Aplikace matematiky

Tomáš Roubíček

Constrained optimization: A general tolerance approach

Aplikace matematiky, Vol. 35 (1990), No. 2, 99–128

Persistent URL: http://dml.cz/dmlcz/104393

CONSTRAINED OPTIMIZATION: A GENERAL TOLERANCE APPROACH

Tomáš Roubíček

Dedicated to the Czechoslovak students who struck for tolerance and nonviolence in the Fall of 1989

(Received May 10, 1988)

Summary. To overcome the somewhat artificial difficulties in classical optimization theory concerning the existence and stability of minimizers, a new setting of constrained optimization problems (called problems with tolerance) is proposed using given proximity structures to define the neighbourhoods of sets. The infimum and the so-called minimizing filter are then defined by means of level sets created by these neighbourhoods, which also reflects the engineering approach to constrained optimization problems. Moreover, an appropriate concept of convergence of filters is developed, and stability of the minimizing filter as well as its approximation by the exterior penalty function technique are proved by using a compactification of the problem.

Keywords: constrained optimization, level sets, minimizing sequences, penalty functions, compactifications.

AMS Classification: 49A27, 65K10, 54D35.

1. MOTIVATION AND DEFINITIONS

It can be roughly said that the paper has three aims. First it tries to suggest a possible way how to put right, by using proper definitions, a frequent misunderstanding appearing in optimization theory in which sometimes an inadequate effort is exerted on questions about the existence of optimal solutions and their stability. Indeed, in optimization problems of technical origin, where the data themselves are not known exactly, engineers certainly do not seek exact solutions, but only ε-ones. Therefore, from their point of view, problems concerning the existence of exact solutions look somewhat artificial. Also the possible instability of the exact solutions, which is mostly regarded as a bad property, may be rather a good one if the optimization problems have a character of inverse problems. The instability means, roughly speaking, that the solutions vary considerably while the data vary only little, or, vice versa, the values of the cost fuction and of the mapping representing some constraints vary negligibly when the solutions change considerably. Yet it is actually a good situation because such changed solutions are almost as good as the original ones.

The philosophy of our "tolerance approach" is to replace the exact solutions not by the ε-ones (as has been often done already in the classical optimization theory), but by a collection of these ε-solutions arising when ε passes to zero from above. It will be shown that such approach has a compactifying character and enables us to treat problems posed "with tolerance" by applying standard methods using a certain "closure" of such problems, called compactification. The compactified problems are then treated as classical optimization problems without tolerance. Besides, the tolerance setting of optimization problems ensures apriori certain stability and approximative properties without any data qualification hypothesis (like compactness, continuity, etc.), which cannot appear within the classical setting of the problems.

Of course, our tolerance approach has limited applicability; e.g. if the cost function to be minimized were a potential of some equation, we would actually have to look for exact solutions (= minimizers) because only such solutions can solve the original equation. Yet, our standpoint will be that the function to be minimized is a cost and then the ε-solutions are almost as good as the exact ones (if the latter ones do exist at all). In the presence of constaints treated also "with tolerance" we shall see that they may be even better (i.e. they may achieve a strictly lower cost than the infimum of the problem in the classical setting without tolerance).

As to the second aim, our tolerance approach can be readily applied to a study of minimizing sequences. In the classical optimization theory we define minimizers, and afterwards the set of all minimizers, to study stability behaviour of the minimizers. If one takes, instead of the minimizers, the minimizing sequences as a more advanced concept for solutions (see e.g. [3, 8, 13]), then one immediately realizes the lack of any notion analogous to the set of all minimizers, which is schematically shown by the following diagram:

$$\begin{array}{ccc} \text{a minimizer} & \in & \text{the set of all minimizers} \\ \updownarrow & & \updownarrow \\ \text{a minimizing sequence} & \in & ???\end{array}$$

Roughly speaking, this gap will be filled in by our definition of the minimizing filter (cf. Definition 1.1 together with Remark 1.3 below).

As for the third aim, our definitions of the minimizing and feasible filters may be considered as a generalization of some (slightly modified) "principles of optimality" due to D. A. Molodcov [4, 5]. In particular, our modification of these optimality principles reduces considerably the "$\varepsilon - \delta$ gymnastics" and enables us to employ systematically the standard methods of general topology, which makes all considerations easier to understand (cf. [5]). On the other hand, the optimality principles and especially their stability from above (see Sec. 2 below) contain a bit more information than our minimizing filters and a lower bound of a net of such filters.

Besides, the compactified problems corresponds to what is called relaxed problems in certain special cases (see e.g. J. Warga [13], cf. also Example 6.2 below). However, since the compactified problems will serve only as an auxiliary tool, we will not investigate them here in details (which makes the essential difference between this paper and the former author's works [9–11]).

We will systematically use the proximity space theory which is a proper tool for our approach to optimization problems; cf. also [9–11]. As this theory is not usual in optimization, all notions needed will be defined here. Generally speaking, proximity structures enable us to define neighbourhoods of subsets, being thus "coarser" than uniformities and "finer" than topologies. The proximity theory originated in general topology by the works by V. A. Efremovich [2] and Yu. Smirnov [12] in early 1950's. For a survey of this theory we refer e.g. to [7] or [1].

Hereafter we use the prefix notation whenever the structures in question may not be clear, e.g. instead of saying that a mapping is continuous with respect to the topologies $\mathscr{T}_1$ and $\mathscr{T}_2$ we say briefly that the mapping is $(\mathscr{T}_1, \mathscr{T}_2)$-continuous, etc.

Let us briefly recall some definitions. A proximity δ on X is a binary relation on 2^X (=the power set of X) such that

$$A_1 \delta A_2 \Rightarrow A_1 \delta A_2 ,$$

$$(A_1 \cup A_2)\, \delta B \Rightarrow A_1 \delta B \quad \text{or} \quad A_2 \delta B ,$$

$$A_1 \cap A_2 \neq \emptyset \Rightarrow A_1 \delta A_2 ,$$

$$A_1 \delta A_2 \Rightarrow A_1 \neq \emptyset \quad \text{and} \quad A_2 \neq \emptyset ,$$

$$A_1 \bar{\delta} A_2 \Rightarrow \exists B\colon A_1 \bar{\delta} B \quad \text{and} \quad (X \setminus B)\, \bar{\delta} A_2 ,$$

where $\bar{\delta}$ means the negation of δ. If $A_1 \delta A_2$ or $A_1 \bar{\delta} A_2$, the sets A_1 and A_2 are said to be near to each other or far from each other, respectively. We will also use the dual relation to $\bar{\delta}$, denoted by $\gg$; it means $A_1 \gg A_2$ iff $(X \setminus A_1)\, \bar{\delta} A_2$. If $A_1 \gg A_2$, we say that A_1 is a δ-proximal neighbourhood of A_2. Since the relation $\gg$ has not got any standard name in general topology, we dare call it tolerance here.

Every proximity δ induces a topology, denoted by $\mathscr{T}_\delta$, by declaring $\{x; \{x\}\, \delta A\}$ to be the $\mathscr{T}_\delta$-closure of A. A typical example of a proximity is the proximity δ_d induced by a metric d on X: $A_1 \delta_d A_2$ iff $\hat{d}(A_1, A_2) = 0$ where $\hat{d}(A_1, A_2) = \inf d(A_1, A_2)$ is the distance between the sets A_1 and A_2 with the convention $\inf \emptyset = +\infty$. Then $A \gg B$ means that there is an ε-neighbourhood B_ε of B such that $A \supset B_\varepsilon$, where $B_\varepsilon = \{x \in X; \hat{d}(\{x\}, B) \leqq \varepsilon\}$, $\varepsilon > 0$. Another example of a proximity is the discrete proximity which makes near only sets with a nonempty intersection; the corresponding tolerance is then called discrete.

Let us recall that $\mathscr{A} \subset 2^X$ is a filter base on X iff $\mathscr{A} \neq \emptyset$, $\emptyset \notin \mathscr{A}$, and $A_1, A_2 \in \mathscr{A} \Rightarrow \exists B \in \mathscr{A}\colon B \subset A_1 \cap A_2$. If, in addition, $A \supset B \in \mathscr{A} \Rightarrow A \in \mathscr{A}$, then $\mathscr{A}$ is a filter on X. For example, the collection $A_\gg = \{B; B \gg A\}$ is a filter whenever A is nonempty. For every filter base $\mathscr{A}$ the collection $\{A; \exists B \in \mathscr{A}\colon B \subset A\}$ is a filter; we say that it is generated by $\mathscr{A}$. For two filters $\mathscr{A}_1$, $\mathscr{A}_2$ on X we say that $\mathscr{A}_1$ is finer than $\mathscr{A}_2$

(or $\mathscr{A}_2$ is coarser than $\mathscr{A}_1$) if $\mathscr{A}_1 \supset \mathscr{A}_2$. A collection $\{\mathscr{A}_i\}_{i\in I}$ of filters on X has an upper bound iff $\bigcap_{i\in J} A_i \neq \emptyset$ for every finite subset J of I and $A_i \in \mathscr{A}_i$. Then the filter generated by the base $\{\bigcap_{i\in J} A_i;\ J \subset I \text{ finite}, A_i \in \mathscr{A}_i\}$ is called the upper bound of $\{\mathscr{A}_i\}_{i\in I}$.

Now we pose several "syntactic" rules by means of which we will write our optimization problems in the "tolerance" notation.

Rules.

i) The optimization problem will be the following formula: "*Minimize* $f(x)$ *on* X *with tolerance* $\gg$ *subject to* $\mathscr{A}$", where f is a function $X \to \bar{R}$, $\bar{R} = R \cup \{+\infty, -\infty\}$, $\gg$ is a tolerance on $\bar{R}$, and $\mathscr{A}$ is a filter on X.

ii) For $C \neq \emptyset$, by saying "C *with tolerance* $\gg$" we will mean the filter $C_{\gg}$.

iii) For $g: X \to Y$ and $\mathscr{A}$ a filter on Y, by saying "$g(x)$ *meets* $\mathscr{A}$" we will mean the filter on X generated by the base $g^{-1}(\mathscr{A})$, i.e. the filter $\{A \subset X;\ \exists B \in \mathscr{A}: g^{-1}(B) \subset A\}$ (if it is a filter at all).

iv) The logical conjunction of statements representing filters on some set will mean the upper bound of the corresponding filters (if it is a filter).

v) If $\gg$ is the discrete tolerance, then instead of "*with tolerance* $\gg$" we will say "*without tolerance*".

vi) "*Minimize* $f(x)$ *on* X *with tolerance* $\gg$" will mean "*Minimize* $f(x)$ *on* X *with tolerance* $\gg$ *subject to* $\{X\}$".

Now, we define notions analogous to the set of all feasible points, to the infimum and to the set of all minimizers in the classical optimization theory.

Definition 1.1. *Let* $(\boldsymbol{P})$ *be an optimization problem according to Rule i), i.e.* "Minimize $f(x)$ on X with tolerance $\gg$ subject to $\mathscr{A}$". *Then we put*

$$\mathscr{F}(\boldsymbol{P}) = \mathscr{A}\,,$$

$$\inf(\boldsymbol{P}) = \sup_{A\in\mathscr{A}} \inf_{x\in A} f(x)\,, \quad \textit{and}$$

$$\mathscr{M}(\boldsymbol{P}) = \{A \subset X;\ \exists \hat{A} \in \mathscr{A},\ B \gg [-\infty, \inf(\boldsymbol{P})]: \hat{A} \cap f^{-1}(B) \subset A\}\,,$$

and call them the feasible filter, the infimum, and the minimizing filter of $(\boldsymbol{P})$, *respectively (provided* $\mathscr{M}(\boldsymbol{P})$ *is a filter at all).*

Remark 1.1. Obviously, $\mathscr{F}(\boldsymbol{P})$ and $\inf(\boldsymbol{P})$ do not depend on the tolerance with which f is to be minimized. Since $\mathscr{A}$ is a filter and the mapping $A \mapsto \inf f(A)$ is monotone, we may define alternatively $\inf(\boldsymbol{P}) = \lim_{A\in\mathscr{A}} \inf_{x\in A} f(x)$. If $\gg$ is the tolerance on $\bar{R}$ such that the corresponding proximity induces the standard compact topology of $\bar{R}$, then it is quite evident that $\inf(\boldsymbol{P})$ is the lowest value α for which $\mathscr{A}$ and the filter generated by $f^{-1}([-\infty, \alpha]_{\gg})$ have the upper bound (which is then equal just to $\mathscr{M}(\boldsymbol{P})$). On the other hand, $\mathscr{M}(\boldsymbol{P})$ may contain the empty set and thus need not be a filter on X if $\gg$ would be, e.g., the discrete tolerance on $\bar{R}$ (i.e. if $f(x)$ would have been minimized on X without tolerance).

Example 1.1. Let us consider the problem in the classical notation:

$$\begin{cases} \text{minimize } f(x) \text{ on } X \\ \text{subject to } x \in A \text{ and} \\ \qquad\qquad g(x) \in B \cap C\,. \end{cases}$$

For simplicity suppose $g(A) \cap B \cap C \neq \emptyset$. Exploiting Rules i)–v), we can paraphraze this problem in terms of tolerance as follows (using also the discrete proximity, for example):

$$(\boldsymbol{P}_{\text{example}}) \begin{cases} \text{minimize } f(x) \text{ on } X \text{ with tolerance } \gg_0 \\ \text{subject to } x \text{ meets } A \text{ with tolerance } \gg_1 \text{ and} \\ \qquad\qquad g(x) \text{ meets } B \text{ with tolerance } \gg_2 \text{ and} \\ \qquad\qquad\qquad\qquad C \text{ without tolerance,} \end{cases}$$

where $\gg_0$, $\gg_1$, and $\gg_2$ are some tolerances on $\bar{R}$, X, and Y, respectively, $g: X \to Y$, $B, C \subset Y$, $A \subset X$. By Definition 1.1, e.g. $\mathscr{F}(\boldsymbol{P}_{\text{example}}) = \{D \subset X;\ \exists \hat{A} \gg_1 A, \hat{B} \gg_2 B\colon \hat{A} \cap g^{-1}(\hat{B} \cap C) \subset D\}$.

We will investigate the abstract optimization problem which would be written in the classical notation as follows: minimize $f(x)$ on X subject to $g(x) \in C$, where $f: X \to \bar{R}$, $g: X \to Y$, $C \subset Y$. For problems of more complicated structure see Sec. 6. To distinguish proximities (or other structures) on different sets, we will employ a subscript (thus δ_X, δ_Y, etc. will mean some proximities on X or Y, respectively). As for the proximity $\delta_{\bar{R}}$, in what follows we will confine ourselves to the case when $\delta_{\bar{R}}$ induces the standard compact topology of $\bar{R}$ (thus $\delta_{\bar{R}}$ is determined uniquely), and the corresponding tolerance on $\bar{R}$ will be denoted by $>$ without causing any discrepancy with the usual ordering of the extended real line; obviously, for $a, b \in \bar{R}$ we have $[-\infty, a] > [-\infty, b]$ if and only if $a > b$ or $a = b = +\infty$. Furthermore, given some fixed proximity δ_Y on Y, $\gg$ will denote the corresponding tolerance, if it is not said otherwise. Using the just introduced tolerance notation, we will investigate the optimization problem

$$(\boldsymbol{P}) \begin{cases} \text{minimize } f(x) \text{ on } X \text{ with tolerance } > \\ \text{subject to } g(x) \text{ meets } C \text{ with tolerance } \gg. \end{cases}$$

To satisfy Rule i), "$g(x)$ meets C with tolerance" must be a filter on X. This is true if and only if

(1.1) $\qquad g(X)$ and C are δ_Y-near to each other, i.e. $g(X)\ \delta_Y C$.

Obviously, (1.1) is weaker than the classical feasibility condition $g(X) \cap C \neq \emptyset$. Being important for all the forthcoming results, (1.1) will be implicitly assumed in what follows.

Note that $\inf(\boldsymbol{P})$ depends on the proximity δ_Y in a monotone manner, namely the finer the proximity δ_Y (i.e. the smaller with respect to the ordering of binary relations on 2^Y by inclusion), the greater the infimum of $(\boldsymbol{P})$. In particular, for the finest proximity on Y (i.e. for the discrete proximity) we obtain the greatest value for the infimum of $(\boldsymbol{P})$, namely $\inf f(g^{-1}(C))$, which is obviously the infimum of the problem treated in the classical sense, that means without tolerance. Let us remark that, for a general tolerance $\gg$, the non-negative quantity $\inf f(g^{-1}(C)) - \inf(\boldsymbol{P})$ is just what is called a duality gap in the classical optimization theory (of course, when the perturbations by means of which a dual problem is constructed are taken in accord with our tolerance), cf. [10; Sec. 6.2]. Note that if $\inf(\boldsymbol{P}) \neq \inf f(g^{-1}(C))$ and C is $\mathscr{T}_{\delta_Y}$-closed, then $\mathscr{M}(\boldsymbol{P})$ is necessarily a free filter on X, which means $\bigcap_{A \in \mathscr{M}(\boldsymbol{P})} A = \emptyset$. Under some qualification hypotheses about the data f, g, C, and δ_Y, we can even ensure that the gap $\inf f(g^{-1}(C)) - \inf(\boldsymbol{P})$ is zero. Suppose that there exists a proximity δ_X on X such that

(1.2) $\qquad A_1 \delta_X A_2 \Rightarrow f(A_1)\, \delta_{\bar{R}}\, f(A_2)\,,$ and

(1.3) $\qquad \forall B \gg g^{-1}(C)\; \exists \hat{C} \gg C \colon g^{-1}(\hat{C}) \subset B\,,$

where the first tolerance $\gg$ corresponds to δ_X while the second to δ_Y. By definition, (1.2) means that f is $(\delta_X, \delta_{\bar{R}})$-proximally continuous. In particular, (1.3) is fulfilled if g^{-1} is a singlevalued (δ_Y, δ_X)-proximally continuous mapping. If the proximities δ_X, δ_Y and $\delta_{\bar{R}}$ are induced by some metrics d_X, d_Y, and $d_{\bar{R}}$, respectively, then (1.2) means recisely that f is uniformly continuous in the usual sense and (1.3) is guaranteed when the (possibly multivalued) mapping g^{-1} is uniformly Hausdorff continuous; it means $\forall \varepsilon > 0\; \exists \eta > 0\; \forall y_1, y_2 \in Y \colon d_Y(y_1, y_2) \leqq \eta \Rightarrow g^{-1}(y_1) \subset \{x \in X;\ \hat{d}_X(\{x\}, g^{-1}(y_2)) \leqq \varepsilon\}$. Also, (1.3) is valid when δ_X induces a compact topology on X, g is continuous and C is closed.

Realize that, for given data f, g, C, δ_Y, (1.2) requires δ_X to be fine enough, while (1.3) requires δ_X to be sufficiently coarse, hence (1.2) together with (1.3) represent actually a qualification hypothesis about the data. It is clear that such δ_X does exist only when δ_Y is fine enough (particularly, it always exists if δ_Y is discrete).

Proposition 1.1. *Let $g^{-1}(C) \neq \emptyset$ and let there exist a proximity δ_X satisfying* (1.2) *and* (1.3). *Then* $\inf(\boldsymbol{P}) = \inf f(g^{-1}(C))$.

Proof. It suffices to show that $\inf(\boldsymbol{P}) \geqq \inf f(g^{-1}(C))$. We may suppose that $\inf f(g^{-1}(C)) > -\infty$, because the converse case is trivial. First, we treat the case $\inf f(g^{-1}(C)) \neq +\infty$. We will show that $\forall \varepsilon > 0\; \exists C_\varepsilon \gg C \colon \inf f(g^{-1}(C_\varepsilon)) \geqq$ $\geqq \inf f(g^{-1}(C)) - \varepsilon$. As $g^{-1}(C)$ is nonempty, by (1.2) we can choose $B_\varepsilon \gg g^{-1}(C)$ such that $\inf f(B_\varepsilon) \geqq \inf f(g^{-1}(C)) - \varepsilon$, and afterwards by (1.3) we can take $C_\varepsilon \gg C$ such that $g^{-1}(C_\varepsilon) \subset B_\varepsilon$. We get $\inf(\boldsymbol{P}) \geqq \inf f(g^{-1}(C_\varepsilon)) \geqq \inf f(g^{-1}(C)) - \varepsilon$.

In case $\inf f(g^{-1}(C)) = +\infty$ we can show in the same way that $\forall \varepsilon > 0\; \exists C_\varepsilon \gg C$: $\inf f(g^{-1}(C_\varepsilon)) \geqq 1/\varepsilon$. □

If (1.2) and (1.3) cannot be satisfied by any δ_X, a non-zero gap can actually appear. It happens typically when X and Y are infinite-dimensional Banach spaces, f is uniformly continuous with respect to the norm of X, δ_Y is induced by the norm of Y, and g is a compact operator. Then the problem with tolerance may offer "better" solutions (i.e. at lower cost) than that without tolerance. The same situation occurs in the relaxed-control theory [13] where g is typically governed by a differential equation and the relaxed controls may achieve lower cost than the ordinary ones.

Let us illustrate Definition 1.1 for $(\boldsymbol{P})$ by the case when δ_Y is induced by a metric d_Y. Then the feasible filter $\mathscr{F}(\boldsymbol{P})$ has a base

$$\{\{x \in X;\ \hat{d}_Y(\{g(x)\}, C) \leqq \varepsilon\};\ \varepsilon > 0\}$$

and, if $\inf(\boldsymbol{P}) \neq -\infty$, the minimizing filter $\mathscr{M}(\boldsymbol{P})$ has a base

$$\{\{x \in X;\ f(x) \leqq \inf(\boldsymbol{P}) + \varepsilon,\ \hat{d}_Y(\{g(x)\}, C) \leqq \varepsilon\};\ \varepsilon > 0\}\ .$$

In particular, both $\mathscr{F}(\boldsymbol{P})$ and $\mathscr{M}(\boldsymbol{P})$ have countable bases. Besides, they may be considered as generalizations (after slight modifications) of some principles of optimality in the sense of D. A. Molodcov [4, 5]. Let us confine ourselves to the more illustrative case of $\mathscr{M}(\boldsymbol{P})$. In the case $Y = R^n$, $C = (R^+)^n$, $R^+ = [0, +\infty[$, $g = (g_1, \ldots, g_n)$, and δ_Y the Euclidean proximity on R^n, Molodcov [4] introduced the principle of optimality which can be rewritten in our notation (with a discrete proximity on X) as follows:

$$\begin{aligned}&\operatorname{Min}(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n, \alpha_1, \ldots, \alpha_n) =\\&= \{x \in X;\ f(x) \leqq \iota(\alpha_1, \ldots, \alpha_n) + \varepsilon_0,\ g_1(x) + \varepsilon_1 \geqq 0, \ldots, g_n(x) +\\&+ \varepsilon_n \geqq 0\}\ ,\end{aligned}$$

where $\iota(\alpha_1, \ldots, \alpha_n) = \inf\{f(x);\ g_1(x) + \alpha_1 \geqq 0, \ldots, g_n(x) + \alpha_n \geqq 0\}$. It is easy to see that $\iota(\alpha_1, \ldots, \alpha_n)$ converges from below to $\inf(\boldsymbol{P})$ if all $\alpha_i \searrow 0$. Clearly, for $\varepsilon_i > 0$ the sets of the form $\operatorname{Min}(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n, \ldots)$ with $\iota(\alpha_1, \ldots, \alpha_n)$ replaced by $\inf(\boldsymbol{P})$ generate just the minimizing filter $\mathscr{M}(\boldsymbol{P})$. Such an approach to optimization problems, admitting "ε-tolerance" both in the cost function and the constraints, is undoubtedly very realistic from the viewpoint of applications in technical practice.

To justify our definition even more, we will show a connection with the standard notion of minimizing or feasible sequences. For a sequence $s = \{s_n\}_{n \in N}$, where $s_n \in X$ and N is the set of natural numbers, we define the so-called sequential filter $\mathscr{S}(s)$ as a filter on X generated by the base $\{\{s_n \in X;\ n \geqq m\};\ m \in N\}$.

Definition 1.2. *A sequence $s = \{s_n\}_{n \in N}$ is called feasible or minimizing for $(\boldsymbol{P})$ if the corresponding sequential filter $\mathscr{S}(s)$ is finer than the feasible or the minimizing filter, respectively.*

If δ_Y is induced by a metric d_Y, a sequence s is feasible if and only if $\lim_{n\to\infty} \hat{d}_Y(\{g(s_n)\}, C) = 0$.

Proposition 1.2. *For any sequence $s = \{s_n\}_{n \in N}$ the following statements are equivalent to each other.*

(a) *s is a minimizing sequence for $(\boldsymbol{P})$,*

(b) *s is feasible and* $\limsup_{n \to \infty} f(s_n) \leqq \inf(\boldsymbol{P})$,

(c) *s is feasible and* $\lim_{n \to \infty} f(s_n) = \inf(\boldsymbol{P})$.

If, in addition, the filter $C_{\gg}$ has a countable base, then they are also equivalent to the statements

(d) *s is feasible and* $\limsup_{n \to \infty} f(s_n) \leqq \limsup_{n \to \infty} f(\tilde{s}_n)$ *for every feasible sequence $\tilde{s} = \{\tilde{s}_n\}_{n \in N}$,*

(e) *s is feasible and* $\lim_{n \to \infty} f(s_n) \leqq \liminf_{n \to \infty} f(\tilde{s}_n)$ *for every feasible sequence $\tilde{s} = \{\tilde{s}_n\}_{n \in N}$.*

Proof. By the definition, (a) is equivalent to: $\forall a > \inf(\boldsymbol{P})\ \forall \tilde{C} \gg C\ \exists m \in N\ \forall n \geqq m$: $f(s_n) \leqq a$ and $g(s_n) \in \tilde{C}$, which is nothing else than (b). Suppose that $\alpha = \liminf_{n \to \infty} f(s_n) < \inf(\boldsymbol{P})$. Then, for every $a > \alpha$ and $\tilde{C} \gg C$, there would exists $s_n \in \operatorname{lev}_a f \cap g^{-1}(\tilde{C})$, in particular the level set $\operatorname{lev}_a f \cap g^{-1}(\tilde{C})$ would be nonempty, hence $\sup_{\tilde{C} \gg C} \inf_{x \in g^{-1}(\tilde{C})} f(x) \leqq \alpha < \inf(\boldsymbol{P})$, which contradicts Definition 1.1. Therefore, (b) $\Rightarrow$ (c). The converse implication is trivial. By the same argument we obtain that $\limsup_{n \to \infty} f(s_n) < \inf(\boldsymbol{P})$ and $\liminf_{n \to \infty} f(s_n) < \inf(\boldsymbol{P})$ is not possible provided s is feasible, thus (c) implies (d) and (e). Moreover, if $\{C_n\}_{n \in N}$ is the countable base of the filter $C_{\gg}$, we can take $\tilde{s}_n \in \operatorname{lev}_{\inf(\boldsymbol{P}) + 1/n} f \cap g^{-1}(C_n)$, which gives a feasible sequence $\{\tilde{s}_n\}_{n \in N}$ with $\lim_{n \to \infty} f(\tilde{s}) = \inf(\boldsymbol{P})$. Then obviously (d) $\Rightarrow$ (b) and (e) $\Rightarrow$ (c). □

Remark 1.2. The statements (d) or (e) are sometimes used for the definition of minimizing sequences; we refer e.g. to J. Warga [13; III.2] who used (e) or E. Polak and Y. Y. Wardi [8] who used (d) for some special problems. However, somewhat different terms are usually used (asymptotically or eventually feasible sequences and asymptotically minimizing sequences or minimizing approximate solutions, etc.). We also refer to E. G. Golshtein [3] who required, in addition, that the feasible sequences mapped by f have a limit in $\bar{R}$ (and then called them generalized plans). For the minimizing sequences this additional requirement is fulfilled automatically, however; see (c). Of course, these standard notions will coincide with that of ours only if the tolerance $\gg$ is taken appropriately (namely if $\gg$ is induced by the metric d_Y used for the definition of these stadard notions).

Remark 1.3. It can be easily demonstrated that, if the filter $C_{\gg}$ has got a countable base, then the feasible (or minimizing) filter of $(\boldsymbol{P})$ is equal just to the intersection of the sequential filters of all feasible (or minimizing) sequences for $(\boldsymbol{P})$. Thus, in this countable case, we have an alternative definition of $\mathscr{F}(\boldsymbol{P})$ and $\mathscr{M}(\boldsymbol{P})$, and this paper can be regarded as a study of the collections of all feasible or minimizing sequences.

Remark 1.4. If the filter $C_{\gg}$ has no countable base, $\inf(\boldsymbol{P})$ need not be attained by any feasible sequence: Take a proximity space (Y, δ_Y) and its subset C such that $(Y \setminus C)\, \delta_Y C$ and the intersection of an arbitrary countable family of δ_Y-proximal neighbourhoods of C is again a δ_Y-proximal neighbourhood of C (such situation does exist). Then take $X = Y$, g the identity on X, $f(x) = 0$ for $x \notin C$ and $f(x) = 1$ for $x \in C$. Since $C \gg C$ is not true, $\inf(\boldsymbol{P}) = 0$. On the other hand, if a sequence $\{s_n\}_{n\in N}$ is feasible, s_n must belong to C for all sufficiently large n, hence $\lim_{n\to\infty} f(s_n) =$ $= 1$. Thus we have got an example of a problem that does not possess any minimizing sequence.

2. LIMES INFERIOR OF A NET OF FILTERS

In the classical optimization theory the solutions (minimizers) are understood as points of X. Therefore, to treat stability or approximation of the set of minimizers, a concept of convergence of subsets of X is needed, which is usually introduced by means of some topology. In our tolerance approach we have defined the minimizing filter instead of the set of minimizers, thus we need a concept of convergence of filters on X. It should be emphasized that no topology on X will be employed.

Again we start with a motivation. Consider a filter $\mathscr{A}$ on X having a countable base $\{A_k\}_{k\in N}$ and a sequence $\{\mathscr{A}^i\}_{i\in N}$ of filters on X such that each $\mathscr{A}^i$ has a countable base $\{A^i_k\}_{k\in N}$. If $\mathscr{A}$ and $\mathscr{A}^i$ are interpreted as some principles of optimality in the sense of [4, 5], then "stability from above" by Molodcov [4; Def. 2] can be written in our notation as: $\forall k \in N\ \exists i_k \in N\ \exists n_k \in N\ \forall i \geqq i_k$: $A^i_{n_k} \subset A_k$. It is evident that such definition requires the indices of different bases to be comparable with each other. If we do not suppose it, that means every base has indices of its own, we come to the following condition:

$$\forall k \in N\ \exists i_k \in N\ \forall i \geqq i_k\ \exists n_i \in N\colon A^i_{n_i} \subset A_k\,.$$

Furthermore, if we replace N by a directed index set $(I, \leqq)$ and avoid the assumption concerning the countable bases, we obtain the following simple condition:

$$(2.1)\qquad \forall A \in \mathscr{A}\ \exists i_A \in I\ \forall i \geqq i_A\colon A \in \mathscr{A}^i\,.$$

Definition 2.1. *A filter $\mathscr{A}$ on X is said to be a lower bound for a net $\{\mathscr{A}^i\}_{i\in I}$ of filters on X if* (2.1) *is fulfilled.*

This definition has a simple interpretation: Let $\mathscr{A}$ and $\mathscr{A}^i$ be filters corresponding to an original and an approximate (or perturbed) problems, respectively, and let $A \in \mathscr{A}$ represent the set of approximate solutions (with a certain accuracy) of the original problem. If $\mathscr{A}$ is a lower bound of the net $\{\mathscr{A}^i\}_{i\in I}$, then the elements of A can be obtained by solving approximately the i-th problem with $i \in I$ sufficiently large and also with accuracy sufficiently large (depending on i). Of course, in such a way we can obtain only some elements of A, but it is the usual situation even in the classical optimization: numerical methods yield only some minimizers, not the whole set of them. However, such particular answer is entirely satisfactory for most optimization problems in technical practice.

It should be also recalled that the ordering of filters on subsets of X by inclusion has an opposite character to the ordering of subsets of X by inclusion (e.g. $A_1 \subset A_2$ iff $\mathscr{A}_1 \supset \mathscr{A}_2$ where $\mathscr{A}_i = \{A \in 2^X;\ A \supset A_i\}$), thus the lower bound in Definition 2.1 corresponds freely to an upper bound in the classical concept of optimization theory (cf. also Lemma 5.1 below).

It is evident that if $\mathscr{A}$ is a lower bound for a net of filters, then every filter coarser than $\mathscr{A}$ is a lower bound, too. It encourages us to look for lower bounds that are as fine as possible. It is interesting that, for any net of filters, there exists the finest lower bound:

Proposition 2.1. *Let $\{\mathscr{A}^i\}_{i\in I}$ be a net of filters on X. Then there is exactly one filter $\mathscr{A}_0$ on X such that:*

i) *$\mathscr{A}_0$ is a lower bound for $\{\mathscr{A}^i\}_{i\in I}$,*

ii) *if $\mathscr{A}$ is another lower bound for $\{\mathscr{A}^i\}_{i\in I}$, then $\mathscr{A} \subset \mathscr{A}_0$.*

Proof. It is evident that there can exist at most one filter $\mathscr{A}_0$ satisfying i) and ii), hence it suffices to construct some $\mathscr{A}_0$ that will satisfy i) and ii).

We will show that the set of all lower bounds for the net $\{\mathscr{A}^i\}_{i\in I}$, let us denote it by $\mathfrak{L}$, has got an upper bound, that means for every $n \in N$ and every choice $A_k \in \in \mathscr{L}_k \in \mathfrak{L}$ with $k = 1, \ldots, n$, the intersection $A_1 \cap A_2 \cap \ldots \cap A_n$ is nonempty. As $\mathscr{L}_k \in \mathfrak{L}$, by (2.1) there is $i_{A_k} \in I$ such that $A_k \in \mathscr{A}^i$ for all $i \geqq i_{A_k}$. As $(I, \geqq)$ is a directed set, there exists $j \geqq i_{A_k}$ such that $j \geqq i_{A_k}$ for all $k = 1, \ldots, n$. Then all A_k belong to the filter $\mathscr{A}^j$ and therefore their intersection cannot be empty, since otherwise $\mathscr{A}^j$ would not be a filter.

Take for $\mathscr{A}_0$ the upper bound of $\mathfrak{L}$. As $\mathscr{A} \subset \mathscr{A}_0$ for every $\mathscr{A} \in \mathfrak{L}$, ii) is immediately satisfied. It remains to show that $\mathscr{A}_0 \in \mathfrak{L}$. By the definition of $\mathscr{A}_0$, for every $A \in \mathscr{A}_0$ there is $n \in N$ and $A_k \in \mathscr{L}_k \in \mathfrak{L}$, $k = 1, \ldots, n$, such that $A_1 \cap \ldots \cap A_n \subset A$. Taking j as above we get $A_k \in \mathscr{A}^i$ for all $i \geqq j$ and $k = 1, \ldots, n$, and thus also $A \in \mathscr{A}^i$ because $\mathscr{A}^i$ are filters on X. This shows, in view of (2.1), that $\mathscr{A}_0$ is a lower bound for $\{\mathscr{A}^i\}_{i\in I}$. □

Definition 2.2. *The finest lower bound for a net $\{\mathscr{A}^i\}_{i\in I}$, i.e. the filter $\mathscr{A}_0$ from Proposition 2.1, will be denoted by* $\operatorname{Lim\,inf}_{i\in I} \mathscr{A}^i$.

Corollary 2.1. *A filter $\mathscr{A}$ is a lower bound for a net $\{\mathscr{A}^i\}_{i\in I}$ if and only if $\mathscr{A} \subset \subset$* $\operatorname{Lim\,inf}_{i\in I} \mathscr{A}^i$.

Proposition 2.2. *Let $\{\mathscr{A}^i\}_{i\in I}$ be a net of filters on X, let $\mathscr{A}$ be its lower bound such that $\forall i \in I\ \exists j \geqq i$: $\mathscr{A}^j \subset \mathscr{A}$. Then $\mathscr{A} =$* $\operatorname{Lim\,inf}_{i\in I} \mathscr{A}^i$.

Proof. In view of Corollary 2.1 we are to prove that $\mathscr{A}$ is finer than any lower bound $\mathscr{A}'$ for $\{\mathscr{A}^i\}_{i\in I}$. By (2.1), for any $A \in \mathscr{A}'$ there is $i \in I$ such that $A \in \mathscr{A}^j$ for every $j \geqq i$. Then $A \in \mathscr{A}$ because of the assumption $\mathscr{A}^j \subset \mathscr{A}$ for some $j \geqq i$. As this holds for every $A \in \mathscr{A}'$, we have proved $\mathscr{A}' \subset \mathscr{A}$. □

Definitions 1.2 and 1.3 and Remark 1.1 encourage us to study how the fact $\mathscr{A} \subset \operatorname{Lim\,inf}_{i \in I} \mathscr{A}^i$ is reflected by the sequences whose sequential filters are finer than $\mathscr{A}$ or $\mathscr{A}^i$.

Proposition 2.3. *Let $\{\mathscr{A}^n\}_{n \in N}$ be a sequence of filters on X, let $\mathscr{A} \subset \operatorname{Lim\,inf}_{n \to \infty} \mathscr{A}^n$ be a filter on X with a countable base, and let $s^n = \{s^n_m\}_{m \in N}$ be sequences on X such that $\mathscr{S}(s^n) \supset \mathscr{A}^n$ for all $n \in N$ (recall that $\mathscr{S}(s^n)$ is the corresponding sequential filter). Then there is a function $\mu: N \to N$ such that $\mathscr{S}(s) \supset \mathscr{A}$ for every sequence $s = \{s^n_{m_n}\}_{n \in N}$ with $m_n \geqq \mu(n)$.*

Proof. Let $\{A_i\}_{i \in N}$ be a countable base of $\mathscr{A}$. We may suppose $A_1 = X$ and $A_i \supset A_{i+1}$ for all $i \in N$. By (2.1), $\forall i \in N\ \exists l_i \in N\ \forall n \geqq l_i$: $A_i \in \mathscr{A}^n$. We may suppose $l_1 = 1$ and $l_i < l_{i+1}$ for all $i \in N$. Then, for every $n \in N$, there exists exactly one $j(n) \in N$ such that $l_{j(n)} \leqq n < l_{j(n)+1}$. Obviously, the function $n \mapsto j(n)$ is non-decreasing and $\lim_{n \to \infty} j(n) = +\infty$. Put $B_n = A_{j(n)}$. The collection $\{B_n\}_{n \in N}$ is again a base of $\mathscr{A}$ with $B_n \supset B_{n+1}$ and, in addition, $B_n \in \mathscr{A}^n$ for all $n \in N$. Since $\mathscr{S}(s^n) \supset$ $\supset \mathscr{A}^n$, there is $\mu(n)$ such that $s^n_m \in B_n$ whenever $m \geqq \mu(n)$. Choosing $m_n \geqq \mu(n)$, we get $s^n_{m_n} \in B_n$, hence also $s^n_{m_n} \in B_m$ whenever $n \geqq m$. Since $\{B_m\}_{m \in N}$ is a base of $\mathscr{A}$, we can see that $\mathscr{S}(s) \supset \mathscr{A}$ for $s = \{s^n_m\}_{n \in N}$ with $m_n \geqq \mu(n)$. □

3. STABILITY OF THE INFIMUM AND THE MINIMIZING FILTER

The notions from Definition 1.1 can be reasonable only if they are stable, i.e. if they vary only a little when the data on which they depend vary also little. We will show in this section that the tolerance approach apriori ensures certain stability whenever the perturbations of the data are "compatible" with the tolerance employed. Let us suppose that $(I, \geqq)$ is a directed index set and, for $i \in I$, we are given mappings $f^i: X \to \bar{R}$, $g^i: X \to Y$, and subsets C^i of Y. The proximity δ_Y and thus also the tolerance $\gg$ does not depend on i. Then we consider the following perturbed optimization problem:

$$(\boldsymbol{P}^i)\begin{cases}\text{minimize } f^i(x) \text{ on } X \text{ with tolerance } > \\ \text{subject to } g^i(x) \text{ meets } C^i \text{ with tolerance } \gg .\end{cases}$$

Of course, we define $\mathscr{F}(\boldsymbol{P}^i)$, $\inf(\boldsymbol{P}^i)$ and $\mathscr{M}(\boldsymbol{P}^i)$ again by means of Definition 1.1, supposing the generalized feasibility condition (1.1) to be valid also for $(\boldsymbol{P}^i)$, i.e. $g^i(X)\, \delta_Y C^i$. As we will need a uniform convergence of f^i and g^i, we need some uniform structures on $\bar{R}$ and Y, respectively. Recall that a filter $\mathscr{U}$ on $Z \times Z$ with the properties $\forall W \in \mathscr{U}$: $\Delta \subset W$ and $\exists V \in \mathscr{U}$: $V \circ V \subset W$ is called a semiuniformity on Z; see [6]. If, in addition, $W^{-1} \in \mathscr{U}$ whenever $W \in \mathscr{U}$, $\mathscr{U}$ is called a uniformity on Z. Here $\Delta = \{(z, z); z \in Z\}$ denotes the diagonal of $Z \times Z$, $W^{-1}\{(z_1, z_2); z_2 W z_1\}$ is the inverse relation to the binary relation W, and $V \circ V = \{(z_1, z_2);\ \exists z_3\colon z_1 V z_3, z_3 V z_2\}$ is the composite relation (as V is a binary relation, we use the infix notation, i.e. $z_1 V z_2$ means $(z_1, z_2) \in V$). For a (semi)uniformity on Z and mappings $\varphi, \varphi^i\colon X \to Z$, we say that φ^i converge $\mathscr{U}$-(semi)uniformly to φ iff $\forall V \in \mathscr{U}\ \exists i_0 \in I\ \forall i \geqq i_0\ \forall x \in X\colon \varphi^i(x)\ V\ \varphi(x)$. Any uniformity $\mathscr{U}$ induces a proximity, denoted by $\delta_{\mathscr{U}}$, by declaring $A \delta_{\mathscr{U}} B$ iff $(A \times B) \cap V \neq \emptyset$ for every $V \in \mathscr{U}$. Of course, the topology $\mathscr{T}_{\mathscr{U}}$ induced by $\mathscr{U}$ is defined as the topology induced by the proximity $\delta_{\mathscr{U}}$.

We say that C is a δ_Y-upper bound for a net $\{C^i\}_{i \in I}$ iff $\forall \tilde{C} \gg C\ \exists i \in I\ \forall j \geqq i\colon C^j \subset \tilde{C}$. Note that, if δ_Y is induced by a metric, this fact is nothing else than the upper Hausdorff semicontinuity of the set-valued mapping $i \mapsto C^i$.

The collections $\mathscr{U}^+_{\bar{R}}$ and $\mathscr{U}_{\bar{R}}$ of all $(\mathscr{T}_{\bar{R}} \times \mathscr{T}_{\bar{R}})$-neighbourhoods of the sets $\Delta^+ = \{(a, b) \in \bar{R} \times \bar{R};\ a \geqq b\}$ and $\Delta = \{(a, b) \in \bar{R} \times \bar{R};\ a = b\}$ form a semi-uniformity on $\bar{R}$ and the standard uniformity on $\bar{R}$, respectively ($\mathscr{T}_{\bar{R}}$ is the standard compact topology on $\bar{R}$). Let us suppose that we have got a uniformity $\mathscr{U}_Y$ on Y inducing the given proximity δ_Y. In applications, when (Y, d_Y) is a metric space and $\delta_Y = \delta_{d_Y}$, it is natural to take the uniformity $\mathscr{U}_Y = \{V \in 2^{Y \times Y};\ \exists \varepsilon > 0\colon d_Y(y_1, y_2) \leqq \varepsilon \Rightarrow y_1 V y_2\}$. Now we can impose assumptions on the perturbed data:

(3.1) $\qquad f^i$ converge $\mathscr{U}^+_{\bar{R}}$-semiuniformly to f,

(3.2) $\qquad g^i$ converge $\mathscr{U}_Y$-uniformly to g,

(3.3) $\qquad C$ is a δ_Y-upper bound for the net $\{C^i\}_{i \in I}$.

Theorem 3.1. *If* (3.2) *and* (3.3) *are valid, then*

$$\mathscr{F}(\boldsymbol{P}) \subset \operatorname*{Lim\,inf}_{i \in I} \mathscr{F}(\boldsymbol{P}^i)\,. \tag{3.4}$$

If, in addition, (3.1) *is valid, then*

$$\inf(\boldsymbol{P}) \leqq \liminf_{i \in I} \inf(\boldsymbol{P}^i)\,. \tag{3.5}$$

The proof of this theorem as well as of the following ones is postponed to Sec. 5 where it will be performed by a suitable compactification. To obtain an estimate also for the minimizing filter, we must strengthen the assumptions:

(3.6) $\qquad f^i$ converge $\mathscr{U}_{\bar{R}}$-uniformly to f.

Theorem 3.2. *If* (3.3), (3.6) *are valid,* $C^i \supset C$, *and* $g^i \equiv g$, *then*

$$\mathscr{F}(\boldsymbol{P}) = \operatorname*{Lim\,inf}_{i \in I} \mathscr{F}(\boldsymbol{P}^i)\,, \tag{3.7}$$

$$\inf(\boldsymbol{P}) = \liminf_{i \in I} \inf(\boldsymbol{P}^i)\,, \quad \textit{and} \tag{3.8}$$

$$\mathscr{M}(\boldsymbol{P}) \subset \operatorname*{Lim\,inf}_{i \in I} \mathscr{M}(\boldsymbol{P}^i)\,. \tag{3.9}$$

It should be emphasized that the above stability results hold for arbitrary data (without any continuity requirement for f or g, etc.), which is caused by admitting tolerance in the definition of the problem. This feature has no analogy in the classical optimization theory. On the other hand, to ensure stability with respect to another perturbation of the data (e.g. stability of $\inf(\boldsymbol{P})$ and $\mathscr{M}(\boldsymbol{P})$ when C^i converge to C regarding the Hausdorff uniformity of $\mathscr{U}_Y$) we would have to impose some qualification hypothesis on the data.

Theorems 3.1 and 3.2 together with Proposition 2.3 can be used for minimizing or feasible sequences. Roughly speaking, taking minimizing or feasible sequences for the perturbed problems, we get a minimizing or feasible sequence for the original problem by means of the diagonalization procedure whenever the members of the sequences in question are chosen large enough:

Corrolary 3.1. *Let $I = N$, (3.4) or (3.9) be fulfilled, let the filter $C_{\gg}$ have a countable base and, for all $n \in N$, let $s^n = \{s^n_m\}_{m \in N}$ be a feasible or minimizing sequence for $(\boldsymbol{P}^n)$. Then there is $\mu: N \to N$ such that every sequence $s = \{s^n_{m_n}\}_{n \in N}$ with $m_n \geqq \geqq \mu(n)$ is feasible or minimizing for $(\boldsymbol{P})$, respectively.*

4. NUMERICAL APPROXIMATION OF THE MINIMIZING FILTER

The tolerance approach to optimization problems, being realistic from the technical standpoint and stable as shown in § 3, is moreover closely related with the usual numerical methods. It may offer better understanding how these methods actually work in the general case when the traditional conditions are not fulfilled.

We will confine ourselves to an exterior penalty function method, which is the simplest method how to treat the constraint $g(x) \in C$, though most of the results stated below are preserved also for more advanced methods like the augmented Lagrangean ones. As usual in the penalty technique, we approximate $(\boldsymbol{P})$ by a family of unconstrained problems $(\boldsymbol{P}_r)$ (again considered here with tolerance) with the cost function augmented by a penalty term multiplied by a parameter $r \in R^+ = [0, +\infty[$:

$(\boldsymbol{P}_r)$ $\quad$ minimize $f_r(x) = f(x) + r\, h(g(x))$ on X with tolerance $>$,

where $h: Y \to \bar{R}$ is a penalty function compatible with C and with the proximity δ_Y in the following manner:

(4.1) $\quad h(C) = 0$,

(4.2) $\quad h$ is $(\delta_Y, \delta_{\bar{R}})$-proximally continuous ,

(4.3) $\quad \forall \tilde{C} \gg C \; \exists \varepsilon > 0: h(Y \setminus \tilde{C}) \geqq \varepsilon$.

As $\bar{R}$ is compact, (4.2) is equivalent to the $(\mathscr{U}_Y, \mathscr{U}_{\bar{R}})$-uniform continuity of h provided $\mathscr{U}_Y$ induces δ_Y. Note that if a function with the properties (4.1)–(4.3) does exist, the level sets $\operatorname{lev}_\varepsilon h = h^{-1}([-\infty, \varepsilon])$ with $\varepsilon > 0$ form a base of the filter $C_{\gg}$, hence we must confine ourselves to the case when this filter has a countable base. Nevertheless it does not represent any restriction for most problems arising in technical applications. In a typical case when δ_Y is induced by a metric d_Y, we can obtain a function h satisfying (4.1)–(4.3) if we put $h(y) = q(\hat{d}_Y(\{y\}, C))$, where $q: R \to R$ is an arbitrary continuous, increasing function with $q(0) = 0$.

Furthermore, we assume f to be bounded from below, i.e.

$$(4.4) \qquad \exists M > -\infty \ \forall x \in X: f(x) \geqq M .$$

Of course, we define $\inf(\boldsymbol{P}_r)$ and $\mathscr{M}(\boldsymbol{P}_r)$ again by means of Definition 1.1. As $(\boldsymbol{P}_r)$ is unconstrained (i.e. $\mathscr{F}(\boldsymbol{P}_r) = \{X\}$) we now have simply $\inf(\boldsymbol{P}_r) = \inf f_r(X)$ and $\mathscr{M}(\boldsymbol{P}_r) = \{A \subset X;\ \exists \varepsilon > 0: \mathrm{lev}_{\inf(\boldsymbol{P}_r)+\varepsilon} f_r \subset A\}$.

Theorem 4.1. *Let* (4.1)–(4.4) *be valid and* $\inf(\boldsymbol{P}) \neq +\infty$. *Then the function* $r \mapsto \inf(\boldsymbol{P}_r)$ *is nondecreasing, and*

$$(4.5) \qquad \inf(\boldsymbol{P}) = \lim_{r\to\infty} \inf(\boldsymbol{P}_r) ,$$

$$(4.6) \qquad \mathscr{M}(\boldsymbol{P}) \subset \operatorname*{Lim\,inf}_{r\to\infty} \mathscr{M}(\boldsymbol{P}_r) .$$

For the proof see again § 5.

There are simple examples showing that generally $\mathscr{M}(\boldsymbol{P}) \neq \operatorname{Lim\,inf}_{r\to\infty} \mathscr{M}(\boldsymbol{P}_r)$. However, the penalty-function approximation of the tolerance constrained optimization problem is so natural that, when used carefully, this method can even yield precisely the minimizing filter. Put $\mathscr{L}_{r,a} = \{A \in 2^X;\ \mathrm{lev}_a f_r \subset A\}$. Clearly, $\mathscr{L}_{r,a}$ is a filter on X if $a > \inf(\boldsymbol{P}_r)$ (particularly if $a > \inf(\boldsymbol{P})$), and $\mathscr{M}(\boldsymbol{P}_r) = \bigcup\{\mathscr{L}_{r,a};\ a > \inf(\boldsymbol{P}_r)\}$.

Theorem 4.2. *Let* (4.1)–(4.4) *be valid*, $\inf(\boldsymbol{P}) \neq +\infty$, *and let* $a: R \to R$ *be an arbitrary decreasing function such that* $\lim_{r\to\infty} a(r) = \inf(\boldsymbol{P})$. *Then*

$$(4.7) \qquad \mathscr{M}(\boldsymbol{P}) = \operatorname*{Lim\,inf}_{r\to\infty} \mathscr{L}_{r,a(r)} .$$

It should be pointed out that the penalized problem $(\boldsymbol{P}_r)$ can be handled by digital computers only when X is a subset of a finite-dimensional linear space. If it is not, we must perform further approximation: instead of X, which is infinite-dimensional, we take some (finite-dimensional) subsets X^k of X, $k \in N$, such that

$$(4.8) \qquad X^{k_1} \subset X^{k_2} \subset X \quad \text{whenever} \quad k_1 \leqq k_2 .$$

Besides, to ensure convergence we must suppose some data qualification, namely: there exists a topology $\mathscr{T}_X$ on X such that

(4.9) $\bigcap_{k\in N} X^k$ is $\mathscr{T}_X$-dense in X, and

(4.10) f and g are $(\mathscr{T}_X, \mathscr{T}_R)$- and $(\mathscr{T}_X, \mathscr{T}_Y)$-continuous, respectively, where $\mathscr{T}_{\bar{R}}$ is the standard compact topology on $\bar{R}$ and $\mathscr{T}_Y$ is the topology induced by δ_Y. It is clear that (4.9) requires $\mathscr{T}_X$ to be coarse enough, while (4.10) conversely requires $\mathscr{T}_X$ to be fine enough, therefore existence of a suitable topology $\mathscr{T}_X$ may be understood actually as a certain data qualification.

Usually we have to approximate also the cost function f and the mapping g, say by methods like numerical quadrature, finite-difference or finite-element methods, etc. Thus we get some mappings $f^k: X^k \to \bar{R}$ and $g^k: X^k \to Y$. For simplicity, we suppose that the penalty function h is simple enough to be evaluated exactly, which is a frequent case indeed. We will assume the following approximative property ($\mathscr{U}_{\bar{R}}$ and $\mathscr{U}_Y$ are the uniformities already used in (3.2) and (3.6)):

(4.11) $\qquad \forall V \in \mathscr{U}_{\bar{R}}\ \exists k_0 \in N\ \forall k \geqq k_0\ \forall x \in X^k: f^k(x)\ V f(x)\,,$

(4.12) $\qquad \forall V \in \mathscr{U}_Y\ \exists k_0 \in N\ \forall k \geqq k_0\ \forall x \in X^k: g^k(x)\ V g(x)\,.$

Now we can approximate the penalized problem $(\boldsymbol{P}_r)$ by the problem:

$(\boldsymbol{P}_r^k)$ $\qquad$ minimize $f_r^k(x) = f^k(x) + r\,h(g^k(x))$ on X^k with tolerance $>$.

We could also define this problem as: "Minimize $f_r^k(x)$ on X with tolerance $>$ subject to x meets X^k without tolerance", which would require formally to define f^k and g^k on the whole space X, however. Thus we have preferably defined $(\boldsymbol{P}_r^k)$ as done above, which causes, on the other hand, that $\mathscr{M}(\boldsymbol{P}_r^k)$ is a filter not on X (if $X \neq X^k$) but on X^k, and thus we are forced to modify it by introducing $\mathscr{M}_X(\boldsymbol{P}_r^k)$ as the filter on X generated by the base $\mathscr{M}(\boldsymbol{P}_r^k)$.

Theorem 4.3. *Let* (4.1)–(4.4), (4.8)–(4.12) *be fulfilled and* $\inf(\boldsymbol{P}) \neq +\infty$. *Then there exists a function* $\varkappa: R^+ \to N$ *such that*

(4.13) $$\inf(\boldsymbol{P}) = \lim_{k\to\infty, r\to\infty, k \geqq \varkappa(r)} \inf(\boldsymbol{P}_r^k)\,, \quad \textit{and}$$

(4.14) $$\mathscr{M}(\boldsymbol{P}) \subset \operatorname*{Lim\,inf}_{k\to\infty, r\to\infty, k\geqq\varkappa(r)} \mathscr{M}_X(\boldsymbol{P}_r^k)\,.$$

We observe that the convergence is ensured only under a "stability condition" $k \geqq \varkappa(r)$. In other words, k must approach infinity sufficiently quickly in comparison with r. We can easily construct examples where the convergence is violated when only $k \to \infty$ and $r \to \infty$. However, in concrete problems the choice of the function $\varkappa$ may require fine knowledge of the properties of the data (cf. Example 4.1 below), and therefore it is surely useful to state an additional data qualification that guarantees the unconditional convergence: Let us suppose that there exist a proximity δ_X on X (then $\mathscr{T}_X$ in (4.9) and (4.10) is induced by δ_X), a set $C_0 \subset Y$, and a point $x_0 \in X$ such that:

(4.15) $\qquad f, f^k$ and g, g^k are $(\delta_X, \delta_{\bar{R}})$- and (δ_X, δ_Y)-proximally continuous, respectively,

(4.16) $\qquad C_0$ is $\mathscr{T}_Y$-open and $g(x_0) \in C_0 \subset C \subset \mathrm{cl}_Y C_0\,,$

(4.17) $\qquad \forall B \gg g^{-1}(C_0)\ \exists \tilde{C} \gg C_0: g^{-1}(\tilde{C}) \subset B\,,$

where $\mathrm{cl}_Y C_0$ means the $\mathscr{T}_Y$-closure of C_0 in Y. Let us remark that (4.16) is ensured if Y is a linear topological space, C is a convex set with a nonempty interior $\mathrm{int}_Y C$ and $g(x_0) \in \mathrm{int}_Y C$ for some $x_0 \in X$ (then (4.16) is satisfied with $C_0 = \mathrm{int}_Y C$). Such condition appears very often in classical optimization theory where it is called the Slater constraint qualification. Similarly as (1.3), the condition (4.17) is guaranteed if g^{-1} is a singlevalued (δ_Y, δ_X)-proximally continuous mapping, or, in the case that δ_X and δ_Y are induced by some metrics, if the (multivalued) mapping g^{-1} is uniformly Hausdorff continuous.

Theorem 4.4. *Let* (4.1)–(4.4), (4.8)–(4.12), (4.15)–(4.17) *be fulfilled, and let* $\inf(\boldsymbol{P}) \neq +\infty$. *Then*

$$\inf(\boldsymbol{P}) = \lim_{k\to\infty, r\to\infty} \inf(\boldsymbol{P}_r^k)\,, \quad \textit{and} \tag{4.18}$$

$$\mathscr{M}(\boldsymbol{P}) \subset \operatorname*{Lim\,inf}_{k\to\infty, r\to\infty} \mathscr{M}_X(\boldsymbol{P}_r^k)\,. \tag{4.19}$$

Note that (4.15)–(4.17) imply (1.2) and (1.3), hence Theorem 4.4 cannot be used in case $\inf(\boldsymbol{P}) < \inf f(g^{-1}(C))$.

Example 4.1. From the proof of Theorem 4.3 in § 5 it is evident that, if one knows an estimate of the discretization error:

$$|\inf(\boldsymbol{P}_r^k) - \inf(\boldsymbol{P}_r)| \leqq \varepsilon(r, k_0) \quad \text{for all} \quad k \geqq k_0\,,$$

then for $\varkappa$ from Theorem 4.3 one can take arbitrary $\varkappa: N \to N$ such that $\lim_{r\to\infty} \varepsilon(r, \varkappa(r)) = 0$. This error estimate can be typically obtained as follows: let X and Y be normed linear spaces, let $\|\cdot\|_X$ and $\|\cdot\|_Y$ be their respective norms, and let f, g, and h be Hölder continuous, i.e. for every $x_1, x_2 \in X$, $y_1, y_2 \in Y$

$$\begin{aligned} |f(x_1) - f(x_2)| &\leqq L\|x_1 - x_2\|_X^{\alpha_f}\,, \\ \|g(x_1) - g(x_2)\|_Y &\leqq L\|x_1 - x_2\|_X^{\alpha_g}\,, \\ |h(y_1) - h(y_2)| &\leqq L\|y_1 - y_2\|_Y^{\alpha_h}\,, \end{aligned}$$

hold for some positive constants L and α's. Moreover, let the following rate-of-error estimates be known:

$$|f(x) - f^k(x)| \leqq L\,.\,k^{-\beta_f}\,, \quad \|g(x) - g^k(x)\|_Y \leqq L\,.\,k^{-\beta_g}$$

for every $x \in X^k$ and some positive β's, and finally let

$$\sup_{x\in X} \inf_{\tilde{x}\in X^k} \|x - \tilde{x}\|_X \leqq L\,.\,k^{-\gamma}\,, \quad \gamma > 0\,. \tag{4.20}$$

The following situation can serve as an example for (4.20) (the notation is standard): Let Ω be a polygonal domain in R^n, X a bounded subset of the Sobolev space $W^{1,2}(\Omega)$, but let X be endowed by the norm $\|\cdot\|_X$ of the space $L^2(\Omega)$. Let $X^k \subset X$ be spaces of the finite-element type, constructed, say, by means of linear triangular elements ($1/k$ is the mesh parametr). Then (4.20) with $\gamma = 1$ follows by the well-known inequality $\inf_{\tilde{x} \in X^k} \|x - \tilde{x}\|_X \leqq (1/k) \|x\|_{W^{1,2}(\Omega)}$.

Now the following estimate can be easily verified:

$$\inf(\boldsymbol{P}_r) - L_0(k^{-\beta_f} + r \cdot k^{-\alpha_h \beta_g}) \leqq \inf(\boldsymbol{P}_r^k) \leqq$$

$$\leqq \inf(\boldsymbol{P}_r) + L_0(k^{-\gamma \alpha_f} + k^{-\beta_f} + r \cdot (k^{-\gamma \alpha_h \alpha_g} + k^{-\alpha_h \beta_g}))$$

with some $L_0 > 0$, and we can obviously put $\varkappa(r) = r^\nu$ with $\nu > \max\{1/(\gamma \alpha_h \alpha_g), 1/(\alpha_h \beta_g)\}$. Note that neither α_f nor β_f occur in this estimate of ν.

5. COMPACTIFICATIONS OF THE PROBLEM ($\boldsymbol{P}$)

The above introduced features suggest that the optimization problems resulting from admitting tolerance behave like classical problems without tolerance but on a compact space X. In this section we explain this fact by constructing a "closure" of the problem ($\boldsymbol{P}$), which yields, in addition, simple and elegant proofs of the above stated assertions. It may also help to obtain further results for problems posed with tolerance because behaviour of the compactified problems introduced below gives a good hint for behaviour of the original problem with tolerance.

Let $(\bar{X}, \mathscr{T}_{\bar{X}})$ be a topological space, $(\bar{Y}, \mathscr{U}_{\bar{Y}})$ a uniform space, $\bar{C} \subset \bar{Y}$, $\bar{f}: \bar{X} \to \bar{R}$, and $\bar{g}: \bar{X} \to \bar{Y}$. We consider the constrained minimization problem in the classical sense, i.e. without tolerance:

$$(\bar{\boldsymbol{P}}) \begin{cases} \text{minimize} & \bar{f}(x) \quad \text{on} \quad \bar{X} \\ \text{subject to} & \bar{g}(x) \in \bar{C}\,. \end{cases}$$

We will use the following assumptions:

(5.1) $\mathscr{T}_{\bar{X}}$ is compact and X is $\mathscr{T}_{\bar{X}}$-dense in $\bar{X}$,

(5.2) $Y \subset \bar{Y}$ and the trace on Y of $\mathscr{U}_{\bar{Y}}$ is a uniformity $\mathscr{U}_Y$ inducing the given proximity δ_Y,

(5.3) $\bar{C}$ is the closure of C in $(\bar{Y}, \mathscr{U}_{\bar{Y}})$,

(5.4) $\bar{f}$ is l.s.c. (lower semicontinuous), $\bar{f}(x) = f(x)$ for all $x \in X$, and $\bar{f}(x) \geqq \geqq \liminf_{\tilde{x} \to x, \tilde{x} \in X} f(x)$ for every $x \in \bar{X}$ (of course, $\tilde{x} \to x$ stands for the convergence in the topology $\mathscr{T}_{\bar{X}}$),

(5.5) $\bar{g}$ is continuous and $\bar{g}(x) = g(x)$ for all $x \in X$,

(5.6) $\bar{f}$ is continuous.

Clearly, if $\bar{f}(x) = f(x)$ for $x \in X$, then (5.6) implies (5.4).

Definition 5.1. *The problem* $(\bar{P})$ *is said to be a compactification of the problem* (P) *if* (5.1)–(5.5) *are valid. If also* (5.6) *is valid, the compactification is called regular.*

The minimum of any compactification $(\bar{P})$ is obviously attained, and we may define $\min(\bar{P}) = \min \bar{f}(\bar{g}^{-1}(\bar{C}))$ and $\operatorname{Argmin}(\bar{P}) = \{x \in \bar{X}; \bar{f}(x) = \min(\bar{P}), \bar{g}(x) \in \bar{C}\}$. First we show that, as for the ability to study the problem (P), all (regular) compactifications (P) are equivalent to one another.

Proposition 5.1. *Let* $(\bar{P})$ *be a compactification of* (P). *Then* $\min(\bar{P}) = \inf(P)$.

Proof. Denote by $\mathscr{U}_{\bar{X}}$ the unique uniformity on $\bar{X}$ inducing the compact topology $\mathscr{T}_{\bar{X}}$, and by $\mathscr{U}_X$ and $\mathscr{T}_X$ the trace on X of $\mathscr{U}_{\bar{X}}$ and $\mathscr{T}_{\bar{X}}$, respectively. Then $f\colon X \to \bar{R}$ is $(\mathscr{T}_X, \mathscr{T}_{\bar{R}})$-l.s.c. and g is $(\mathscr{U}_X, \mathscr{U}_Y)$-uniformly continuous. Then we may use the arguments of the proof of Theorem 2 in [10] to show that the collection $\{\operatorname{lev}_{\alpha+\varepsilon} f \cap g^{-1}(V(C));\ \varepsilon > 0,\ V \in \mathscr{U}_Y\}$ is a filter base on X if and only if $\alpha \geqq \min(\bar{P})$, where $V(C) = \{x \in X;\ \exists x_0 \in C\colon xVx_0\}$ with $V \in \mathscr{U}_Y$ is a $\mathscr{U}_Y$-uniform neighbourhood of C. Since $\mathscr{U}_Y$ induces δ_Y, the $\mathscr{U}_Y$-uniform neighbourhoods of C coincide with the δ_Y-proximal ones. In view of Definition 1.1 we get $\min(\bar{P}) = \inf(P)$. □

Let us denote by $\mathscr{N}(S)$ the collection of all $\mathscr{T}_{\bar{X}}$-neighbourhoods of a set $S \subset \bar{X}$ (the dependence on $\mathscr{T}_{\bar{X}}$ will not be explicitly indicated). If S is nonempty, then $\mathscr{N}(S)$ is a filter on $\bar{X}$, and, since X is $\mathscr{T}_{\bar{X}}$-dense in $\bar{X}$, the trace on X of $\mathscr{N}(S)$, i.e. the collection $\mathscr{N}(S)|_X = \{A \cap X;\ A \in \mathscr{N}(S)\}$, is a filter on X.

Proposition 5.2. *Let* $(\bar{P})$ *be a compactification of* (P). *Then* $\mathscr{N}(\bar{g}^{-1}(\bar{C}))|_X$ *is the feasible filter* $\mathscr{F}(P)$. *If the compactification* $(\bar{P})$ *is regular, then* $\mathscr{N}(\operatorname{Arg\,min}(\bar{P}))|_X$ *is the minimizing filter* $\mathscr{M}(P)$.

Proof. First we prove that $\bar{g}^{-1}(\bar{C})$ is nonempty. Suppose the contrary, i.e. $\bar{g}(\bar{X}) \cap \bar{C} = \emptyset$. Since $\bar{g}$ is continuous and $\bar{X}$ compact, $\bar{g}(\bar{X})$ is compact, too. Since $\bar{C}$ is closed, $\bar{g}(\bar{X})$ and $\bar{C}$ are $\delta_{\bar{Y}}$-far from each other (see [1; (5.3.24)]), where $\delta_{\bar{Y}}$ is the proximity on $\bar{Y}$ induced by $\mathscr{U}_{\bar{Y}}$. Since the trace on Y of $\delta_{\bar{Y}}$ is just δ_Y, $g(X)$ and C are δ_Y-far from each other, which contradicts (1.1).

As the assertion concerning the feasible filter can be obtained from the assertion concerning the minimizing filter if $f \equiv +\infty$, we will prove only the latter one. It is now evident that, in view of (1.1), (5.1)–(5.5), $\operatorname{Arg\,min}(\bar{P})$ is nonempty. Taking again $\mathscr{U}_X$ as the trace on X of the only uniformity on $\bar{X}$ inducing the compact topology $\mathscr{T}_{\bar{X}}$, we may use the arguments of the proof of Theorem 3 in [10] together with the fact that $\mathscr{U}_Y$ induces δ_Y to show that $\mathscr{N}(\operatorname{Arg\,min}(\bar{P}))|_X$ coincides with the filter $\mathscr{A} = \{V(A);\ V \in \mathscr{U}_X,\ A \in \mathscr{M}(P)\}$. By definition, $\mathscr{A}$ is coarser than $\mathscr{M}(P)$. Exploiting the regularity of the compactification used, hence the $(\mathscr{U}_X, \mathscr{U}_{\bar{R}})$-uniform continuity of f, we will show that $\mathscr{A}$ is finer than $\mathscr{M}(P)$. Let $A \in \mathscr{M}(P)$, i.e. $A \supset f^{-1}([-\infty, a]) \cap g^{-1}(\tilde{C})$ for some $\tilde{C} \gg C$ and $[-\infty, a] > [-\infty, \inf(P)]$. Thanks to the properties of the tolerances $>$ and $\gg$, we can take $B = f^{-1}([-\infty, b]) \cap g^{-1}(\hat{C})$ with some b

and $\hat{C}$ such that $\tilde{C} \gg \hat{C} \gg C$ and $[-\infty, a] > [-\infty, b] > [-\infty, \inf(P)]$. Obviously $B \in \mathcal{M}(\boldsymbol{P})$. Because of the $(\mathscr{U}_X, \mathscr{U}_{\bar{R}})$- and $(\mathscr{U}_X, \mathscr{U}_Y)$-uniform continuity of f and g, respectively, there exists $V \in \mathscr{U}_X$ such that $V(B) \subset A$, thus $A \in \mathscr{A}$. □

Of course, if we desire to use the compactifications for proofs it is necessary to guarantee the existence of at least one compactification. For this reason we must recall what the Smirnov compactification of a proximity space (Z, δ) is; see [12] or also e.g. [1, 7]. It is quite evident that the filter base on $Z \times Z$

$$\{\bigcup_{i=1}^{n} A_i \times A_i;\ A_i \subset Z,\ n \in N,\ \forall i \leqq n\ \exists B_i\colon A_i \gg B_i \text{ and } \bigcup_{i=1}^{n} B_i = Z\}$$

(with $\gg$ related to δ) generates a uniformity on Z, let us denote it by $\mathscr{U}_\delta$, which is precompact (i.e. $\forall V \in \mathscr{U}_\delta\ \exists$a finite set $A \subset Z$: $V(A) = Z$) and induces the proximity δ. Moreover, it is the only precompact uniformity on X inducing δ, which justifies the notation $\mathscr{U}_\delta$. As $\mathscr{U}_\delta$ is precompact, the completion of the uniform space $(Z, \mathscr{U}_\delta)$ is compact and it is called the Smirnov compactification of the proximity space (Z, δ), or briefly δ-compactification of Z.

Proposition 5.3. *Every problem $(\boldsymbol{P})$ admits at least one regular compactification $(\bar{\boldsymbol{P}})$.*

Proof. Let δ_X be the discrete proximity on X, i.e. $A\delta_X B$ only if $A \cap B \neq \emptyset$. As δ_X is the finest proximity on X, f and g are $(\delta_X, \delta_{\bar{R}})$- and (δ_X, δ_Y)-proximally continuous, respectively. Therefore there exist (even unique) continuous extensions $\bar{f}\colon \bar{X} \to \bar{R}$ and $\bar{g}\colon \bar{X} \to \bar{Y}$ of f and g, respectively, where $\bar{X}$ and $\bar{Y}$ are the δ_X- and δ_Y-compactifications of X and Y, respectively; note that f and g are also $(\mathscr{U}_{\delta_X}, \mathscr{U}_{\delta_{\bar{R}}})$- and $(\mathscr{U}_{\delta_X}, \mathscr{U}_{\delta_Y})$-uniformly continuous, respectively, and $\bar{f}$ and $\bar{g}$ are nothing else than their continuous extensions to the corresponding completions, (cf. [1; (6.2.11)]). Of course, $\bar{R}$, being compact in its standard proximity $\delta_{\bar{R}}$, coincides with its Smirnov compactification, which is important to preserve the standard ordering of $\bar{R}$. Finally, take the closure of C in $\bar{Y}$ for $\bar{C}$. One of the regular compactifications of $(\boldsymbol{P})$ has been just constructed. □

Remark 5.1. The compactification $(\boldsymbol{P})$ used for the above proof is the "largest" one in the sense that it uses the finest compactification $\bar{X}$ of X ($\bar{X}$ can be identified there with the set of all ultrafilters on X). This compactification does not require any continuity properties of f and g. If f and g do satisfy some continuity requirements, "smaller" compactifications can be admitted. For instance, if f and g are continuous with respect to some completely regular topology on X, then there is a regular compactification $(\bar{\boldsymbol{P}})$ for which $\bar{X}$ is the well-known Stone-Čech compactification of the completely regular topological space X. It is even the "smallest" regular compactification, in general. If f and g are uniformly continuous with respect to some uniformity on X, then there is even a smaller compactification $(\bar{\boldsymbol{P}})$. It uses for $\overline{X}$ the Smirnov compactification of X regarding the proximity induced by the uniformity considered, and it is again generally the smallest possible regular compactification under these uniform continuity requirements. Such compactification has been used in [10]. Of course, if f and g are continuous with respect to a compact topology on X, then $(\boldsymbol{P})$ admits the absolutely smallest compactification, using $\overline{X} = X$.

Although for some purposes a particular choice of a compactification of $(\boldsymbol{P})$ is not important (and thus we can use universally the largest compactification from the proof of Proposition 5.3), sometimes we need smaller compactification because some properties of the data cannot be transferred to compactifications which are too large; cf. the proof of Theorem 4.4 below. Small compactifications are also advantageous to study neccesary and sufficient conditions of optimality; for the unconstrained case we refer to [9] where the Smirnov compactification of a normed linear space together with the Ekeland ε-variational principle has been used.

Now we go on to the proofs of the results stated in §§ 3, 4 by exploiting the compactification. The idea is very simple: first transfer the properties of the data from the orignal problem to its compactification, then exploit good behaviour of the compactified problem by standard techniques (only we must realize that the extended spaces need not satisfy the first countability axiom), and afterwards return to the original problem. As for the infimum, the return to the original problem is straightforward thanks to Proposition 5.1, while for the minimizing or the feasible filter we need the following assertion:

Lemma 5.1. *Let $(\overline{X}, \mathscr{T}_{\overline{X}})$ be a compact space, X a $\mathscr{T}_{\overline{X}}$-dense subset of $\overline{X}$, let $A \subset \overline{X}$ be $\mathscr{T}_{\overline{X}}$-closed and nonempty, and let $\{A^i\}_{i\in I}$ be a net of nonempty subsets of $\overline{X}$. Then*

$$A \supset \operatorname*{Lim\,sup}_{i\in I} A^i \Rightarrow \mathcal{N}(A)|_X \subset \operatorname*{Lim\,inf}_{i\in I} (\mathcal{N}(A^i)|_X)\,.$$

If, in addition, $A^j \supset \operatorname*{Lim\,sup}_{i\in I} A^i$ for all $j \in I$, then

$$\mathcal{N}(\operatorname*{Lim\,sup}_{i\in I} A^i)|_X = \operatorname*{Lim\,inf}_{i\in I} (\mathcal{N}(A^i)|_X)\,,$$

where $\operatorname{Lim\,sup}_{i\in I} A^i$ *has the usual meaning, i.e. it consists of all $\mathscr{T}_{\overline{X}}$-cluster points of all nets $\{x^i\}_{i\in I}$ with $x^i \in A^i$.*

Proof. Take $S \in \mathcal{N}(A)|_X$, hence $S = \bar{S} \cap X$ for some $\bar{S} \in \mathcal{N}(A)$. Since A is compact, the set $\overline{X} \setminus \bar{S}$ and A are disconnected, and thus there is $B \in \mathcal{N}(A)$ such that $\bar{S} \in \mathcal{N}(B)$; in other words, $\bar{S} \gg B \gg A$ with $\gg$ related with the only proximity inducing the compact topology $\mathscr{T}_{\overline{X}}$. In addition, we may and will suppose B to be open, hence $\overline{X} \setminus B$ compact. We show that $A^i \subset B$ for sufficiently large $i \in I$. Suppose the contrary. Then for every $i \in I$ there would exist $j(i) \geqq i$ and $x^{j(i)} \in A^{j(i)} \setminus B$.

Thanks to the compactness, the net $\{x^{j(i)}\}_{i\in I}$ would have a cluster point $x^\infty \in \bar{X} \setminus B$; therefore $x^\infty \notin A$, but simultaneously $x^\infty \in \operatorname{Lim\,sup}_{i\in I} A^i \subset A$, which is a contradiction. Thus $A^i \subset B$ for $i \in I$ large enough, hence $S \in \mathcal{N}(A^i)|_X$. Since S has been taken arbitrarily, we have proved that $\mathcal{N}(A)|_X$ is a lower bound of the net $\{\mathcal{N}(A^i)|_X\}_{i\in I}$ due to Definition 2.1. If $A = \operatorname{Lim\,sup}_{i\in I} A^i \subset A^j$, then $\mathcal{N}(A)|_X \supset \supset \mathcal{N}(A^j)|_X$, and the second assertion to be proved follows immediately from Proposition 2.2. □

Now we will prove the assertions from §§ 3, 4. To transfer the properties (3.1), (3.2), and (3.6) to the extended functions, we need still the following lemma:

Lemma 5.2. *Let $(\bar{X}, \mathcal{T}_{\bar{X}})$ be a topological space, $(\bar{Y}, \mathcal{U}_{\bar{Y}})$ be a (semi)uniform space, X a $\mathcal{T}_{\bar{X}}$-dense subset of $\bar{X}$, $\bar{g}, \bar{g}^i: \bar{X} \to \bar{Y}$ continuous mappings, $i \in I$, and let $\bar{g}^i|_X$ converge $\mathcal{U}_{\bar{Y}}$-(semi)uniformly to $\bar{g}|_X$, where $\bar{g}^i|_X$ and $\bar{g}|_X$ are the restrictions to X of $\bar{g}_i$ and $\bar{g}$, respectively. Then $\bar{g}^i$ converge $\mathcal{U}_{\bar{Y}}$-(semi)uniformly to $\bar{g}$.*

Proof. The assertion follows from the facts that the closed elements V from $\mathcal{U}_{\bar{Y}}$ form a base of $\mathcal{U}_{\bar{Y}}$, and, if $\bar{g}^i(x)\, V \bar{g}(x)$ for all $x \in X$ and V is $\mathcal{T}_{\bar{Y}} \times \mathcal{T}_{\bar{Y}}$-closed, then $\bar{g}^i(x)\, V \bar{g}(x)$ for all $x \in \bar{X}$ because the mapping $x \mapsto (\bar{g}^i(x), \bar{g}(x))$ from $\bar{X}$ to $\bar{Y} \times \bar{Y}$ is continuous. □

Proof of Theorem 3.1. We take such compactifications $(\bar{\boldsymbol{P}})$ and $(\bar{\boldsymbol{P}}^i)$ of $(\boldsymbol{P})$ and $(\boldsymbol{P}^i)$, respectively, that use the common spaces $(\bar{X}, \mathcal{T}_{\bar{X}})$ and $(\bar{Y}, \mathcal{U}_{\bar{Y}})$, and, moreover, the trace on Y of the uniformity $\mathcal{U}_{\bar{Y}}$ is coarser than the uniformity $\mathcal{U}_Y$ from (3.2). For simplicity we may suppose them to be regular. Such compactifications do exist, cf. the construction via the discrete proximity on X used in the proof of Proposition 5.3 (note that the uniformity $\mathcal{U}_{\delta_Y}$, being the coarsest uniformity inducing δ_Y, is surely coarser than $\mathcal{U}_Y$. Then (3.2) implies the $\mathcal{U}_Y$-uniform convergence of $\bar{g}^i|_X$ to $\bar{g}|_X$, and by Lemma 5.2 we can see that

$$\bar{g}^i \text{ converge } \mathcal{U}_{\bar{Y}}\text{-uniformly to } \bar{g}\,. \tag{5.7}$$

From (3.1) we get again by Lemma 5.2 that

$$\bar{f}^i \text{ converge } \mathcal{U}_{\bar{R}}^+\text{-semiuniformly to } \bar{f}\,. \tag{5.8}$$

Now we will prove that

$$\forall V \in \mathcal{U}_{\bar{Y}}\ \exists i_0 \in I\ \forall i \geqq i_0:\ \bar{C}^i \subset V(\bar{C})\,. \tag{5.9}$$

Take some $W \in \mathcal{U}_{\bar{Y}}$ such that $W \circ W \subset V$, and put $D = W(\bar{C}) \cap Y$. Since the trace on Y of $\mathcal{U}_{\bar{Y}}$ induces δ_Y we have $D \gg C$, and then $C^i \subset D$ for all $i \in I$ large enough because of (3.3). Then $\bar{C}^i \subset \operatorname{cl}_{\bar{Y}} D \subset V(\bar{C})$.

Now we will prove that $\bar{g}^{-1}(\bar{C}) \supset \operatorname{Lim\,sup}_{i\in I} (\bar{g}^i)^{-1} (\bar{C}^i)$. Take $x^i \in (\bar{g}^i)^{-1} (\bar{C}^i)$ and a $\mathcal{T}_{\bar{X}}$-cluster point x^∞ of the net $\{x^i\}_{i\in I}$. In view of (5.7) and the continuity of $\bar{g}$ we can see that $\bar{g}(x^\infty)$ is a $\mathcal{T}_{\bar{Y}}$-cluster point of $\{\bar{g}(x^i)\}_{i\in I}$, and by (5.9) we have $\bar{g}(x^\infty) \in \bar{C}$, from which (3.4) follows by Lemma 5.1 and by Proposition 5.2.

As (3.5) is trivial if $\min(\bar{\boldsymbol{P}}) = -\infty$, we suppose $\min(\bar{\boldsymbol{P}}) > -\infty$. First we treat the case $\min(\bar{\boldsymbol{P}}) \neq +\infty$. Suppose, for a moment, that $\liminf_{i\in I} \min(\bar{\boldsymbol{P}}^i) < \min(\bar{\boldsymbol{P}})$. Then there are $\varepsilon > 0$ and a cofinal subset J of I such that $\min(\bar{\boldsymbol{P}}^i) \leqq \min(\bar{\boldsymbol{P}}) - \varepsilon$ for all $i \in J$. Take $x^i \in \operatorname{Arg\,min}(\bar{\boldsymbol{P}}^i)$, and a cluster point x^∞ of the net $\{x^i\}_{i\in J}$. We have clearly $\bar{f}^i(x^i) \leqq \min(\bar{\boldsymbol{P}}) - \varepsilon$, and by (5.8) and the continuity of $\bar{f}$, also $\bar{f}(x^\infty) \leqq$ $\leqq \min(\bar{\boldsymbol{P}}) - \varepsilon$. On the other hand, we have already shown that $\bar{g}(x^\infty) \in \bar{C}$. Therefore $\bar{f}(x^\infty) \geqq \min(\bar{\boldsymbol{P}})$, a contradiction. Thus $\liminf_{i\in I} \min(\bar{\boldsymbol{P}}^i) \geqq \min(\bar{\boldsymbol{P}})$, from which (3.5) follows by Proposition 5.1. In the case $\min(\bar{\boldsymbol{P}}) = +\infty$ we obtain a contradiction analogously, supposing $\min(\bar{\boldsymbol{P}}^i) \leqq 1/\varepsilon$ for some $\varepsilon > 0$ and all $i \in J$. □

Proof of Theorem 3.2. Take the regular compactifications $(\bar{\boldsymbol{P}})$ and $(\bar{\boldsymbol{P}}^i)$ as in the previous proof. As $\bar{C}^i \supset \bar{C}$ and $\bar{g}^i = \bar{g}$, we have $(\bar{g}^i)^{-1}(\bar{C}^i) \supset g^{-1}(\bar{C})$, which implies (3.7) again by Lemma 5.1 and Proposition 5.2. By (3.6) and Lemma 4.2 we get that

(5.10) $\quad \bar{f}^i$ converge $\mathscr{U}_{\bar{R}}$-uniformly to $\bar{f}$.

Take some $x \in \operatorname{Arg\,min}(\bar{\boldsymbol{P}})$. Then $\bar{g}(x) \in C^i$ for all $i \in I$, and for every $\varepsilon > 0$ we have $\bar{f}^i(x) \leqq \bar{f}(x) + \varepsilon$ provided i is sufficiently large and $\min(\bar{\boldsymbol{P}}^i) \neq -\infty$. Therefore $\min(\bar{\boldsymbol{P}}^i) \leqq \min(\bar{\boldsymbol{P}}) + \varepsilon$, which shows that $\limsup_{i\in I} \min(\bar{\boldsymbol{P}}^i) \leqq \min(\bar{\boldsymbol{P}})$. Since $\liminf_{i\in I} \min(\bar{\boldsymbol{P}}^i) \geqq \min(\bar{\boldsymbol{P}})$, which has been already proved above, we obtain (3.8) again by Proposition 5.1. In case $\min(\bar{\boldsymbol{P}}) = -\infty$ we get from (5.10) that $\min(\bar{\boldsymbol{P}}^i) \leqq -1/\varepsilon$, and (3.8) follows analogously.

We will show that

(5.11) $$\operatorname{Arg\,min}(\bar{\boldsymbol{P}}) \supset \operatorname*{Lim\,sup}_{i\in I} \operatorname{Arg\,min}(\bar{\boldsymbol{P}}^i).$$

Take $x^i \in \operatorname{Arg\,min}(\bar{\boldsymbol{P}}^i)$ and a $\mathscr{T}_{\bar{X}}$-cluster point x^∞ of the net $\{x^i\}_{i\in I}$. Since $\bar{g}^i(x^i) =$ $= \bar{g}(x^i) \in \bar{C}^i$, due to (5.9) and the continuity of $\bar{g}$ we again obtain $\bar{g}(x^\infty) \in \bar{C}$. Since $\bar{f}^i(x^i) = \min(\bar{\boldsymbol{P}}^i)$, by (5.10) and the continuity of $\bar{f}$ we can see that $\bar{f}(x^\infty)$ is a cluster point of $\{\bar{f}^i(x^i)\}_{i\in I}$. Therefore $\bar{f}(x^\infty) = \min(\bar{\boldsymbol{P}})$ because $\lim_{i\in I} \min(\bar{\boldsymbol{P}}^i) = \min(\bar{\boldsymbol{P}})$. In other words, $x^\infty \in \operatorname{Arg\,min}(\bar{\boldsymbol{P}})$, and (5.11) has been proved. Then (3.9) follows from (5.11) again by Lemma 5.1 and Proposition 5.2. □

Proof of Theorem 4.1. Let us take a regular compactification $(\bar{\boldsymbol{P}})$ of $(\boldsymbol{P})$ and denote by δ_X the trace on X of the (unique) proximity on $\bar{X}$ inducing the compact topology $\mathscr{T}_{\bar{X}}$. Then f and g are $(\delta_X, \delta_{\bar{R}})$- and (δ_X, δ_Y)-proximally continuous, respectively. Thus $f_r = f + r \, . \, h \circ g$ is $(\delta_X, \delta_{\bar{R}})$-proximally continuous because of (4.2) and the proximal continuity of the binary operation "extended addition" $+ : [M, +\infty] \times$ $\times [0, +\infty] \to \bar{R}$; cf. (4.4) and observe that $h \geqq 0$ as a consequence of (4.1)–(4.3). Therefore we can extend f_r to $\bar{f}_r : \bar{X} \to \bar{R}$ by continuity, obtaining the problem:

$(\bar{\boldsymbol{P}}_r)$ $\quad$ minimize $\bar{f}_r(x)$ on $\bar{X}$,

which is a regular compactification of $(\boldsymbol{P}_r)$ in the sense of Definition 5.1. Thanks to (4.2), $h : Y \to \bar{R}$ is $(\mathscr{U}_Y, \mathscr{U}_{\bar{R}})$-uniformly continuous (note that $\mathscr{U}_{\bar{R}}$ is the coarsest uniformity on $\bar{R}$ inducing $\delta_{\bar{R}}$), and we can extend h to $\bar{h}: \bar{Y} \to \bar{R}$ by continuity. As a consequence of (4.1)–(4.3) we have $\bar{h}(\bar{C}) = 0$ and $\bar{h}(\bar{Y} \setminus \bar{C}) > 0$. Since the function $\bar{f} + r . \bar{h} \circ \bar{g}: \bar{X} \to \bar{R}$ is continuous and coincides with $\bar{f}_r$ on the $\mathscr{T}_{\bar{X}}$-dense subset X, we can see that $\bar{f}_r = \bar{f} + r . \bar{h} \circ \bar{g}$.

Since $\bar{h} \geqq 0$, the function $r \mapsto \min(\bar{\boldsymbol{P}}_r)$ is nondecreasing, and therefore the limit $L = \lim_{r\to\infty} \min(\bar{\boldsymbol{P}}_r)$ does exist. Obviously, $\bar{f}_r(x) = \min(\bar{\boldsymbol{P}})$ for every $x \in \operatorname{Arg\,min}(\bar{\boldsymbol{P}})$ (which is nonempty), thus $\min(\bar{\boldsymbol{P}}_r) \leqq \min(\bar{\boldsymbol{P}})$ and $L \leqq \min(\bar{\boldsymbol{P}})$. We take $x_r \in \operatorname{Arg\,min}(\bar{\boldsymbol{P}}_r)$ and a $\mathscr{T}_{\bar{X}}$-cluster point x_∞ of the net $\{x_r\}_{r\in R^+}$. By the assumptions $\inf(\boldsymbol{P}) \neq +\infty$ and $M > -\infty$ (see (4.4)) we can estimate: $\bar{h}(\bar{g}(x_r)) \leqq (L - M)/r \leqq (\inf(\boldsymbol{P}) - M)/r = \mathcal{O}(1/r)$ for $r \to +\infty$. Thus $\bar{h}(\bar{g}(x_\infty)) = 0$, and therefore $\bar{g}(x_\infty) \in \bar{C}$ and $\bar{f}(x_\infty) \geqq \min(\bar{\boldsymbol{P}})$. Moreover, $\bar{f}(x_\infty)$ is a $\mathscr{T}_{\bar{R}}$-cluster point of $\{\bar{f}(x_r)\}_{r\in R^+}$ and $\bar{f}(x_r) \leqq \bar{f}_r(x_r) = \min(\bar{\boldsymbol{P}}_r) \leqq L \leqq \min(\bar{\boldsymbol{P}})$, which gives $L = \min(\bar{\boldsymbol{P}})$. By Proposition 5.1 we get (4.5). It is also clear that $x_\infty \in \operatorname{Arg\,min}(\bar{\boldsymbol{P}})$, hence we have demonstrated that $\operatorname{Arg\,min}(\bar{\boldsymbol{P}}) \supset \operatorname{Lim\,sup}_{r\to\infty} \operatorname{Arg\,min}(\bar{\boldsymbol{P}}_r)$, from which (4.6) follows immediately by Lemma 5.1 and Proposition 5.2. □

Proof of Theorem 4.2. Take again a regular compactification $(\bar{\boldsymbol{P}})$ of $(\boldsymbol{P})$. As $\operatorname{lev}_{a(r)} \bar{f}_r \cap X = \operatorname{lev}_{a(r)} f_r$, the filter $\mathscr{L}_{r,a(r)}$ is just the trace on X of the filter $\{A \in 2^{\bar{X}};\ A \supset \operatorname{lev}_{a(r)} \bar{f}_r\}$. We show that $\operatorname{Arg\,min}(\bar{\boldsymbol{P}}) \supset \operatorname{Lim\,sup}_{r\to\infty} \operatorname{lev}_{a(r)} \bar{f}_r$. Take $x_r \in \operatorname{lev}_{a(r)} \bar{f}_r$ and a $\mathscr{T}_{\bar{X}}$-cluster point x_∞ of the net $\{x_r\}_{r\in R^+}$. By (4.4) and the assumption $\inf(\boldsymbol{P}) \neq +\infty$ we obtain the estimate $\bar{h}(\bar{g}(x_r)) \leqq (a(r) - M)/r = \mathcal{O}(1/r)$ for $r \to +\infty$, hence $\bar{h}(\bar{g}(x_\infty)) = 0$ and thus $\bar{g}(x_\infty) \in \bar{C}$. Moreover, $\bar{f}(x_r) \leqq \bar{f}_r(x_r) \leqq a(r)$, which implies $\bar{f}(x_\infty) \leqq \min(\bar{\boldsymbol{P}})$ because $a(r) \searrow \min(\bar{\boldsymbol{P}})$ and $\bar{f}$ is continuous. Thus $x_\infty \in \operatorname{Arg\,min}(\bar{\boldsymbol{P}})$. By Lemma 5.1 we obtain $\mathscr{M}(\boldsymbol{P}) \subset \operatorname{Lim\,inf}_{r\to\infty} \mathscr{L}_{r,a(r)}$.

On the other hand, for $a > \min(\bar{\boldsymbol{P}})$ we have $\operatorname{lev}_a \bar{f}_r \in \mathscr{N}(\operatorname{Arg\,min}(\bar{\boldsymbol{P}}))$ because $\operatorname{lev}_a \bar{f}_r \in \mathscr{N}(\operatorname{lev}_{\min(\bar{\boldsymbol{P}})} \bar{f}_r)$, which is a consequence of the continuity of $\bar{f}_r$, and $\operatorname{Arg\,min}(\bar{\boldsymbol{P}}) \subset \operatorname{lev}_{\min(\bar{\boldsymbol{P}})} \bar{f}_r$. It implies $\mathscr{L}_{r,a(r)} \subset \mathscr{M}(\boldsymbol{P})$, and, by Proposition 2.2, also (4.7). □

Proof of Theorem 4.3. Take a regular compactification $(\bar{\boldsymbol{P}})$ of $(\boldsymbol{P})$ such that f^k and g^k are proximally continuous with respect to the trace on X^k of the proximity of the compact space $(\bar{X}, \mathscr{T}_{\bar{X}})$; such $(\bar{\boldsymbol{P}})$ does exist, the compactification used in the proof of Proposition 5.3 can serve as an example for it. Furthermore, take the closure of X^k in $\bar{X}$ for $\bar{X}^k$. We can extend f^k and g^k continuously to $\bar{X}^k$, denoting the extensions respectively by $\bar{f}^k: \bar{X}^k \to \bar{R}$ and $\bar{g}^k: \bar{X}^k \to \bar{Y}$, and consider the problem

$$(\bar{\boldsymbol{P}}_r^k) \qquad \text{minimize} \quad \bar{f}_r^k(x) = \bar{f}^k(x) + r . \bar{h}(\bar{g}^k(x)) \quad \text{on} \quad \bar{X}^k .$$

By the way, $(\bar{\boldsymbol{P}}_r^k)$ is a regular compactification of $(\boldsymbol{P}_r^k)$ according to Definition 5.1.

Now we will show (4.13). Suppose, for a moment, that r is fixed. Take some $x_r^k \in \operatorname{Arg\,min}(\bar{\boldsymbol{P}}_r^k)$ and a cluster point x_r^∞ of the net $\{x_r^k\}_{k\in N}$. Thanks to (4.10) and (4.2), $\operatorname{lev}_{\inf(\boldsymbol{P}_r)+\varepsilon} f_r$ is a $\mathscr{T}_X$-open nonempty subset of X for every $\varepsilon > 0$. By (4.9) with (4.8), $X^k \cap \operatorname{lev}_{\inf(\boldsymbol{P}_r)+\varepsilon} f_r$ is nonempty provided k is large enough. In other words, there is $x_\varepsilon \in X^k$ (for k large enough) such that $f_r(x_\varepsilon) \leqq \inf(\boldsymbol{P}_r) + \varepsilon$. The condition (4.12) remains valid if $\mathscr{U}_Y$ is replaced by the (unique) precompact uniformity $\mathscr{U}_Y^*$ inducing δ_Y because $\mathscr{U}_Y^* \subset \mathscr{U}_Y$. Due to (4.2), h is $(\mathscr{U}_Y^*, \mathscr{U}_{\bar{R}})$-uniformly continuous, and by (4.11) and (4.12) we can see that f_r^k converge for $k \to \infty$ to f_r $\mathscr{U}_{\bar{R}}$-uniformly on every X^k (recall that r is fixed). In particular, $f_r^k(x) - f_r(x)$ converge for $k \to \infty$ to zero $\mathscr{U}_R$-uniformly on $\{x; |f_r(x)| \leqq L \text{ or } \forall k\colon |f_r^k(x)| \leqq L\}$ where $L \neq +\infty$ (the bound L had to be introduced due to the fact that, thanks to the points $-\infty$ and $+\infty$, $\mathscr{U}_{\bar{R}}$ restricted to R is strictly coarser than the standard additive uniformity $\mathscr{U}_R$ on R). We obviously have the following apriori estimates: $M \leqq f_r(x_\varepsilon) \leqq \min(\bar{\boldsymbol{P}}) + \varepsilon$ (for M see (4.4)) and $M - 1 \leqq \bar{f}_r^k(x_r^k) \leqq f_r^k(x_\varepsilon) \leqq \min(\bar{\boldsymbol{P}}) + \varepsilon + 1$ (we have used also (4.11), (4.12) and $\min(\bar{\boldsymbol{P}}_r) \leqq \min(\bar{\boldsymbol{P}})$) provided k is large enough. Take $L = \max(|M|, |\min(\bar{\boldsymbol{P}}) + \varepsilon|) + 1$. Then from these estimates we obtain respectively $f_r^k(x_\varepsilon) - f_r(x_\varepsilon) \to 0$ and $\bar{f}_r^k(x_r^k) - \bar{f}_r(x_r^k) \to 0$ for $k \to \infty$. Taking into account also the estimates $f_r(x_\varepsilon) \leqq \inf(\boldsymbol{P}_r) + \varepsilon$ and $\bar{f}_r^k(x_r^k) \leqq f_r^k(x_\varepsilon)$ stated above and the continuity of $\bar{f}_r$, we get eventually $\bar{f}_r(x_r^\infty) \leqq \min(\bar{\boldsymbol{P}}_r) + \varepsilon$. As $\varepsilon > 0$ is arbitrary, $\bar{f}_r(x_r^\infty) = \min(\bar{\boldsymbol{P}}_r)$. As this holds for every cluster point x_r^∞ of $\{x_r^k\}_{k\in N}$, there exists the limit $\lim_{k\to\infty} \min(\bar{\boldsymbol{P}}_r^k)$ and equals $\min(\bar{\boldsymbol{P}}_r)$. In particular, $\forall r\ \exists \varkappa(r)\ \forall k \geqq \varkappa(r)$: $\min(\bar{\boldsymbol{P}}_r) - 1/r \leqq \min(\bar{\boldsymbol{P}}_r^k) \leqq \min(\bar{\boldsymbol{P}}_r) + 1/r$. By Theorem 4.1 and Proposition 5.1 we then obtain (4.13).

Now we want to prove (4.14). First we prove

$$(5.12) \qquad \operatorname{Arg\,min}(\bar{\boldsymbol{P}}) \supset \operatorname*{Lim\,sup}_{k\to\infty, r\to\infty, k\geqq\varkappa(r)} \operatorname{Arg\,min}(\bar{\boldsymbol{P}}_r^k).$$

Take $x_r^k \in \operatorname{Arg\,min}(\bar{\boldsymbol{P}}_r^k)$ and a $\mathscr{T}_{\bar{X}}$-cluster point x_∞ of the net $\{x_r^k\}_{k\in N, r\in R^+, k\geqq\varkappa(r)}$. We know an apriori estimate: $M - 1 \leqq \bar{f}^k(x_r^k) \leqq \min(\bar{\boldsymbol{P}}) + 1$ for k and r large enough, hence (4.11) implies similarly as above that $\bar{f}^k(x_r^k) - \bar{f}(x_r^k) \to 0$ for $k \to \infty$. Since $\bar{f}^k(x_r^k) \leqq \min(\bar{\boldsymbol{P}}_r^k) \to \min(\bar{\boldsymbol{P}})$, we get $\limsup_{k\to\infty, r\to\infty, k\geqq\varkappa(r)} \bar{f}(x_r^k) \leqq \min(\bar{\boldsymbol{P}})$, and therefore $\bar{f}(x_\infty) \leqq \min(\bar{\boldsymbol{P}})$ thanks to the continuity of $\bar{f}$. Similarly, from the apriori estimate $0 \leqq \bar{h}(\bar{g}^k(x_r^k)) \leqq (\min(\bar{\boldsymbol{P}}) - M + 1)/r$ we obtain $\bar{h}(\bar{g}(x_\infty)) = 0$. Therefore we get $x_\infty \in \operatorname{Arg\,min}(\bar{\boldsymbol{P}})$, hence (5.12) is proved. By compactness, for any $\mathscr{T}_{\bar{X}}$-neighbourhood A of $\operatorname{Arg\,min}(\bar{\boldsymbol{P}}_r^k)$ there is $a > \min(\bar{\boldsymbol{P}}_r^k)$ such that $A \supset \operatorname{lev}_a \bar{f}_r^k$, hence also $A \cap X \supset \operatorname{lev}_a f_r^k$, which shows that the trace on X of the filter $\mathscr{N}(\operatorname{Arg\,min}(\bar{\boldsymbol{P}}_r^k))$ is coarser than $\mathscr{M}_X(\boldsymbol{P}_r^k)$. Then (4.14) follows from (5.12) by Proposition 5.2 and Lemma 5.1. □

Proof of Theorem 4.4. Thanks to (4.15) there exists the regular compactification $(\bar{\boldsymbol{P}})$ of $(\boldsymbol{P})$ such that $\bar{X}$ is the δ_X-compactification of X (δ_X is the proximity used in (4.15) and (4.17)). Moreover, we can define $(\bar{\boldsymbol{P}}_r^k)$ as in the proof of Theorem 4.3.

Put $\bar{C}_0 = \operatorname{int}_{\bar{Y}} \bar{C}$, where $\operatorname{int}_{\bar{Y}}$ denotes the interior in the topology induced by $\bar{\mathscr{U}}_Y$. From (4.16) and the fact that $\mathscr{U}_Y$ induces on Y the topology $\mathscr{T}_Y$ we can easily show that $\bar{C} \subset \operatorname{cl}_{\bar{Y}} \bar{C}_0$. As g is continuous, the set $A_0 = \bar{g}^{-1}(\bar{C}_0)$ is open in $\bar{X}$. Besides, we will show that $A_0 \subset \bar{g}^{-1}(\bar{C}) \subset \operatorname{cl}_{\bar{X}} A_0$. The first inclusion is evident. To verify the second one, suppose that there is $x \in \bar{g}^{-1}(\bar{C}) \setminus \operatorname{cl}_{\bar{X}} A_0$. Then there are $A_1 \gg A_0$ and $A_2 \gg \{x\}$ with $A_1 \cap A_2 = \emptyset$, where $\gg$ is related with the unique proximity of the compact space $\bar{X}$ (which is just the prolongation of δ_X). As $\bar{g}(x) \in \bar{C}$ and $\bar{g}$ is continuous, $A_2 \cap g^{-1}(\tilde{C})$ is nonempty for every $\tilde{C} \gg C$. On the other hand, we can put $B = A_1 \cap X$ into (4.17), thus obtaining $\tilde{C} \gg C$ such that $g^{-1}(\tilde{C}) \subset B \subset A_1$, which contradicts $A_1 \cap A_2 = \emptyset$, however. Therefore we have proved $\bar{g}^{-1}(\bar{C}) \subset \subset \mathrm{cl}_{\bar{X}} A_0$.

In the proof of Theorem 4.3 we have shown in particular that $\lim\inf_{k\to\infty} \min(\bar{\boldsymbol{P}}_r^k) \geqq \geqq \min(\bar{\boldsymbol{P}}_r)$, and it is quite clear that it holds uniformly with respect to $r \in R^+$. In view of (4.5) we then obtain $\lim\inf_{k\to\infty, r\to\infty} \min(\bar{\boldsymbol{P}}_r^k) \geqq \min(\bar{\boldsymbol{P}})$. Thus we have to show that $\min(\bar{\boldsymbol{P}}) \geqq \lim\sup_{k\to\infty, r\to\infty} \min(\bar{\boldsymbol{P}}_r^k)$. Take $\varepsilon > 0$ and put $A_\varepsilon = \{x \in X;\ f(x) < \min(\boldsymbol{P}) + \varepsilon\}$. We will prove that $\bar{g}^{-1}(\hat{C}) \cap A_\varepsilon \neq \emptyset$ for some $\mathscr{T}_{\bar{Y}}$-open $\hat{C}$ such that $\bar{C}_0 \gg \hat{C}$. As $\bar{C}_0$ is open, the union of all open $\hat{C}$ with $\bar{C}_0 \gg \hat{C}$ is just $\bar{C}_0$, and therefore the union of $\bar{g}^{-1}(\hat{C})$ is just A_0. Due to (4.16), A_0 is nonempty because $x_0 \in A_0$. Also $A_0 \cap A_\varepsilon$ is nonempty because any $x \in \operatorname{Arg\,min}(\bar{\boldsymbol{P}})$ belongs simultaneously to the interior of A_ε and to the closure of A_0 (since we have already proved $\bar{g}^{-1}(\bar{C}) \subset \mathrm{cl}_{\bar{X}} A_0$). Therefore there is an open $\hat{C} = \hat{C}_\varepsilon$ such that $\bar{C}_0 \gg \hat{C}_\varepsilon$ and $\bar{g}^{-1}(\hat{C}_\varepsilon) \cap A_\varepsilon = \hat{A}_\varepsilon$ is nonempty. The set $\hat{A}_\varepsilon$ is open due to the continuity of $\bar{f}$ and $\bar{g}$. As the trace on X of $\mathscr{T}_{\bar{X}}$ is just the topology $\mathscr{T}_X$ used in (4.9), the union $\bigcup_{k\in N} X^k$ is $\mathscr{T}_{\bar{X}}$-dense in $\bar{X}$. By (4.8) there is $k_\varepsilon \in N$ such that X^k intersects $\hat{A}_\varepsilon$ whenever $k \geqq k_\varepsilon$. In other words, for $k \geqq k_\varepsilon$ there exists $x \in X^k$ such that $f(x) \leqq \inf(\boldsymbol{P}) + \varepsilon$ and $g(x) \in \hat{C}_\varepsilon \gg \bar{C}$. Moreover, due to (4.12), we have $g^k(x) \in C$ provided k_ε is chosen sufficiently large, because the proximities induced on Y by $\mathscr{U}_{\bar{Y}}$ and $\mathscr{U}_Y$ are the same, namely δ_Y. Taking k_ε large enough, by (4.11) we obtain $f^k(x) \leqq \inf(\boldsymbol{P}) + 2\varepsilon$. Thus $f_r^k(x) = f^k(x) \leqq \inf(\boldsymbol{P}) + 2\varepsilon$, and we can see that $\inf(\boldsymbol{P}_r^k) \leqq \inf(\boldsymbol{P}) + 2\varepsilon$. As $\varepsilon > 0$ is arbitrary, we have proved $\lim\sup_{k\to\infty, r\to\infty} \min(\bar{\boldsymbol{P}}_r^k) \leqq \min(\bar{\boldsymbol{P}}_r)$. Hence (4.18) is proved.

The assertion (4.19) can be proved analogously as (4.14), but using (5.12) without the stability condition $k \geqq \varkappa(r)$. □

6. MISCELLANEOUS REMARKS AND EXAMPLES

Remark 6.1. Although the general structure of $(\boldsymbol{P})$ may cover all constrained optimization problems, it is worth noticing how the tolerance approach can be applied to problems with a more concrete structure because some specific phenomena can appear there. Let us consider a minimization problem with a collection of functional constraints:

$$(\hat{\boldsymbol{P}})\ \begin{cases} \text{minimize } f(x) \text{ with tolerance } > \\ \text{subject to } g_i(x) \text{ meets } C_i \text{ with tolerance } \gg_i,\ i \in I\,, \end{cases}$$

where $f: X \to \bar{R}$, $g_i: X \to Y_i$, $C_i \subset Y_i$, I is an index set and $\gg_i$ are the tolerances corresponding respectively to some given proximities δ_{Y_i} on Y_i. In view of Definition 1.1, this problem is equivalent to the problem $(\boldsymbol{P})$ in the sense $\mathscr{F}(\boldsymbol{P}) = \mathscr{F}(\hat{\boldsymbol{P}})$,

$\inf(\boldsymbol{P}) = \inf(\hat{\boldsymbol{P}})$, $\mathscr{M}(\boldsymbol{P}) = \mathscr{M}(\hat{\boldsymbol{P}})$ provided one takes the data for $(\boldsymbol{P})$ as follows: $Y = \prod_{i\in I} Y_i$, $C = \prod_{i\in I} C_i$, $g = (g_i)_{i\in I}$, and $\delta_Y = \prod_{i\in I} \delta_{Y_i}$ (by the standard definition of the product of proximities, δ_Y is the coarsest proximity on Y that makes all canonical projections $Y \to Y_i$ proximally continuous). If I is not countable we can get an example for δ_Y non-metrizable (except trivial cases such that Y_i are singletons etc.).

Suppose δ_{Y_i} are induced by some metrics d_{Y_i} and I is countable, say $I = N$. It is then natural to endow also Y by some metric, e.g. by $d_Y = \sum_{i\in N} 2^{-i} d_{Y_i}/(1 + d_{Y_i})$. Yet, it is known (see [1; (7.3.39)]) that, even if I were finite, d_Y would not induce the proximity δ_Y defined above provided at least two of the metric spaces (Y_i, d_{Y_i}), $i \in I$, are not precompact. Nevertheless, we can use the proximity induced by d_Y, although in general it is strictly finer than δ_Y, with the same effect as δ_Y because both the proximities generate the same proximal neighbourhoods of the set C, which is caused by the fact that C is not a general subset of Y but has got the special form of the product $\prod_{i\in I} C_i$.

Remark 6.2. Let us mention the situation that occurs in optimal control problems. Let X, Y, Z be the sets of controls, observations and states, respectively, let $f_0\colon X \times Z \to \bar{R}$ be a cost function, $A\colon X \to Z$ a state operator, and $g_0\colon X \times Z \to Y$ an observation mapping. Furthermore, the observation space Y is endowed by a proximity δ_Y ($\gg$ will again denote the corresponding tolerance on Y). We will write briefly "$z = A(x)$" instead of "(x, z) meets the graph of A without tolerance", and consider the following optimal control problem with tolerance:

$$(\boldsymbol{P}_0)\begin{cases}\text{minimize } f_0(x, z) \text{ on } X \times Z \text{ with tolerance } > .\\ \text{subject to } z = A(x) \text{ and}\\ \qquad\qquad g_0(x, z) \text{ meets } C \text{ with tolerance } \gg .\end{cases}$$

Such a formulation of the optimal control problem is in harmony with the very realistic approach of J. Warga [13; Sec. III.1] who distinguishes, on the one hand, the "absolute" constraint formed by the state equation $z = A(x)$ which is supposed to be governed by the laws of nature and should be fulfilled exactly (i.e. without tolerance) because otherwise we would move "out of the world", and, on the other hand, the "desired" constraint $g_0(x, z) \in C$ given by some technical or engineering requirements which may be satisfied only with a certain accuracy (in our notation, with tolerance). The optimal control problem is effectively treated after a transformation into the problem $(\boldsymbol{P})$, representing then a mathematical programming problem on the space of controls X, by means of the substitution

$$(6.1) \qquad f(x) = f_0(x, A(x)) \quad \text{and} \quad g(x) = g_0(x, A(x)) .$$

It is not difficult to see that, if f and g defined by (6.1) are taken for $(\boldsymbol{P})$, this mathematical-programming transformation actually leads to the problem $(\boldsymbol{P})$ which is equivalent to the original optimal control problem $(\boldsymbol{P}_0)$ in the following sense:

$\mathscr{F}(\boldsymbol{P}) = \mathrm{Pr}_X \mathscr{F}(\boldsymbol{P}_0)$, $\inf(\boldsymbol{P}) = \inf(\boldsymbol{P}_0)$, and $\mathscr{M}(\boldsymbol{P}) = \mathrm{Pr}_X \mathscr{M}(\boldsymbol{P}_0)$, where $\mathrm{Pr}_X: X \times Z \to X$ denotes the canonical projector.

Nevertheless, the problem $(\boldsymbol{P}_0)$ possesses certain specific features. Although the state operator should be treated without tolerance, sometimes it is sensible to consider an approximation A^i of A, which can arise from numerical evaluation of the state operator, or from errors in coefficients appearing in the state equation which may be obtained, say, by some measurements, etc. Then the above stated results will be preserved under the following data qualification: there exists a uniformity $\mathscr{U}_Z$ on Z such that A^i converge $\mathscr{U}_Z$-uniformly to A and simultaneously $f_0(x, \cdot): Z \to \bar{R}$ are $(\mathscr{U}_Z, \mathscr{U}_{\bar{R}})$-uniformly equi-continuous with respect to $x \in X$ (i.e. $\forall V \in \mathscr{U}_{\bar{R}}\ \exists W \in \mathscr{U}_Z\ \forall x \in X\ \forall z_1, z_2 \in Z: z_1 W z_2 \Rightarrow f_0(x, z_1)\ V f_0(x, z_2)$), and similarly $g_0(x, \cdot): Z \to Y$ are $(\mathscr{U}_Z, \mathscr{U}_Y)$-uniformly equi-continuous. From these assumptions we immediately obtain (3.6) and (3.2) for f and g given by (6.1) and $f^i = f_0 \circ A^i$, $g^i = g_0 \circ A^i$. We may conclude that the tolerance admitted for the constraint and the cost function allows us to admit some "tolerance" also for the state operator, originally considered without tolerance.

Remark 6.3. Some of the conditions used above may be sometimes too strong, e.g. (3.2) or (4.12) provided X and X^k are unbounded subsets of a normed linear space. However, often it is possible to employ a certain coercivity of the problem using some concept of boundedness of the subsets of X, and to weaken the mentioned conditions by restricting them to bounded subsets only. We will not deal with this idea in detail because it is rather standard.

Example 6.1. Though in applications the tolerances will be mostly metrizable, it is worth giving a simple and quite natural example in which the tolerance need not be metrizable. Combining the problems from Remarks 6.1 and 6.2, we will consider the state-constrained optimal control problem for a dynamical system:

$$(\hat{\boldsymbol{P}}_0) \begin{cases} \text{minimize } f_0(x, z) \text{ on } X \times Z \text{ with tolerance } > \\ \text{subject to } \mathrm{d}z/\mathrm{d}t = F(x(t), z(t), t) \text{ for a.a. } t \in [0, T]\,, \\ \qquad z \text{ absolutely continuous, } z(0) = z_0\,, \\ \qquad z(t) \text{ meets } C \text{ with tolerance } \gg \text{ for all } t \in [0, T)\,, \\ \qquad z(T) \text{ meets } \{z_d\} \text{ with tolerance } \gg\,, \end{cases}$$

where $X = \{x: [0, T] \to B \text{ measurable}\}$, B is a bounded subset of R^m, $Z = (R^n)^{[0,T]}$, $F: R^m \times R^n \times [0, T] \to R^n$ determines the dynamics of the controlled system (we will suppose $F(x, z, \cdot)$ measurable, $F(\cdot, \cdot, t)$ Lipschitz continuous, and $|F(x, z, t)| \leqq \text{const.}\,(1 + |z|)$), $C \subset R^n$, $z_0 \in R^n$, $T > 0$ is a finite time horizon, z_d is a desired final state, and $\gg$ is the Euclidean tolerance on R^n (i.e. the tolerance corresponding to the proximity induced by the Euclidean metric). By Remark 6.1, the collection of the state constraints is equivalent to one constraint "z meets C_0 with tolerance $\gg_0$"

where $C_0 = C^{[0,T)} \times \{z_d\}$ and $\gg_0$ is the tolerance on $(R^n)^{[0,T]}$ corresponding to the proximity obtained by the product of the Euclidean proximities on R^n parametrized by $t \in [0, T]$. As $[0, T]$ is not countable, this proximity is not metrizable.

However, the non-metrizable tolerance is rather formal here. Realizing that admissible trajectories z are equi-Lipschitz continuous (note that F has at most linear growth in z independently of x and t, and B and $[0, T]$ are bounded), we can easily see that, if restricted to the subset of Z containing admissible trajectories only, the tolerance $\gg_0$ yields the same proximal neighbourhoods of C as the tolerance induced e.g. by the Chebyshev metric $d_Z(z_1, z_2) = \sup_{0 \leqq t \leqq T} |z_1(t) - z_2(t)|$.

Although we can use equivalently the metrizable tolerance in this particular example, from the viewpoint of numerical solution the non-metrizable tolerance $\gg_0$ seems to be more advantageous: we need not check the constraint within the whole trajectory, but only at a finite number of time levels (not prescribed in advance, however).

Example 6.2. It may be said that the well-known relaxed-control theory (see J. Warga [13]) can serve as a very concrete example of compactification of optimal control problems. If we confine ourselves to the preceding example, the set of controls X is then imbedded in a natural way into the space $(L^1(0, T; C^0(\bar{B})))^*$ by assigning to x a linear continuous functional on $L^1(0, T; C^0(\bar{B}))$ defined by $\varphi \mapsto \int_0^T \varphi(t, x(t))\, dt$, where $\bar{B}$ denotes the closure of B in R^n, $C^0(\cdot)$ the space of all continuous functions, $L^1(0, T; \cdot)$ the space of all Bochner integrable functions on $[0, T]$, and the star denotes the topological dual. Then X is precompact in the uniformity related with the weak-star topology of $(L^1(0, T; C^0(\bar{B})))^*$ and, under some additional assumptions on F, the state operator $x \mapsto z$ from Example 6.1 is uniformly continuous when taking the uniformity on the space of states Z coarse enough, say that induced by the Chebyshev metric d_Z from the Example 6.1. Moreover, this weak-star uniformity is metrizable on X and we can obtain a compactification simply by forming the completion of X with respect to this metric. The elements of $\bar{X}$, called relaxed controls, can be then identified with the functions on $[0, T]$ whose values are random measures on $\bar{B}$, i.e. positive Borel measures μ on $\bar{B}$ such that $\mu(\bar{B}) = 1$. Moreover, this compactification is generally the coarsest (i.e. smallest) one. On the other hand, the cases when the compactification can be constructed as a metric completion and the elements of the compactified sets can be indentified in a similar manner as it was done for the relaxed controls are rather exceptional and in general the compactified spaces will not be metrizable (their elements being called generalized solutions in the author's former works [9–11]).

Acknowledgements. The author would like to express his gratitude to Professors Pavol Brunovský and Ľubica Holá for many valuable comments that made the paper more comprehensible to readers.

References

[1] *Á. Császár:* General Topology. Akademiai Kiadó, Budapest, 1978.

[2] *A. V. Efremovich:* The geometry of proximity (in Russian). Mat. Sbornik 31 (73) (1952), 189—200.

[3] *E. K. Golshtein:* Duality Theory in Mathematical Programming and Its Applications (in Russian). Nauka, Moscow, 1971.

[4] *D. A. Molodcov:* Stability and regularization of principles of optimality (in Russian). Zurnal vycisl. mat. i mat. fiziki 20 (1980), 1117—1129.

[5] *D. A. Molodcov:* Stability of Principles of Optimality (in Russian). Nauka, Moscow, 1987.

[6] *L. Nachbin:* Topology and Order. D. van Nostrand Comp., Princeton, 1965.

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

[8] *E. Polak, Y. Y. Wardi:* A study of minizing sequences. SIAM J. Control Optim. 22 (1984), 599—609.

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

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

[11] *T. Roubíček:* Stable extensions of constrained optimization problems. J. Math. Anal. Appl. 141 (1989), 120—135.

[12] *Yu. M. Smirnov:* On proximity spaces (in Russian). Mat. Sbornik 31 (73) (1952), 534—574.

[13] *J. Warga:* Optimal Control of Differential and Functional Equations. Academic press, New York, 1972.

Souhrn

OPTIMALIZACE S OMEZENÍMI: OBECNĚ TOLERANČNÍ PŘÍSTUP

TOMÁŠ ROUBÍČEK

Pro překonání poněkud umělých těžkostí v klasické teorii optimalizace, týkajících se existence a stability řešení, se navrhuje nové pojetí optimalizačních úloh s omezeními (nazvanými úlohami s tolerancí) za použití daných proximitních struktur pro zadání okolí množin. Infimum a takzvaný minimalizující filtr se potom definují pomocí úrovňových množin indukovaných těmito okolími, což také odráží inženýrské chápání optimalizačních úloh s omezeními. Navíc je rozvinut odpovídající koncept konvergence filtrů, a dokázána stabilita minimalizujícího filtru jakož i jeho aproximace technikou vnější pokutové funkce použitím kompaktifikace úlohy.

Резюме

УСЛОВНАЯ ОПТИМИЗАЦИЯ: ОБЩИЙ ТОЛЕРАНТНОСТНЫЙ ПОДХОД

TOMÁŠ ROUBÍČEK

Для преодоления несколько искуственных трудностей в классической теории оптимизации, касающихся существования и устойчивости решений, предлагается новая постановка проблем условной оптимизации (названных здесь проблемами с толерантностью), использующая структуры близости для определения окрестностей множеств. Нижняя грань и так-называемый минимизирующий фильтр определяются затем посредством множеств уровня порожденныж этими окресностями, что тоже отражает инженерное понимание проблем условной оптимизации. Далее развивается подходящая концепция сходимости фильтров и при помощи компактификации проблемы доказывается устойчивость минимизирующего фильтра и его приближение методом внешнего штрафа.

Author's address: Ing. *Tomáš Roubíček,* CSc., Ústav teorie informace a automatizace ČSAV, Pod vodárenskou věží 4, 182 08 Praha 8.