal{V}$, and $G^{\mathcal{V}}(p)$ is the closure of G(p) in $X^{\mathcal{V}}$. Then the multivalued mapping $P \to X^{\mathcal{V}}$, assigning to p the set of minimizers of the corresponding extended problem, is upper semicontinuous and the infima of both the original and the extended problems are the same (for all families of the problems satisfying the hypotheses) if and only if $X^{\mathcal{V}}$ is the Smirnov compactification of X regarding to the proximity induced by $\mathcal{U}$. Then also $f^{\mathcal{V}}$ is the continuous extension of f.

2. Optimization problems treated with tolerance

Let us consider an abstract optimization problem with functional constraints:

$$f(x) \rightarrow \inf \mid x \in X,\ g(x) \in C,$$

where $f: X \rightarrow \bar{R}$ is a cost function, $g: X \rightarrow Y$, and $C \subset Y$. We want to follow the philosophy of Introduction: to minimize f only "up to $\varepsilon > 0$" and to fulfil the constraint also only "up to $\varepsilon > 0$". For this we need some structures on $\bar{R}$ and Y. We confine ourselves to a metric case (for more general case see [5]) supposing we are given by some metric d on Y inducing the relation $\gg$ on 2^Y. Naturally, we employ on $\bar{R}$ both the standard ordering and the only proximity inducing the standard compact topology of $\bar{R}$. We define the binary relation $>$ on $\bar{R}$ as follows: $a > b$ iff $[-\infty, a]$ is a proximal neighbourhood of $[-\infty, b]$. Clearly this relation has got the obvious meaning but the only exception: $+\infty > +\infty$. We will include the relations $>$ and $\gg$ into the data determining our problem, believing that it is more natural from the engineering point of view. Then our problem is determined by the 5-touple $(f, g, C, >, \gg)$, and we will write it symbolically as follows:

P: $f(x) \rightarrow \inf$ with tolerance given by $>$ $\mid x \in X,\ g(x) \in C$ with tolerance given by $\gg$.

From now on we suppose a generalized controlability condition: g(X) and C are near to each other (in the classical setting of the problem without tolerance we would have to suppose that g(X) and C intersect each other). To define notions analogous with the sets of solutions or of feasible points, we will investigate here the collection of level sets ($\alpha \in \bar{R}$):

$$\mathcal{B}_\alpha = \{f^{-1}([-\infty, a]) \cap g^{-1}(\tilde{C}) \mid a > \alpha,\ \tilde{C} \gg C\}.$$

It can be proved that there exists $\hat{\alpha} \in \bar{R}$ such that $\mathcal{B}_\alpha$ is a filter base on X if and only if $\alpha \geqslant \hat{\alpha}$. Particularly, $\mathcal{B}_\alpha$ is a filter base for $\alpha = +\infty$. It is natural to declare $\hat{\alpha}$ as the infimum of the problem, denoted by inf(P), and the filter generated on X by $\mathcal{B}_\alpha$ with $\alpha = \hat{\alpha}$ (or $\alpha = +\infty$) as a minimizing (or a feasible) filter, denoted by $\mathcal{M}(P)$ (or $\mathcal{F}(P)$), respectively. In other words, $\mathcal{M}(P)$ has got a base consisting of the sets:

$$\{x \in X \mid f(x) \leqslant \inf(P) + \varepsilon,\ \mathrm{dist}(g(x), C) \leqslant \varepsilon\}$$

with $\varepsilon > 0$; $\mathrm{dist}(y, C) = \inf\{d(y, \tilde{y}) \mid \tilde{y} \in C\}$. For a given accuracy ε this set contains just the elements for which the constraint is fulfilled "up to ε" and, at the same time, the cost function takes the lowest possible value (with regards to all $\varepsilon > 0$) "up to ε". These "ε-solutions" are obviously of the engineering or economical interest. Let us also mention the "principles of optimality" by D.A.Molodcov [9] since some of them can be considered as patterns from which the minimizing and the feasible filters can be obtained after a slight modification and generalization. Note that inf(P) depends on the chosen metric d: the finer the metric d, the greater the infimum of P. Particularly, for the discrete metric d we get the greatest value $\inf(P)=\inf f(g^{-1}(C))$, which is the classical infimum of the problem treated without any tolerance.

Let us define by the usual manner, cf. [8]: a sequence $s=\{x_n\}_{n\in N}$ is feasible if $\lim_{n\to\infty}\text{dist}(g(x_n),C))=0$, and minimizing if it is feasible and $\limsup_{n\to\infty} f(x_n) \leqslant \limsup_{n\to\infty} f(\tilde{x}_n)$ for every feasible sequence $\tilde{s}=\{\tilde{x}_n\}_{n\in N}$. A sequence s generates on X the so-called sequencial filter $\mathcal{S}(s)$ by means of the base $\{\{x_n | n\geqslant m\} | m\in N\}$. The following assertion justifies our terminology: a sequence s is minimizing or feasible if and only if the sequencial filter $\mathcal{S}(s)$ is finer than the minimizing or feasible filter, respectively.

On the contrary to the classical setting of the optimization problem without tolerance, our notions defined "with tolerance" are stable with respect to some perturbations of the data f, g, and C, which makes them sensible. Generalizing the "stability from above" of the optimality principles by D.A.Molodcov[9], we say that a filter $\mathcal{F}$ is a lower bound for a sequence of filters $\{\mathcal{F}_n\}_{n\in N}$ if $\forall A\in\mathcal{F}\ \exists m\in N\ \forall n\geqslant m$: $A\in\mathcal{F}_n$. It is interesting that there always exists (just one) finest lower bound, denote it by $\liminf_{n\to\infty}\mathcal{F}_n$. The notion of the lower bound can be used for certain stability of the sequences by means of the following assertion: if $\liminf_{n\to\infty}\mathcal{F}_n\supset\mathcal{F}$, $\mathcal{F}$ has a countable base, $s^n=\{x_k^n\}_{k\in N}$ be sequences such that $\mathcal{S}(s^n)\supset\mathcal{F}_n$, then there is $\varkappa:N\to N$ such that for every $s=\{x_{k(n)}^n\}_{n\in N}$ with $k(n)\geqslant\varkappa(n)$ we have $\mathcal{S}(s)\supset\mathcal{F}$. Hence, for perturbed problems P_n we get a stability of the minimizing or feasible sequences if we prove that $\mathcal{M}(P)$ or $\mathcal{F}(P)$ is a lower bound for $\{\mathcal{M}(P_n)\}_{n\in N}$ or $\{\mathcal{F}(P_n)\}_{n\in N}$, respectively. Indeed we have the following results: If f_n converge to f uniformly (it suffices from below), g_n converge to g uniformly with respect to the metric d, and C_n converge "from above" to C in the sense $\forall \tilde{C} \gg C\ \exists m\in N\ \forall n\geq m$: $C_n \subset \tilde{C}$, then $\liminf_{n\to\infty} \inf(P_n) \geq \inf(P)$ and $\liminf_{n\to\infty} \mathcal{F}(P_n) \supset \mathcal{F}(P)$, where P_n means naturally the perturbed problem arising from P by replacing f, g, C with f_n, g_n, C_n, respectively. Under stronger assumptions we can prove even more: If f_n converge to f uniformly, $g_n \equiv g$, C_n converge to C "from above", and $C_n \supset C$, then $\inf(P_n) \to \inf(P)$ (i.e. the infimum depends continuously on the data), $\liminf_{n\to\infty} \mathcal{M}(P_n) \supset \mathcal{M}(P)$ and $\liminf_{n\to\infty} \mathcal{F}(P_n) = \mathcal{F}(P)$.

The tolerance approach is closely related with some usual numerical methods which avoid the implicit constraint, especially with the exterior penalty function method. Let $h: R \to R$ be an increasing, continuous function with $h(0)=0$. We can pose the penalized, unconstrained problem $(r\in R^+)$:

$P_r: f_r(x) \to \inf$ with tolerance given by $>$ | $x\in X$,

where $f_r(x) = f(x) + r.h(\mathrm{dist}(g(x),C))$. Applying the above introduced definitions to (P_r), we get clearly: $\inf(P_r) = \inf f_r(X)$, $\mathcal{F}(P_r) = \{X\}$, and $\mathcal{M}(P_r) = \{A\subset X \mid \exists a > \inf(P_r): A \supset f_r^{-1}([-\infty,a])$. Supposing f to be bounded from below and $\inf(P) \neq +\infty$, we can obtain the convergence results: $\inf(P_r) \nearrow \inf(P)$ and $\liminf_{r\to\infty} \mathcal{M}(P_r) \supset \mathcal{M}(P)$. Moreover, if we define for $\varepsilon > 0$ the filters $\mathcal{M}_\varepsilon(P_r) = \{A\subset X \mid A \supset f_r^{-1}([-\infty$

, $\inf(P_r)+\varepsilon]$) and the nonincreasing nonnegative function $e:R_+\to R$ by $e(r)=\inf(P)-\inf(P_r)$, we have even $\mathcal{M}(P) = \liminf_{r\to\infty,\varepsilon\to o,\varepsilon>e(r)} \mathcal{M}(P_r)$. Convergence results can be obtained also for a large class of the augmented Lagrangeans methods; we refer to [5]. We may observe that there is a large class of dual problems that have the same supremum depending, in fact, only on the chosen proximity on Y. Besides, this supremum is equal to inf(P), it means that there is no (i.e. zero) duality gap. It is caused simply by consistency of the setting of P with tolerance and of the construction of the dual problem by the same tolerance, using the Lindberg perturbational theory of duality. On the contrary, in standard approach such consistency is missing, hence it is not surprising that additional requirements must be imposed on the data to ensure zero duality gap.

3. Compactification of the optimization problems

We may have observed that the optimization problems posed with tolerance exhibit good behavior like problems with continuous mappings f and g, closed C, and compact X, which will be explained in what follows. Let $\bar{f}:\bar{X}\to\bar{R}$, $\bar{g}:\bar{X}\to\bar{Y}$, $\bar{C}\subset\bar{Y}$. The optimization problem:

$$\bar{P}: \quad \bar{f}(x) \to \inf \;\Big|\; x\in\bar{X},\ \bar{g}(x)\in\bar{C}$$

(considered in the classical sense, i.e. without tolerance) will be called a compactification of P if X is a compact topological space with a topology $\bar{\sigma}$, X is a $\bar{\sigma}$-dense subset of $\bar{X}$, $Y \subset \bar{Y}$, $\bar{Y}$ is a uniform space with a uniformity $\bar{\mathcal{Y}}$ whose trace on Y induces the same proximity as the given metric d, $\bar{f}$ and $\bar{g}$ are continuous and their restrictions on X are just f and g, respectively, and $\bar{C}$ is the closure of C in $\bar{Y}$. Every problem P admits at least one compactification that can be constructed as follows: take the discrete proximity on X (i.e. every disjoint subsets are far from each other) and the proximity induced by d on Y; then f and g are obviously proximally continuous (i.e. they map sets that are near to each other onto sets being again near to each other) and can be thus extended continuously on the corresponding Smirnov compactifications of X and Y which are taken for $\bar{X}$ and $\bar{Y}$, respectively. This is the "largest" compactification in the sense that it uses the finest compactification of X (the elements of $\bar{X}$ can be then identified just with the ultrafilters on X). Nevertheles, P may admit generally a large amount of the compactifications. E.g. if X is a uniform space and f and g are uniformly continuous, we can take for $\bar{X}$ the Smirnov compactification of X regarding to the proximity induced by the considered uniformity; cf. [4]. This compactification is even the "smalest" one in the sense that, under these uniform-continuity assumptions, we cannot use any strictly coarser compactification of X. If X is a normed linear space (considered as a uniform space), we can extend on this compactification also the neccessary conditions for the minimum of $\bar{P}$ by using of the Ekeland ε-variational principle; for the unconstrained case see [3].

All the compactifications of P has the following common properties: $\inf(P)=\min(\bar{P})$, and $\mathcal{M}(P)$ or $\mathcal{F}(P)$ is the trace on X of the $\bar{\tau}$-neighbourhood filter of the set of minimizers of $\bar{P}$ or of the set of feasible points of $\bar{P}$, respectively. It enables to prove the results stated in Sec.2 and also some other results by the following way: first, transfer the properties of the data to the extended, compactified problem; then use standard techniques exploiting the compactness; and afterwards transfer the obtained results back to the original problem. Thus via compactification we can study either the problems posed with tolerance, or the classical minimizing and feasible sequences. Alternatively we may declare the compactified problem as a natural extension of the original problem, obtaining thus certain generalized solutions of it; cf. [4]. It is analogical with what is made in the relaxed-control theory [8] where the situation is, however, rather simple from the viewpoint of our general approach because X, the space of all measurable functions from $[0,T]$ to a compact subset S of R^n, is precompact in the uniformity induced by imbedding this set properly into the dual space of $L^1(0,T;C^o(S))$ (L^1 and C^o means the space of integrable and of continuous functions, respectively) endowed with the weak* uniformity; for details see [8].

References

[1] V.A.Efremovič: The geometry of proximity I. (in Russian) Mat. Sb. 31 (73) (1952), 189-200.

[2] S.A. Naimpally, B.D. Warrack: Proximity Spaces. Cambridge Univ. Press, Cambridge, 1970.

[3] T. Roubíček: A generalized solution of a nonconvex minimization problem and its stability. Kybernetika 22 (1986), 289-298.

[4] T. Roubíček: Generalized solutions of constrained optimization problems. SIAM J. Control Optim. 24 (1986), 951-960.

[5] T. Roubíček: Constrained optimization: a general tolerance approach. Aplikace Matematiky (submitted).

[6] T. Roubíček: Stable extensions of constrained optimization problems. J. Math. Anal. Appl. (accepted).

[7] Ju.M.Smirnov: On proximity spaces (in Russian). Mat. Sb. 31 (73) (1952), 543-574. English transl. in: AMS Transl. Ser.2, 38, 5-35.

[8] J. Warga: Optimal Control of Differential and Functional Equations. Academic Press, New York, 1972.

[9] D.A.Molodcov: Stability and regularization of principles of optimality (in Russian). Žurnal vyčisl. mat. i mat. fiz. 20 (1980), 1117-1129.