Research Article

Well-Posedness and Primal-Dual Analysis of Some Convex Separable Optimization Problems

Stefan M. Stefanov

Department of Informatics, South-West University "Neofit Rilski", 2700 Blagoevgrad, Bulgaria

Correspondence should be addressed to Stefan M. Stefanov; stephanovsm@abv.bg

We focus on some convex separable optimization problems, considered by the author in previous papers, for which problems, necessary and sufficient conditions or sufficient conditions have been proved, and convergent algorithms of polynomial computational complexity have been proposed for solving these problems. The concepts of well-posedness of optimization problems in the sense of Tychonov, Hadamard, and in a generalized sense, as well as calmness in the sense of Clarke, are discussed. It is shown that the convex separable optimization problems under consideration are calm in the sense of Clarke. The concept of stability of the set of saddle points of the Lagrangian in the sense of Gol'shtein is also discussed, and it is shown that this set is not stable for the "classical" Lagrangian. However, it turns out that despite this instability, due to the specificity of the approach, suggested by the author for solving problems under consideration, it is not necessary to use modified Lagrangians but only the "classical" Lagrangians. Also, a primal-dual analysis for problems under consideration in view of methods for solving them is presented.

1. Introduction

1.1. Statement of Problems under Consideration: Preliminary Results. In this paper, we study well-posedness and present primal-dual analysis of some convex separable optimization problems, considered by the author in previous papers. For the sake of convenience, in this subsection we recall main results of earlier papers that are used in this study.

In paper [1], the following convex separable optimization problem was considered:

(C)

$$\min \quad \left\{ c(x) = \sum_{j \in J} c_j \left(\mathbf{x}_j \right) \right\}, \tag{1}$$

$$\text{subject to} \quad \sum_{j \in J} d_j \left(x_j \right) \le \alpha, \tag{2}$$

$$a_j \le x_j \le b_j, \quad j \in J, \tag{3}$$

where $c_j(x_j)$ are twice differentiable strictly convex functions, and $d_j(x_j)$ are twice differentiable convex functions, defined on the open convex sets X_j in $\mathbb{R}$, $j \in J$, respectively, $d'_j(x_j) > 0$ for every $j \in J$, $\mathbf{x} = (x_j)_{j \in J}$, and $J \equiv \{1, \dots, n\}$.

Assumptions for problem (C) are as follows.

(I.1) $a_j \le b_j$ for all $j \in J$. If $a_k = b_k$ for some $k \in J$, then the value $x_k := a_k = b_k$ is determined *a priori*.

(II.1) $\sum_{j \in J} d_j(a_j) \le \alpha$. Otherwise, the constraints (2) and (3) are inconsistent and the feasible set X, defined by (2)-(3), is empty. In addition to this assumption, we suppose that $\alpha \le \sum_{j \in J} d_j(b_j)$ in some cases which are specified below.

(III.1) (Slater's constraint qualification) There exists a point $\overline{\mathbf{x}} = (\overline{x}_1, \dots, \overline{x}_n) \in X$ such that $\sum_{j \in J} d_j(\overline{x}_j) < \alpha$.

Under these assumptions, the following characterization theorem (necessary and sufficient condition) for problem (C) was proved in [1].

Denote by $h_j^{\le}$, $h_j^{=}$, $h_j^{\ge}$ the values of x_j, for which $c'_j(x_j) = 0$ for the three problems under consideration in this paper, respectively.

Theorem 1 (characterization of the optimal solution to problem (C)). *Under the above assumptions, a feasible solution*

$\mathbf{x}^* = (x_j^*)_{j \in J} \in X$ *is an optimal solution to problem* (C) *if and only if there exists a* $\lambda \in \mathbb{R}_+^1$ *such that*

$$x_j^* = a_j, \quad j \in J_a^\lambda \overset{\text{def}}{=} \left\{ j \in J : \lambda \geq -\frac{c_j'(a_j)}{d_j'(a_j)} \right\} \tag{4}$$

$$x_j^* = b_j, \quad j \in J_b^\lambda \overset{\text{def}}{=} \left\{ j \in J : \lambda \leq -\frac{c_j'(b_j)}{d_j'(b_j)} \right\} \tag{5}$$

$$x_j^* : \lambda d_j'(x_j^*) = -c_j'(x_j^*),$$

$$j \in J^\lambda \overset{\text{def}}{=} \left\{ j \in J : -\frac{c_j'(b_j)}{d_j'(b_j)} < \lambda < -\frac{c_j'(a_j)}{d_j'(a_j)} \right\}. \tag{6}$$

The following polynomial algorithm for solving problem (C) with strictly convex differentiable functions $c_j(x_j)$, $j \in J$, was suggested in [1].

Algorithm 2.

(0) Initialization: $J := \{1, \ldots, n\}$, $k := 0$, $\alpha^{(0)} := \alpha$, $n^{(0)} := n$, $J^{(0)} := J$, $J_a^\lambda := \emptyset$, $J_b^\lambda := \emptyset$, initialize $h_j^\leq$, $j \in J$.

If $\sum_{j \in J} d_j(a_j) \leq \alpha$ go to Step (1), else go to Step (9).

(1) Construct the sets J_a^0, J_b^0, J^0 by using (4), (5), (6) with $\lambda = 0$.

Calculate

$$\delta(0) := \sum_{j \in J_a^0} d_j(a_j) + \sum_{j \in J_b^0} d_j(b_j)$$
$$+ \sum_{j \in J^0} d_j(h_j^\leq) - \alpha. \tag{7}$$

If $\delta(0) \leq 0$ then $\lambda := 0$, go to Step (8)

else if $\delta(0) > 0$, then

if $\alpha \leq \sum_{j \in J} d_j(b_j)$ go to Step (2)
else if $\alpha > \sum_{j \in J} d_j(b_j)$ go to Step (9)
(there does not exist $\lambda^* > 0$ such that $\delta(\lambda^*) = 0$).

(2) $J^{\lambda(k)} := J^{(k)}$. Calculate $\lambda^{(k)}$ by using the explicit expression of λ, calculated from the equality $\sum_{j \in J^{\lambda(k)}} d_j(x_j) = \alpha^{(k)}$, where $x_j, j \in J^{\lambda(k)}$, are given by (6). Go to Step (3).

(3) Construct the sets $J_a^{\lambda(k)}$, $J_b^{\lambda(k)}$, $J^{\lambda(k)}$ through (4), (5), (6) (with $j \in J^{(k)}$ instead of $j \in J$) and find their cardinal numbers $|J_a^{\lambda(k)}|$, $|J_b^{\lambda(k)}|$, $|J^{\lambda(k)}|$, respectively. Go to Step (4).

(4) Calculate

$$\delta(\lambda^{(k)}) := \sum_{j \in J_a^{\lambda(k)}} d_j(a_j) + \sum_{j \in J_b^{\lambda(k)}} d_j(b_j)$$
$$+ \sum_{j \in J^{\lambda(k)}} d_j(x_j^*) - \alpha^{(k)}, \tag{8}$$

where $x_j^*, j \in J^{\lambda(k)}$, are calculated from (6) with $\lambda = \lambda^{(k)}$. Go to Step (5).

(5) If $\delta(\lambda^{(k)}) = 0$ or $J^{\lambda(k)} = \emptyset$ then $\lambda := \lambda^{(k)}$, $J_a^\lambda := J_a^\lambda \cup J_a^{\lambda(k)}$, $J_b^\lambda := J_b^\lambda \cup J_b^{\lambda(k)}$, $J^\lambda := J^{\lambda(k)}$, go to Step (8)

else if $\delta(\lambda^{(k)}) > 0$ go to Step (6)

else if $\delta(\lambda^{(k)}) < 0$ go to Step (7).

(6) $x_j^* := a_j$ for $j \in J_a^{\lambda(k)}$, $\alpha^{(k+1)} := \alpha^{(k)} - \sum_{j \in J_a^{\lambda(k)}} d_j(a_j)$, $J^{(k+1)} := J^{(k)} \setminus J_a^{\lambda(k)}$,

$n^{(k+1)} := n^{(k)} - |J_a^{\lambda(k)}|$, $J_a^\lambda := J_a^\lambda \cup J_a^{\lambda(k)}$, $k := k + 1$. Go to Step (2).

(7) $x_j^* := b_j$ for $j \in J_b^{\lambda(k)}$, $\alpha^{(k+1)} := \alpha^{(k)} - \sum_{j \in J_b^{\lambda(k)}} d_j(b_j)$, $J^{(k+1)} := J^{(k)} \setminus J_b^{\lambda(k)}$,

$n^{(k+1)} := n^{(k)} - |J_b^{\lambda(k)}|$, $J_b^\lambda := J_b^\lambda \cup J_b^{\lambda(k)}$, $k := k + 1$. Go to Step (2).

(8) $x_j^* := a_j$ for $j \in J_a^\lambda$; $x_j^* := b_j$ for $j \in J_b^\lambda$; assign x_j^* the value calculated from (6) for $j \in J^\lambda$. Go to Step (10).

(9) Problem (C) has no optimal solution because $X = \emptyset$ or there does not exist $\lambda > 0$ satisfying Theorem 1.

(10) End.

It is proved in [1] that this algorithm is convergent.

Theorem 3 (convergence of Algorithm 2). *Let $\{\lambda^{(k)}\}$ be the sequence generated by Algorithm 2. Then,*

(i) *if $\delta(\lambda^{(k)}) > 0$, then $\lambda^{(k)} \leq \lambda^{(k+1)}$,*

(ii) *if $\delta(\lambda^{(k)}) < 0$, then $\lambda^{(k)} \geq \lambda^{(k+1)}$.*

In paper [2], the following two convex separable optimization problems were considered:
(C$^=$)

$$\min \quad \left\{ c(\mathbf{x}) = \sum_{j \in J} c_j(x_j) \right\} \tag{9}$$

subject to $\sum_{j \in J} d_j x_j = \alpha,$ \hfill (10)

$$a_j \leq x_j \leq b_j, \quad j \in J, \tag{11}$$

and
(C$^\geq$)

$$\min \quad \left\{ c(\mathbf{x}) = \sum_{j \in J} c_j(x_j) \right\} \tag{12}$$

subject to $\sum_{j \in J} d_j x_j \geq \alpha$ \hfill (13)

$$a_j \leq x_j \leq b_j, \quad j \in J, \tag{14}$$

where for both problems, $c_j(x_j)$ are twice differentiable convex functions, defined on the open convex sets X_j in $\mathbb{R}$,

$j \in J$, respectively, $d_j > 0$, for every $j \in J$, $\mathbf{x} = (x_j)_{j \in J}$, and $J = \{1, \ldots, n\}$.

Assumptions for problem $(C^=)$ are as follows.

(I.2) $a_j \leq b_j$ for each $j \in J$.

(II.2) $\sum_{j \in J} d_j a_j \leq \alpha \leq \sum_{j \in J} d_j b_j$. Otherwise the constraints (10), (11) are inconsistent and $X^L = \emptyset$, where X^L is defined by (10)-(11).

Under these assumptions, the following characterization theorem (necessary and sufficient condition) for problem $(C^=)$ is proved in [2].

Theorem 4 (characterization of the optimal solution to problem $(C^=)$). *A feasible solution $\mathbf{x}^* = (x_j^*)_{j \in J} \in X^L$ is an optimal solution to problem $(C^=)$ if and only if there exists a $\lambda \in \mathbb{R}^1$ such that*

$$x_j^* = a_j, \quad j \in J_a^\lambda \overset{\text{def}}{=} \left\{ j \in J : \lambda \geq -\frac{c_j'(a_j)}{d_j} \right\}, \qquad (15)$$

$$x_j^* = b_j, \quad j \in J_b^\lambda \overset{\text{def}}{=} \left\{ j \in J : \lambda \leq -\frac{c_j'(b_j)}{d_j} \right\}, \qquad (16)$$

$$x_j^* : \lambda d_j = -c_j'(x_j^*),$$

$$j \in J^\lambda \overset{\text{def}}{=} \left\{ j \in J : -\frac{c_j'(b_j)}{d_j} < \lambda < -\frac{c_j'(a_j)}{d_j} \right\}. \qquad (17)$$

Assumptions for problem $(C^\geq)$ are as follows.

(I.3) $a_j \leq b_j$ for all $j \in J$.

(II.3) $\alpha \leq \sum_{j \in J} d_j b_j$. Otherwise the constraints (13), (14) are inconsistent and $X^\geq = \emptyset$, where $X^\geq$ is defined by (13)-(14).

Under these assumptions, the following theorem (sufficient condition) for problem $(C^\geq)$ is proved in [2].

Theorem 5 (sufficient condition for optimal solution to problem $(C^\geq)$). *Let x_j^*, $j \in J$ be components of the optimal solution to problem $C^=$. Then:*

(i) *If $\lambda = -c_j'(x_j^*)/d_j \leq 0$, then x_j^*, $j \in J$, solve problem $(C^\geq)$ as well.*

(ii) *If $\lambda = -c_j'(x_j^*)/d_j > 0$, then $x_j^\geq$, $j \in J$ defined as follows:*

$$x_j^\geq = b_j, \ j \in J_b^\lambda,$$

$$x_j^\geq = \min\{b_j, h_j^\geq\}, \ j \in J^\lambda,$$

$$x_j^\geq = \min\{b_j, h_j^\geq\} \text{ for all } j \in J_a^\lambda \text{ such that } c_j'(a_j) < 0,$$

$$x_j^\geq = a_j \text{ for all } j \in J_a^\lambda, \text{ such that } c_j'(a_j) \geq 0$$

solve problem $(C^\geq)$.

The following polynomial algorithm for solving problem $(C^=)$ with strictly convex differentiable functions $c_j(x_j)$ was suggested in [2].

Algorithm 6.

(1) Initialization: $J := \{1, \ldots, n\}$, $k := 0$, $\alpha^{(0)} := \alpha$, $n^{(0)} := n$, $J^{(0)} := J$, $J_a^\lambda := \emptyset$, $J_b^\lambda := \emptyset$, initialize $h_j^=$, $j \in J$.

If $\sum_{j \in J} d_j a_j \leq \alpha \leq \sum_{j \in J} d_j b_j$ go to Step (2), else go to Step (9).

(2) $J^{\lambda(k)} := J^{(k)}$. Calculate $\lambda^{(k)}$ by using the explicit expression of λ, calculated from the equality constraint $\sum_{j \in J^{\lambda(k)}} d_j x_j = \alpha^{(k)}$, where x_j, $j \in J^{\lambda(k)}$, are given by (17). Go to Step (3).

(3) Construct the sets $J_a^{\lambda(k)}$, $J_b^{\lambda(k)}$, $J^{\lambda(k)}$ through (15), (16), (17) (with $J^{(k)}$ instead of J) and find their cardinalities $|J_a^{\lambda(k)}|$, $|J_b^{\lambda(k)}|$, $|J^{\lambda(k)}|$, respectively. Go to Step (4).

(4) Calculate

$$\delta\left(\lambda^{(k)}\right) := \sum_{j \in J_a^{\lambda(k)}} d_j a_j + \sum_{j \in J_b^{\lambda(k)}} d_j b_j$$

$$+ \sum_{j \in J^{\lambda(k)}} d_j x_j^* - \alpha^{(k)}, \qquad (18)$$

where x_j^*, $j \in J^{\lambda(k)}$, are calculated from (17) with $\lambda = \lambda^{(k)}$. Go to Step (5).

(5) If $\delta(\lambda^{(k)}) = 0$ or $J^{\lambda(k)} = \emptyset$ then $\lambda := \lambda^{(k)}$, $J_a^\lambda := J_a^\lambda \cup J_a^{\lambda(k)}$, $J_b^\lambda := J_b^\lambda \cup J_b^{\lambda(k)}$, $J^\lambda := J^{\lambda(k)}$, go to Step (8)

else if $\delta(\lambda^{(k)}) > 0$ go to Step (6)

else if $\delta(\lambda^{(k)}) < 0$ go to Step (7).

(6) $x_j^* := a_j$ for $j \in J_a^{\lambda(k)}$, $\alpha^{(k+1)} := \alpha^{(k)} - \sum_{j \in J_a^{\lambda(k)}} d_j a_j$, $J^{(k+1)} := J^{(k)} \setminus J_a^{\lambda(k)}$,

$n^{(k+1)} := n^{(k)} - |J_a^{\lambda(k)}|$, $J_a^\lambda := J_a^\lambda \cup J_a^{\lambda(k)}$, $k := k + 1$. Go to Step (2).

(7) $x_j^* := b_j$ for $j \in J_b^{\lambda(k)}$, $\alpha^{(k+1)} := \alpha^{(k)} - \sum_{j \in J_b^{\lambda(k)}} d_j b_j$, $J^{(k+1)} := J^{(k)} \setminus J_b^{\lambda(k)}$,

$n^{(k+1)} := n^{(k)} - |J_b^{\lambda(k)}|$, $J_b^\lambda := J_b^\lambda \cup J_b^{\lambda(k)}$, $k := k + 1$. Go to Step (2).

(8) $x_j^* := a_j$ for $j \in J_a^\lambda$; $x_j^* := b_j$ for $j \in J_b^\lambda$; assign x_j^* the value, calculated from (17) for $j \in J^\lambda$. Go to Step (10).

(9) The problem has no optimal solution because $X^L = \emptyset$.

(10) End.

It is proved in [2] that this algorithm is convergent.

Theorem 7 (convergence of Algorithm 6). *Let $\{\lambda^{(k)}\}$ be the sequence generated by Algorithm 6. Then,*

(i) *if $\delta(\lambda^{(k)}) > 0$, then $\lambda^{(k)} \leq \lambda^{(k+1)}$,*

(ii) *if $\delta(\lambda^{(k)}) < 0$, then $\lambda^{(k)} \geq \lambda^{(k+1)}$.*

The following algorithm for solving problem $(C^{\geq})$ with strictly convex differentiable functions $c_j(x_j)$ is suggested in [2].

Algorithm 8.

(1) Initialization: $J := \{1, \ldots, n\}$, $k := 0$, $\alpha^{(0)} := \alpha$, $n^{(0)} := n$, $J^{(0)} := J$, $J_a^{\lambda} := \emptyset$, $J_b^{\lambda} := \emptyset$, initialize $h_j^{\geq}$, $j \in J$.

If $\sum_{j \in J} d_j a_j \leq \alpha \leq \sum_{j \in J} d_j b_j$ then go to Step (2), else go to Step (9).

Steps (2)–(7) are the same as Steps (2)–(7) of Algorithm 6, respectively.

(8) If $\lambda \leq 0$ then $x_j^{\geq} := a_j$ for $j \in J_a^{\lambda}$; $x_j^{\geq} := b_j$ for $j \in J_b^{\lambda}$;

assign $x_j^{\geq}$ the value, calculated through (17) for $j \in J^{\lambda}$, go to Step (10)

else if $\lambda > 0$ then

$x_j^{\geq} := b_j$ for $j \in J_b^{\lambda}$,

$x_j^{\geq} := \min\{b_j, h_j^{\geq}\}$ for $j \in J^{\lambda}$,

if $j \in J_a^{\lambda}$ and $c_j'(a_j) < 0$ then $x_j^{\geq} := \min\{b_j, h_j^{\geq}\}$

else if $j \in J_a^{\lambda}$ and $c_j'(a_j) \geq 0$ then $x_j^{\geq} := a_j$,

go to Step (10).

(9) Problem $(C^{\geq})$ has no optimal solution because $X^{\geq} = \emptyset$ or there do not exist $x_j^* \in [a_j, b_j]$, $j \in J$, such that $\sum_{j \in J} d_j x_j^* = \alpha$.

(10) End.

Since Algorithm 8 is based on Theorem 5 and Algorithm 6, and since the "iterative" Steps (2)–(7) of Algorithms 6 and 8 are the same, then the "convergence" of Algorithm 8 follows from Theorem 7 as well.

1.2. Organization of the Paper. The rest of the paper is organized as follows. In Section 2, the concepts of well-posedness of optimization problems in the sense of Tychonov, Hadamard, and in a generalized sense, as well as calmness in the sense of Clarke, are discussed. It is shown in Section 2.3 that the convex separable optimization problems under consideration are calm in the sense of Clarke. In Section 3, the concept of stability of the set of saddle points of the Lagrangian in the sense of Gol'shtein is also discussed and it is shown that this set is not stable for the "classical" Lagrangian. However, it is explained that despite this instability, due to the specificity of the approach, suggested by the author in previous papers for solving problems under consideration, it is not necessary to use modified Lagrangians but only the "classical" Lagrangians. In Section 4, primal-dual analysis of the problems under consideration in view of methods for solving them is presented. Main results of well-posedness and primal-dual analysis are included in Section 2.3 and in Sections 3 and 4.

2. Well-Posedness of Optimization Problems

Questions of existence of solutions and how they depend on problem's parameters are usually important for many problems of mathematics, not only in optimization. The term *well-posedness* refers to the existence and uniqueness of a solution and its continuous behavior with respect to data perturbations, which is referred to as *stability*. In general, a problem is said to be *stable* if

$$\varepsilon(\delta) \longrightarrow 0 \quad \text{when } \delta \longrightarrow 0, \tag{19}$$

where δ is a given tolerance of the problem's data, $\varepsilon(\delta)$ is the accuracy with which the solution can be determined, and $\varepsilon(\delta)$ is a continuous function of δ. Besides these conditions, accompanying robustness properties in the convergence of sequence of approximate solutions are also required.

Problems which are not well-posed are called *ill-posed*, or, sometimes, *improperly posed*.

2.1. Tychonov and Hadamard Well-Posedness: Well-Posedness in the Generalized Sense. Recall that f is a *proper function* if $f(\mathbf{x}) < \infty$ for at least one $\mathbf{x} \in \mathbb{R}^n$ and $f(\mathbf{x}) > -\infty$ for all $\mathbf{x} \in \mathbb{R}^n$, or, in other words, if

$$\text{dom } f \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) < \infty\} \tag{20}$$

is a nonempty set on which $f(\mathbf{x}) > -\infty$, where dom f is the *effective domain* of f. Otherwise, f is *improper*.

Definition 9. Let X be a space with either a topology or a convergence structure associated and let $f : X \to \overline{\mathbb{R}} \equiv \mathbb{R} \cup \{+\infty\}$ be a proper extended real-valued function. Consider the problem

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in X. \end{aligned} \tag{21}$$

The problem (21) is *Tychonov well-posed* if and only if f has a unique global minimum point on X towards which every minimizing sequence converges.

An equivalent definition is as follows: problem (21) is *Tychonov well-posed* if and only if there exists a unique $\mathbf{x}_0 \in X$ such that $f(\mathbf{x}_0) \leq f(\mathbf{x})$ for all $\mathbf{x} \in X$ and

$$f(\mathbf{x}_n) \longrightarrow f(\mathbf{x}_0) \quad \text{implies } \mathbf{x}_n \longrightarrow \mathbf{x}_0. \tag{22}$$

There are two ways to cope with ill-posedness.

The first one is to change the statement of the problem.

The second one is the so-called *Tychonov regularization*. A parametric functional is constructed such that if it approaches 0, the solution of the "regularized" problem converges to the exact solution of the original problem.

Consider the problem

$$\min_{\mathbf{x} \in X \subset \mathbb{R}^n} f(\mathbf{x}) = f(\mathbf{x}^*). \tag{23}$$

Associate the following problem with (23):

$$\min_{\mathbf{x} \in X} \left[f(\mathbf{x}) + n_k(\mathbf{x}) \right] = z_k(\mathbf{x}^*(n_k(\mathbf{x}))), \tag{24}$$

where $n_k(\mathbf{x})$ is perturbation in the input data and $\mathbf{x}^*(n_k(\mathbf{x}))$ is an optimal solution to the perturbed problem.

Let

$$\int_{\mathbb{R}^n} n_k^2(\mathbf{x})\, d\mathbf{x} \le c_k^2 < \infty. \tag{25}$$

If

$$\left| z_k\left(\mathbf{x}^*\left(n_k(\mathbf{x})\right)\right) - f(\mathbf{x}^*) \right| \longrightarrow 0, \tag{26}$$

when $c_k \rightarrow 0$, then problem (23) is stable with respect to perturbation $n_k(\mathbf{x})$.

A parametric function $F(\mathbf{x}, \Delta, f(\mathbf{x}))$ with a parameter Δ is called a *regularizing function* for problem (23) with respect to perturbation $n_k(\mathbf{x})$ if the following conditions are satisfied.

(1) $F(\mathbf{x}, \Delta, f(\mathbf{x}))$ is defined for all $\mathbf{x} \in X$ and $\Delta > 0$.

(2) If $\mathbf{x}^*(n_k(\mathbf{x}), \Delta)$ is an optimal solution to problem

$$\min_{\mathbf{x} \in X} F\left(\mathbf{x}, \Delta, z_k(\mathbf{x})\right)$$
$$= F\left(\mathbf{x}^*\left(n_k(\mathbf{x}), \Delta\right), \Delta, z_k(\mathbf{x})\right), \tag{27}$$

then there exists a function $\Delta_k = \Delta_k(c_k)$ such that

$$F\left(\mathbf{x}^*\left(n_k(\mathbf{x}), \Delta_k\right), \Delta_k, z_k(\mathbf{x})\right) \longrightarrow f(\mathbf{x}^*), \tag{28}$$

when $c_k \rightarrow 0$.

Following Tychonov, an ill-posed problem is said to be *regularizable* if there exists at least one regularizing function for it.

The concept of Tychonov well-posedness can be extended to problems without the uniqueness of the optimal solution.

Definition 10. Let X be a space with either a topology or a convergence structure associated, and $f : X \rightarrow \overline{\mathbb{R}} \equiv \mathbb{R} \cup \{+\infty\}$ be a proper real-valued function. Problem (21) is said to be *well-posed in the generalized sense* if and only if $\arg\min_{\mathbf{x} \in X} f(\mathbf{x}) \ne \emptyset$ and every sequence $\{\mathbf{u}_n\} \subset X$ such that $f(\mathbf{u}_n) \rightarrow \inf\{f(\mathbf{x}) : \mathbf{x} \in X\}$ has some subsequence $\{\mathbf{v}_n\} \rightarrow \mathbf{u}$ with $\mathbf{u} \in \arg\min_{\mathbf{x} \in X} f(\mathbf{x})$.

Problem (21) is Tychonov well-posed if and only if it is well-posed in the generalized sense and $\arg\min_{\mathbf{x} \in X} f(\mathbf{x})$ is a singleton.

Hadamard well-posedness is primarily connected with problems of mathematical physics (boundary value problems for partial differential equations) and can be extended to mathematical programming problems. We do not discuss this topic here.

As recent studies in the calculus of variations, optimal control, and numerical methods of optimization show, uniqueness and continuity are often too restrictive to be adopted as the standards of well-posedness. It turns out that practical concepts concerning well-posedness are some forms of semicontinuity in the problem's data and solution mapping, along with potential multivaluedness in this mapping.

2.2. Calmness in the Sense of Clarke. Let X be a Banach space.

Definition 11. Let Y be a subset of X. A function $f : Y \rightarrow \mathbb{R}$ is said to satisfy a *Lipschitz condition* on Y provided that,

for some nonnegative scalar K, the following inequality holds true:

$$\left| f(\mathbf{y}) - f(\mathbf{y}') \right| \le K \left\| \mathbf{y} - \mathbf{y}' \right\|, \tag{29}$$

for all points $\mathbf{y}, \mathbf{y}' \in Y$; this is also referred to as a Lipschitz condition of *rank K*. We say that f is *Lipschitz* (of rank K) *near* $\mathbf{x}$ if for some $\varepsilon > 0$, f satisfies a Lipschitz condition (of rank K) on the set $\mathbf{x} + \varepsilon B$ (i.e., within an ε-neighborhood of $\mathbf{x}$), where B is the open unit ball around $\mathbf{0}$.

A function f, which satisfies a Lipschitz condition, sometimes is said to be *Lipschitz continuous*.

Consider the following general mathematical programming problem:

(P)

$$\min \quad f(\mathbf{x})$$
$$\text{subject to} \quad g_i(\mathbf{x}) \le 0, \quad i = 1, \ldots, t,$$
$$h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, m, \tag{30}$$
$$\mathbf{x} \in C, \quad C \subset X,$$

where g_i, h_j are real-valued functions on X.

Let $\mathbf{g}$ and $\mathbf{h}$ be the functions $\mathbf{g} = [g_1, \ldots, g_t] : X \rightarrow \mathbb{R}^t$, $\mathbf{h} = [h_1, \ldots, h_m] : X \rightarrow \mathbb{R}^m$.

Let (P) be imbedded in a parametrized family $P(\mathbf{p}, \mathbf{q})$ of mathematical programs, where $\mathbf{p} \in \mathbb{R}^t$, $\mathbf{q} \in \mathbb{R}^m$:

$P(\mathbf{p}, \mathbf{q})$

$$\min \quad f(\mathbf{x})$$
$$\text{subject to} \quad \mathbf{g}(\mathbf{x}) + \mathbf{p} \le \mathbf{0}$$
$$\mathbf{h}(\mathbf{x}) + \mathbf{q} = \mathbf{0} \tag{31}$$
$$\mathbf{x} \in C.$$

Denote by A the feasible region of problem $P(\mathbf{p}, \mathbf{q})$.

Definition 12 (Clarke [3]). The *value function* $V : \mathbb{R}^t \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is defined via $V(\mathbf{p}, \mathbf{q}) = \inf\{P(\mathbf{p}, \mathbf{q})\}$ (i.e., the value of the problem $P(\mathbf{p}, \mathbf{q})$). If there are no feasible points for $P(\mathbf{p}, \mathbf{q})$, then the infimum is over the empty set and $V(\mathbf{p}, \mathbf{q})$ is assigned the value $+\infty$.

Definition 13 (Clarke [3]). Let $\mathbf{x}$ solve (P). The problem (P) is *calm* at $\mathbf{x}$ provided that there exist positive ε and M such that for all $(\mathbf{p}, \mathbf{q}) \in \varepsilon B$, for all $\mathbf{x}' \in \mathbf{x} + \varepsilon B$ which are feasible for $P(\mathbf{p}, \mathbf{q})$, one has

$$f(\mathbf{x}') - f(\mathbf{x}) + M \left\| (\mathbf{p}, \mathbf{q}) \right\| \ge 0, \tag{32}$$

where B is the open unit ball in X and $\| (\mathbf{p}, \mathbf{q}) \|$ is the Euclidean norm of $(\mathbf{p}, \mathbf{q})$.

Let U be an open convex subset of X.

Theorem 14 (Roberts and Varberg [4], Clarke [3]; Lipschitz condition from boundedness of a convex function). *Let f be a convex function, bounded above on a neighborhood of some point of U. Then, for any $\mathbf{x}$ in U, f is Lipschitz near $\mathbf{x}$.*

Recall that *limit superior* of a bounded sequence $\{x_n\}$ in $\mathbb{R}$, denoted $\limsup\{x_n\}$ or $\overline{\lim}\{x_n\}$, equals the infimum of all numbers $q \in \mathbb{R}$ for which at most a finite number of elements of $\{x_n\}$ (strictly) exceed q. Similarly, *limit inferior* of $\{x_n\}$ is given by $\liminf\{x_n\} \equiv \underline{\lim}\{x_n\} \equiv \sup\{q :$ at most a finite number of elements of $\{x_n\}$ are (strictly) less than $q\}$.

A bounded sequence always has a unique limit superior and limit inferior.

Theorem 15 (Clarke [3], Calmness). *Let* $V(\mathbf{0}, \mathbf{0})$ *be finite and suppose that*

$$\liminf_{(\mathbf{p}, \mathbf{q}) \to (\mathbf{0}, \mathbf{0})} \frac{V(\mathbf{p}, \mathbf{q}) - V(\mathbf{0}, \mathbf{0})}{\|(\mathbf{p}, \mathbf{q})\|} > -\infty \qquad (33)$$

(this is true in particular if V *is Lipschitz near* $(\mathbf{0}, \mathbf{0})$*). Then, for any solution* $\mathbf{x}$ *to* (P)*, problem* (P) *is calm at* $\mathbf{x}$.

Sometimes problem (P) is said to be *calm* provided V satisfies the hypothesis of Theorem 15.

Slater-Type Conditions. Suppose that (P) has no equality constraints (i.e., $m = 0$), that the functions g_i, $i = 1, \ldots, t$, are convex, and that C is a convex set. Recall that *Slater's condition* (*Slater's constraint qualification*), then is: there exists a point $\bar{\mathbf{x}}$ in C such that $g_i(\bar{\mathbf{x}}) < 0$, $i = 1, \ldots, t$ ($\bar{\mathbf{x}}$ is called a *strictly feasible point*).

For $\mathbf{p} \in \mathbb{R}^t$, let $V(\mathbf{p})$ be the infimum in the problem $P(\mathbf{p})$ in which the constraints $g_i(\mathbf{x}) \le 0$ of problem (P) are replaced by $g_i(\mathbf{x}) + p_i \le 0$.

Theorem 16 (Clarke [3]; Lipschitz property of the value function from Slater's condition). *If* C *is bounded and* f *is Lipschitz on* C, *then Slater's condition (i.e., the existence of a strictly feasible point) implies that* V *is Lipschitz near* $\mathbf{0}$.

Theorems 15 and 16 mean that Slater's constraint qualification implies calmness of problem $P(\mathbf{p})$ in this case.

Theorem 17 (Clarke [3]; Calmness of a problem subject to inequality constraints). *Let* P *incorporate only inequality constraints* $g_i(\mathbf{x}) \le 0$ *and the abstract constraint* $\mathbf{x} \in C$ *and suppose that the value function* $V(\mathbf{p})$ *is finite for* $\mathbf{p}$ *near* $\mathbf{0}$. *Then, for almost all* $\mathbf{p}$ *in a neighborhood of* $\mathbf{0}$, *the problem* $P(\mathbf{p})$ *is calm.*

Remark 18. In the case of problem (P), in which equality constraints exist, it is a consequence of Ekeland's theorem that $P(\mathbf{p}, \mathbf{q})$ is calm for all $(\mathbf{p}, \mathbf{q})$ in a dense subset of any open set upon which V is bounded and lower semicontinuous.

Consider the following way of perturbing problem (P): $P(\boldsymbol{\alpha})$

$$\min \{f(\mathbf{x}, \boldsymbol{\alpha}) : \mathbf{g}(\mathbf{x}, \boldsymbol{\alpha}) \le \mathbf{0}, \mathbf{h}(\mathbf{x}, \boldsymbol{\alpha}) = \mathbf{0}, (\mathbf{x}, \boldsymbol{\alpha}) \in D\}, \quad (34)$$

where $\boldsymbol{\alpha}$ is a vector of k real components. The value function V then would be a function of $\boldsymbol{\alpha}$: $V(\boldsymbol{\alpha}) = \inf P(\boldsymbol{\alpha})$.

This is a special case of problem $P(\mathbf{p}, \mathbf{q})$ with $k = t + m$, $\boldsymbol{\alpha} = (\mathbf{p}, \mathbf{q})$, $f(\mathbf{x}, \boldsymbol{\alpha}) = f(\mathbf{x})$, $\mathbf{g}(\mathbf{x}, \boldsymbol{\alpha}) = \mathbf{g}(\mathbf{x}) + \mathbf{p}$, $\mathbf{h}(\mathbf{x}, \boldsymbol{\alpha}) = \mathbf{h}(\mathbf{x}) + \mathbf{q}$, $D = C \times \mathbb{R}^{t+m}$. At least when the dependence of f, $\mathbf{g}$, and $\mathbf{h}$ on $\boldsymbol{\alpha}$ is locally Lipschitz, we can consider problem

$P(\mathbf{p}, \mathbf{q})$ with $k = t + m$, $\boldsymbol{\alpha} = (\mathbf{p}, \mathbf{q})$, $f(\mathbf{x}, \boldsymbol{\alpha}) = f(\mathbf{x})$, $\mathbf{g}(\mathbf{x}, \boldsymbol{\alpha}) = \mathbf{g}(\mathbf{x}) + \mathbf{p}$, $\mathbf{h}(\mathbf{x}, \boldsymbol{\alpha}) = \mathbf{h}(\mathbf{x}) + \mathbf{q}$, and $D = C \times \mathbb{R}^{t+m}$ rather than problem $P(\boldsymbol{\alpha})$. Hence, the methods and results, considered above, can be applied to perturbed family $P(\boldsymbol{\alpha})$ as well.

Constraint qualifications (regularity conditions) can be classified into two categories: on the one hand, Mangasarian-Fromowitz and Slater-type conditions and their extensions, and, on the other hand, constraint qualifications called *calmness*. It turns out that calmness is the weakest of these conditions, since it is implied by all the others (see, e.g., Theorem 16).

2.3. Well-Posedness of Problems (C), $(C^=)$, and $(C^\ge)$

2.3.1. Existence of Solutions.
The question of existence of solutions to problems (C), $(C^=)$, and $(C^\ge)$ has been discussed in Theorems 1, 4, and 5, respectively. Steps (0), (1), and (9) of Algorithm 2 and Steps (1) and (9) of Algorithms 6 and 8, respectively, refer to these results.

2.3.2. Uniqueness of Solution.
The question of uniqueness of the optimal solution to problems under consideration is also important.

If $c(\mathbf{x}) \equiv \sum_{j \in J} c_j(x_j)$ defined by (1) (by (9), (12), resp.) is a strictly convex function, then problem (C) (problem $(C^=)$ or problem $(C^\ge)$, resp.) has a unique optimal solution in the feasible region X (X^L, $X^\ge$, resp.) in case problem (C) (problem $(C^=)$ or problem $(C^\ge)$, resp.) has feasible solutions; that is, x_j^*, $j \in J^\lambda$, are uniquely determined from (6) [(17)] in the interval $[a_j, b_j]$ in this case. If the parameters a_j, b_j, and so forth of particular problems of the form (C) ($(C^=)$ and $(C^\ge)$, resp.) are generated in intervals where the functions $c_j(x_j)$ are strictly convex, then problem (C) (problem $(C^=)$ or problem $(C^\ge)$, resp.), if it has feasible solutions, has a *unique* optimal solution.

In the general case, if functions $c_j(x_j)$ are convex but not necessarily strictly convex, then, as it is known, a convex programming problem has more than one optimal solution and the set of optimal solutions to such a problem is convex. Further, the optimal value of the objective function is the same for all optimal solutions to problem (C) (problem $(C^=)$ or problem $(C^\ge)$, resp.) if it has more than one optimal solution. If, for example, (6) ((17), resp.) is a linear equation of x_j^*, then x_j^*, $j \in J^\lambda$, are also uniquely determined from (6) (from (17), resp.).

2.3.3. Calmness of the Problems (of the Optimal Solutions).
Let $(C(\mathbf{p}))$, $(C^=(\mathbf{p}, q))$, and $(C^\ge(\mathbf{p}))$ be the parametrized families of mathematical programs associated with problems (C), $(C^=)$, and $(C^\ge)$, respectively.

Feasible regions of problems (C) and $(C^\ge)$ are nonempty by the assumption; this is satisfied when $\sum_{j \in J} d_j(a_j) \le \alpha$ and $\sum_{j \in J} d_j b_j \ge \alpha$, respectively. Without loss of generality, feasible regions

$X(\mathbf{p})$:

$$\sum_{j \in J} d_j (x_j) + p_0 \le \alpha$$

$$a_j \le x_j + p_j \le b_j, \quad j \in J, \tag{35}$$

and $X^{\ge}(\mathbf{p})$:

$$\sum_{j \in J} d_j x_j + p_0 \ge \alpha$$

$$a_j \le x_j + p_j \le b_j, \quad j \in J, \tag{36}$$

of problems $(C(\mathbf{p}))$ and $(C^{\ge}(\mathbf{p}))$, respectively, are also nonempty in a neighborhood of $\mathbf{p} = \mathbf{0}$.

Since the value function $V(\mathbf{p})$, associated with problems $(C(\mathbf{p}))$ and $(C^{\ge}(\mathbf{p}))$, is finite near $\mathbf{0}$ (according to Definition 12 and the assumption that the corresponding feasible set is nonempty) then both problems are calm according to Theorem 17.

An alternative proof of calmness of problem (C) is the following.

The objective function $c(\mathbf{x})$ of problem (C) (1)–(3) is convex (and, therefore, Lipschitz in accordance with Theorem 14), and Slater's constraint qualification is satisfied by the assumption. From Theorem 16, it follows that the value function $V(\mathbf{p})$ is Lipschitz, and problem (C) is calm at any solution $\mathbf{x}^*$ of problem (C) according to Theorem 15.

Consider the parametrized family $(C^{=}(\mathbf{p}, q))$ in which problem $(C^{=})$ is imbedded as follows:

$(C^{=}(\mathbf{p}, q))$

$$\min \quad \left\{ c(\mathbf{x}) = \sum_{j \in J} c_j (x_j) \right\} \tag{37}$$

$$\text{subject to} \quad \mathbf{x} \in X^L (\mathbf{p}, q),$$

where $X^L(\mathbf{p}, q)$ is defined as follows:

$$\sum_{j \in J} d_j x_j + q = \alpha$$

$$a_j \le x_j + p_j \le b_j, \quad j \in J. \tag{38}$$

As it has been pointed out, $X^L \ne \emptyset$ if

$$\sum_{j \in J} d_j a_j \le \alpha \le \sum_{j \in J} d_j b_j, \tag{39}$$

whereas $X^L(\mathbf{p}, q) \ne \emptyset$ if

$$\sum_{j \in J} d_j (a_j - p_j) \le \alpha - q \le \sum_{j \in J} d_j (b_j - p_j). \tag{40}$$

Without loss of generality, assume that there exists a $(\mathbf{p}, q)$ such that $X^L(\mathbf{p}, q) \ne \emptyset$. This is satisfied, for example, when $\sum_{j \in J} d_j p_j = q$ in addition to the requirement $X^L \ne \emptyset$. Then the value function $V(\mathbf{p}, q)$, associated with $(C^{=}(\mathbf{p}, q))$, is finite by Definition 12.

Theorem 19 (Convexity of the infimum of a convex function subject to linear equality constraints). *Let f be a convex function and S be a convex set in $\mathbb{R}^n$. Then, function*

$$h(\mathbf{y}) \stackrel{\text{def}}{=} \inf_{\mathbf{x}} \{ f(\mathbf{x}) : A\mathbf{x} = \mathbf{y}, A_{m \times n}, \mathbf{x} \in S, \mathbf{y} \in \mathbb{R}^m \} \tag{41}$$

is convex.

Proof. Let $\mathbf{x} = \lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$, $\lambda \in [0, 1]$, $\mathbf{x}_1, \mathbf{x}_2 \in S$, $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^m$. Therefore $\mathbf{x} \in S$ as a convex combination of elements of the convex set S. Then,

$$h(\lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2)$$

$$= \inf_{\mathbf{x}} \{ f(\mathbf{x}) : A\mathbf{x} = \lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2 \}$$

$$= \inf_{\mathbf{x}_1, \mathbf{x}_2} \{ f(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2) : A(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2)$$

$$= \lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2 \}$$

$$\le \inf_{\mathbf{x}_1, \mathbf{x}_2} \{ f(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2) : A\mathbf{x}_1 = \mathbf{y}_1, A\mathbf{x}_2 = \mathbf{y}_2 \}$$

$$\le \inf_{\mathbf{x}_1, \mathbf{x}_2} \{ \lambda f(\mathbf{x}_1) + (1 - \lambda) f(\mathbf{x}_2) : A\mathbf{x}_1 = \mathbf{y}_1, A\mathbf{x}_2 = \mathbf{y}_2 \}$$

$$= \inf_{\mathbf{x}_1} \{ \lambda f(\mathbf{x}_1) : A\mathbf{x}_1 = \mathbf{y}_1 \}$$

$$+ \inf_{\mathbf{x}_2} \{ (1 - \lambda) f(\mathbf{x}_2) : A\mathbf{x}_2 = \mathbf{y}_2 \}$$

$$= \lambda h(\mathbf{y}_1) + (1 - \lambda) h(\mathbf{y}_2). \tag{42}$$

We have used that f is a convex function, the property that

$$\{ \mathbf{x}_1, \mathbf{x}_2 \in S : A\mathbf{x}_1 = \mathbf{y}_1, A\mathbf{x}_2 = \mathbf{y}_2 \}$$

$$\subset \{ \mathbf{x}_1, \mathbf{x}_2 \in S : \lambda A\mathbf{x}_1 + (1 - \lambda) A\mathbf{x}_2 = \lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2 \} \tag{43}$$

and the fact that $X \subset Y$ implies $\inf_{\mathbf{x} \in Y} f(\mathbf{x}) \le \inf_{\mathbf{x} \in X} f(\mathbf{x})$.

Therefore, $h(\mathbf{y})$ is a convex function by definition. $\square$

For problem $(C^{=}(\mathbf{p}, q))$, matrix A of Theorem 19 consists of a single row, that is, $m = 1$, and convex set S is the n-dimensional parallelepiped

$$S = \{ \mathbf{x} \in \mathbb{R}^n : a_j \le x_j \le b_j, j \in J \}. \tag{44}$$

The value function associated with problem $(C^{=}(\mathbf{p}, q))$ is

$$V(\mathbf{p}, q) = \inf_{\mathbf{x} \in X^L(\mathbf{p}, q)} f(\mathbf{x}). \tag{45}$$

From Theorem 19 and the assumption that $X^L(\mathbf{p}, q) \ne \emptyset$, it follows that $V(\mathbf{p}, q)$ is convex and finite, respectively, and

from Theorem 14 it follows that it is Lipschitz. Then, problem $(C^=(\mathbf{p}, q))$ is calm according to Theorem 15.

In the general case, if the mathematical program is not convex and equality constraints exist, we can use the approach of the Remark 18.

3. On the Stability of the Set of Saddle Points of the Lagrangian

3.1. The Concept of Stability of Saddle Points of the Lagrangian. Besides well-posedness of the optimization problems, stability of methods for solving these problems is also important.

Let $\Phi(\mathbf{x}, \mathbf{y})$ be a convex function of $\mathbf{x} \in X$ and a concave function of $\mathbf{y} \in Y$, where X and Y are convex and closed sets.

Recall the definition of a saddle point.

A point $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ is said to be a *saddle point* of function $\Phi(\mathbf{x}, \mathbf{y})$, $\mathbf{x} \in X$, $\mathbf{y} \in Y$, if the following inequalities hold:

$$\Phi(\widehat{\mathbf{x}}, \mathbf{y}) \leq \Phi(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \leq \Phi(\mathbf{x}, \widehat{\mathbf{y}}) \tag{46}$$

for all $\mathbf{x} \in X$, $\mathbf{y} \in Y$, that is, if

$$\Phi(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) = \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y} \Phi(\mathbf{x}, \mathbf{y}) = \max_{\mathbf{y} \in Y} \min_{\mathbf{x} \in X} \Phi(\mathbf{x}, \mathbf{y}). \tag{47}$$

This means that $\Phi(\mathbf{x}, \mathbf{y})$ attains at the saddle point $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ its maximum with respect to $\mathbf{y}$ for fixed $\widehat{\mathbf{x}}$ and $\Phi(\mathbf{x}, \mathbf{y})$ attains at $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ its minimum with respect to $\mathbf{x}$ for fixed $\widehat{\mathbf{y}}$.

Set

$$\chi(\mathbf{x}) = \max_{\mathbf{y} \in Y} \Phi(\mathbf{x}, \mathbf{y}), \qquad \psi(\mathbf{y}) = \min_{\mathbf{x} \in X} \Phi(\mathbf{x}, \mathbf{y}). \tag{48}$$

Denote by X^* and Y^* the sets of optimal solutions to the optimization problems

$$\min_{\mathbf{x} \in X} \chi(\mathbf{x}),$$
$$\max_{\mathbf{y} \in Y} \psi(\mathbf{y}), \tag{49}$$

respectively, that is,

$$X^* \stackrel{\text{def}}{=} \left\{ \mathbf{x} : \chi(\mathbf{x}) = \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y} \Phi(\mathbf{x}, \mathbf{y}) \right\}$$

$$\equiv \left\{ \mathbf{x} : \max_{\mathbf{y} \in Y} \Phi(\mathbf{x}, \mathbf{y}) = \Phi(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \right\},$$

$$Y^* \stackrel{\text{def}}{=} \left\{ \mathbf{y} : \psi(\mathbf{y}) = \max_{\mathbf{y} \in Y} \min_{\mathbf{x} \in X} \Phi(\mathbf{x}, \mathbf{y}) \right\} \tag{50}$$

$$\equiv \left\{ \mathbf{y} : \min_{\mathbf{x} \in X} \Phi(\mathbf{x}, \mathbf{y}) = \Phi(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \right\}.$$

Let X^*, Y^* be bounded sets. Then,

$$\chi(\mathbf{x}^*) = \psi(\mathbf{y}^*), \quad \mathbf{x}^* \in X^*, \ \mathbf{y}^* \in Y^*, \tag{51}$$

that is,

$$\chi(\mathbf{x}^*) \stackrel{\text{def}}{=} \max_{\mathbf{y} \in Y} \Phi(\mathbf{x}^*, \mathbf{y}) = \psi(\mathbf{y}^*)$$

$$\stackrel{\text{def}}{=} \min_{\mathbf{x} \in X} \Phi(\mathbf{x}, \mathbf{y}^*) = \Phi(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}). \tag{52}$$

This means that $X^* \times Y^*$ is the set of saddle points of $\Phi(\mathbf{x}, \mathbf{y})$ and

$$\Phi(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) = \Phi(\mathbf{x}^*, \mathbf{y}^*). \tag{53}$$

Consider the sets

$$X_{\mathbf{y}^*} \stackrel{\text{def}}{=} \{ \mathbf{x} : \Phi(\mathbf{x}, \mathbf{y}^*) = \Phi(\mathbf{x}^*, \mathbf{y}^*) \},$$
$$Y_{\mathbf{x}^*} \stackrel{\text{def}}{=} \{ \mathbf{y} : \Phi(\mathbf{x}^*, \mathbf{y}) = \Phi(\mathbf{x}^*, \mathbf{y}^*) \}, \tag{54}$$

that is, $X_{\mathbf{y}^*}$ and $Y_{\mathbf{x}^*}$ denote the sets of arguments of $\Phi(\mathbf{x}, \mathbf{y})$ with $\mathbf{y} = \mathbf{y}^*$ and $\mathbf{x} = \mathbf{x}^*$, respectively, for which the value of $\Phi(\mathbf{x}, \mathbf{y})$ is equal to its value at the saddle point.

In the general case, $X^* \subset X_{\mathbf{y}^*}$, $Y^* \subset Y_{\mathbf{x}^*}$; that is, the sets $X_{\mathbf{y}^*}$, $Y_{\mathbf{x}^*}$ contain sets X^*, Y^*, respectively.

Definition 20. If $X^* = X_{\mathbf{y}^*}$ and $Y^* = Y_{\mathbf{x}^*}$, then the set of saddle points of $\Phi(\mathbf{x}, \mathbf{y})$ is said to be *stable*.

If the set of saddle points of $\Phi(\mathbf{x}, \mathbf{y})$ is stable, then from

$$\lim_{k \to \infty} \Phi\left(\mathbf{x}^{(k)}, \mathbf{y}^{*(k)}\right) = \Phi(\mathbf{x}^*, \mathbf{y}^*), \quad \mathbf{x}^{(k)} \in X, \ \mathbf{y}^{*(k)} \in Y \tag{55}$$

it follows that

$$\delta\left(\mathbf{x}^{(k)}, X^*\right) \longrightarrow 0, \quad k \longrightarrow \infty, \tag{56}$$

and from

$$\lim_{k \to \infty} \Phi\left(\mathbf{x}^{*(k)}, \mathbf{y}^{(k)}\right) = \Phi(\mathbf{x}^*, \mathbf{y}^*), \quad \mathbf{x}^{*(k)} \in X, \ \mathbf{y}^{(k)} \in Y \tag{57}$$

it follows that

$$\delta\left(\mathbf{y}^{(k)}, Y^*\right) \longrightarrow 0, \quad k \longrightarrow \infty, \tag{58}$$

where $\delta(\mathbf{x}, X) \stackrel{\text{def}}{=} \min_{\mathbf{z} \in X} \|\mathbf{x} - \mathbf{z}\|$ is the distance from $\mathbf{x}$ to the set X. The implications written above mean that convergence of $\Phi(\mathbf{x}, \mathbf{y}^*)$ to $\Phi(\mathbf{x}^*, \mathbf{y}^*)$ with respect to $\mathbf{x}^{(k)}$ and convergence of $\Phi(\mathbf{x}^*, \mathbf{y})$ to $\Phi(\mathbf{x}^*, \mathbf{y}^*)$ with respect to $\mathbf{y}^{(k)}$ implies convergence of sequence $(\{\mathbf{x}^{(k)}\}, \{\mathbf{y}^{(k)}\})$ to the set $X^* \times Y^*$ of saddle points of $\Phi(\mathbf{x}, \mathbf{y})$.

The concept of stability, introduced by Definition 20, is important for constructing iterative gradient algorithms for finding saddle points of the Lagrangian associated with an optimization problem.

The set of saddle points of the Lagrangian associated with the problem

$$\min_{\mathbf{x} \in S} f(\mathbf{x}), \tag{59}$$

$$S = \{ \mathbf{x} \in \mathbb{R}^n : \varphi_i(\mathbf{x}) \leq 0, i = 1, \ldots, m, \mathbf{x} \in X \} \tag{60}$$

is not stable according to Definition 20. Concerning the dual variables this can be proved as follows.

Let rth constraint of (60) be satisfied as an equality at $\mathbf{x}^*$, that is,

$$\varphi_r\left(\mathbf{x}^*\right) = 0 \qquad (61)$$

for some r, $1 \leq r \leq m$. Then, the Lagrangian $L(\mathbf{x}^*, \boldsymbol{\lambda})$ of problem (59)-(60) does not depend on λ_r and therefore $L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = L(\mathbf{x}^*, \boldsymbol{\lambda})$ is satisfied for every λ_r. Hence, it is impossible to determine λ_r by using the relation $L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = L(\mathbf{x}^*, \boldsymbol{\lambda})$.

In order to avoid this difficulty, so-called *modified Lagrangians* are used instead of the "classical" Lagrangian. Modified Lagrangians are usually *nonlinear* functions of $\boldsymbol{\lambda}$ and the set of their saddle points is stable and it coincides, under some assumptions, with the set of saddle points of the "classical" Lagrangian for the same problem. This is important to ensure convergence of iterative gradient algorithms (see, e.g., Gol'shtein [5]).

3.2. About the Stability of the Set of Saddle Points for the Approach Considered in this Paper. Consider problem (C) (problem $(C^=)$ and problem $(C^\geq)$, resp.). Obviously, the Lagrange multiplier λ, associated with the constraint (2) ((10) and (13), resp.), is not involved in the equality

$$L\left(\mathbf{x}^*, \mathbf{u}^*, \mathbf{v}^*, \lambda^*\right) = L\left(\mathbf{x}^*, \mathbf{u}, \mathbf{v}, \lambda\right) \qquad (62)$$

when $\delta(\lambda^*) = 0$, that is, when $\sum_{j \in J} d_j(x_j) = \alpha$. For problem (C), λ^* (>0) is either determined *uniquely* from $\delta(\lambda^*) = 0$ when $\delta(0) > 0$ or we set $\lambda^* := 0$ when $\delta(0) \leq 0$ (Algorithm 2). Although the set of saddle points of the Lagrangian $L(\mathbf{x}, \mathbf{u}, \mathbf{v}, \lambda)$, associated with problem (C) (problem $(C^=)$ and problem $(C^\geq)$, resp.) is not stable in the sense of Gol'shtein, the specificity of the approach suggested (the algorithms are not of gradient type and λ^* is determined *uniquely* in all cases for the three problems under consideration) overcomes this "weakness" of the classical Lagrangian. That is why it is not necessary to use modified Lagrangians for problems (C), $(C^=)$, and $(C^\geq)$.

On the one hand, we need a closed form expression of λ at Step (2) of the algorithms suggested. However, it is this feature of the algorithms that allows us to use classical Lagrangians instead of modified Lagrangians in the approach suggested. Moreover, the method for finding λ, and, therefore, for finding x_j^*, $j \in J$, in the corresponding problem (C), $(C^=)$, and $(C^\geq)$, is *exact* although it is an iterative method.

As it usually happens, the disadvantage in one aspect turns out to be an advantage in another aspect and vice versa.

All conclusions in this section have been drawn under the assumption that the objective function $c(\mathbf{x})$ and the constraint function(s) $d_i(\mathbf{x})$ of the three problems under consideration (C), $(C^=)$, $(C^\geq)$ are *nondegenerate*, that is, $\mathbf{c}'(\mathbf{x}^*) \not\equiv \mathbf{0}$, $\mathbf{d}_i'(\mathbf{x}^*) \not\equiv \mathbf{0}$; otherwise, the application of the Karush-Kuhn-Tucker theorem with differentiability is void of meaning.

Some optimality criteria for degenerate mathematical programs are given, for example, in the book of Karmanov [6].

4. Primal-Dual Analysis

Some of the main characteristics of the approach, suggested for solving problems (C), $(C^=)$, and $(C^\geq)$, are following.

Since the method, proposed for problem (C), uses values of the first derivatives of functions $c_j(x_j)$, $j \in J$, we can consider it as a *first-order method*. Also, this method is a *saddle point method* or, more precisely, a *dual variables saddle point method* because it is based on convergence with respect to the Lagrange multiplier (dual variable) λ associated with the single constraint (2).

At Step (2) of Algorithm 2, we use the expression of $\lambda^{(k)}$, calculated from the equality $\delta(\lambda^{(k)}) = 0$, where x_j^* are determined from (6), $j \in J^{\lambda(k)} = J^{(k)}$. As it was proved, under the assumptions for problem (C), we can always determine $\lambda = \lambda(\mathbf{x}^*)$ from $\delta(\lambda) = 0$ as an implicit function of $\mathbf{x}^*$. For example, when $d_j(x_j)$, $j \in J$, are linear functions, the explicit expression of λ is always available for Algorithm 2. There are also many other examples of functions, for which it is possible to obtain closed form expressions of λ, and, therefore, the suggested approach is applicable and gives good results.

Analogous commentary is valid for the method suggested for solving problem $(C^=)$.

When the (optimal) Lagrange multiplier λ^* associated with (2) is known, then problem (C) (1)–(3) can be replaced by the following separable convex optimization problem

$$\min \quad \left\{ \sum_{j \in J} \left[c_j\left(x_j\right) + \lambda^* d_j\left(x_j\right) \right] - \lambda^* \alpha \right\} \qquad (63)$$

subject to $\quad \mathbf{x} \in A \stackrel{\text{def}}{=} \left\{ \mathbf{x} \in \mathbb{R}^n : a_j \leq x_j \leq b_j, j \in J \right\}.$

The problem, dual to problem (C), is

$$\max \quad \Psi(\lambda)$$
$$\text{subject to} \quad \lambda \in \mathbb{R}_+^1, \qquad (64)$$

where

$$\Psi(\lambda) = \min_{\mathbf{x} \in A} \left\{ \sum_{j \in J} \left[c_j\left(x_j\right) + \lambda d_j\left(x_j\right) \right] - \lambda \alpha \right\}. \qquad (65)$$

Problem $(C^=)$ can be considered similarly; $d_j(x_j) = d_j x_j$ and $\lambda \in \mathbb{R}^1$ for it.

Thus, using the Lagrangian duality and Theorem 1 for problem (C) (Theorem 4 for problem $(C^=)$) we have replaced the *multivariate* problem (C) (problem $(C^=)$) of $\mathbf{x} \in \mathbb{R}^n$ by the *single-variable* optimization problem for finding $\lambda \in \mathbb{R}_+^1$ ($\lambda \in \mathbb{R}^1$, resp.).

Since Algorithm 8 is based on Theorem 5 and Algorithm 6, and since the "iterative" Steps (2)–(7) of Algorithms 6 and 8 are the same, then primal-dual analysis for problem $(C^\geq)$ is similar to that for problems (C) and $(C^=)$.

5. Bibliographical Notes

The definition of Tychonov well-posedness is given by Tychonov in [7]. Tychonov and Hadamard well-posedness

and well-posedness in the generalized sense are considered, for example, in the work of Cavazzuti and Morgan [8], in the book of Dontchev and Zolezzi [9], in works of Hadamard [10, 11], and so forth. Other questions regarding stability, ill-posed problems, and Tychonov regularization can be found in [12–19], and so forth. Sometimes Tychonov is written as Tykhonov, Tikhonov or Tychonoff in references. Well-posedness is also discussed in the book of Rockafellar and Wets [20].

Calmness in the sense of Clarke is proposed and studied in the works of Clarke [3, 21, 22]. Concerning Theorem 14, see also the paper of Roberts and Varberg [4].

Stability of the set of saddle points of Lagrangians is considered in the papers of Gol'shtein [5], Gol'shtein and Tret'iakov [23], and so forth.

Various convex separable optimization problems are considered and convergent polynomial algorithms for solving them are proposed in Stefanov [1, 2, 24–29].

References

[1] S. M. Stefanov, "Convex separable minimization subject to bounded variables," *Computational Optimization and Applications*, vol. 18, no. 1, pp. 27–48, 2001.

[2] S. M. Stefanov, "Convex separable minimization problems with a linear constraint and bounded variables," *International Journal of Mathematics and Mathematical Sciences*, no. 9, pp. 1339–1363, 2005.

[3] F. H. Clarke, *Optimization and Nonsmooth Analysis*, vol. 5 of *Classics in Applied Mathematics*, SIAM, Philadelphia, Pa, USA, 1990.

[4] A. W. Roberts and D. E. Varberg, "Another proof that convex functions are locally Lipschitz," *The American Mathematical Monthly*, vol. 81, pp. 1014–1016, 1974.

[5] E. G. Gol'shtein, "Generalized gradient method for finding saddle points," *Economics and Mathematical Methods*, vol. 8, no. 4, pp. 569–579, 1972 (Russian).

[6] V. G. Karmanov, *Mathematical Programming*, Mir, Moscow, Russia, 1989.

[7] A. N. Tychonov, "On the stability of the functional optimization problems," *USSR Computational Mathematics and Mathematical Physics*, vol. 6, no. 4, pp. 631–634, 1966 (Russian).

[8] E. Cavazzuti and J. Morgan, "Well-posed saddle point problems," in *Optimization: Theory and Algorithms*, J. B. Hiriart-Urruty, W. Oettli, and J. Stoer, Eds., vol. 86 of *Lecture Notes in Pure and Applied Mathematics*, pp. 61–76, Marcel Dekker, New York, NY, USA, 1983.

[9] A. L. Dontchev and T. Zolezzi, *Well-Posed Optimization Problems*, vol. 1543 of *Lecture Notes in Mathematics*, Springer, Berlin, Germany, 1993.

[10] J. Hadamard, "Sur les problemes aux dérives partielles et leur signification physique," *Bulletin Princeton University*, vol. 13, pp. 49–52, 1902.

[11] J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Dover, New York, NY, USA, 1953.

[12] A. N. Tychonov, "Stability of inverse problems," *Doklady Akademii Nauk SSSR*, vol. 39, no. 5, pp. 195–198, 1943 (Russian).

[13] A. N. Tihonov, "On the solution of ill-posed problems and the method of regularization," *Doklady Akademii Nauk SSSR*, vol. 151, no. 3, pp. 501–504, 1963 (Russian).

[14] A. N. Tihonov, "On the regularization of ill-posed problems," *Doklady Akademii Nauk SSSR*, vol. 153, no. 1, pp. 49–52, 1963 (Russian).

[15] A. N. Tihonov, "Regularisation methods for optimal control problems," *Doklady Akademii Nauk SSSR*, vol. 162, no. 4, pp. 763–765, 1965 (Russian).

[16] A. N. Tihonov, "ll-posed problems of optimal planning and stable methods for solving them," *Doklady Akademii Nauk SSSR*, vol. 164, no. 3, pp. 507–510, 1965 (Russian).

[17] A. N. Tychonov, "Ill-posed optimal planning problems," *USSR Computational Mathematics and Mathematical Physics*, vol. 6, no. 1, pp. 81–89, 1966 (Russian).

[18] A. N. Tychonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, John Wiley & Sons, New York, NY, USA, 1977.

[19] A. N. Tychonov and V. Y. Arsenin, *Methods for Solving Ill-Posed Problems*, Nauka, Moscow, Russia, 1986.

[20] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, vol. 317 of *Fundamental Principles of Mathematical Sciences*, Springer, Berlin, Germany, 1998.

[21] F. H. Clarke, *Necessary conditions for nonsmooth problems in optimal control and the calculus of variations [Ph.D. thesis]*, University of Washington, Seattle, Wash, USA, 1973.

[22] F. H. Clarke, "A new approach to Lagrange multipliers," *Mathematics of Operations Research*, vol. 1, no. 2, pp. 165–174, 1976.

[23] E. G. Gol'shtein and N. V. Tret'iakov, "Modified Lagrangians," *Economics and Mathematical Methods*, vol. 10, no. 3, pp. 568–591, 1974 (Russian).

[24] S. M. Stefanov, "Method for solving a convex integer programming problem," *International Journal of Mathematics and Mathematical Sciences*, no. 44, pp. 2829–2834, 2003.

[25] S. M. Stefanov, "Convex quadratic minimization subject to a linear constraint and box constraints," *Applied Mathematics Research eXpress*, no. 1, pp. 17–42, 2004.

[26] S. M. Stefanov, "Polynomial algorithms for projecting a point onto a region defined by a linear constraint and box constraints in $\mathbb{R}^n$," *Journal of Applied Mathematics*, no. 5, pp. 409–431, 2004.

[27] S. M. Stefanov, "An efficient method for minimizing a convex separable logarithmic function subject to a convex inequality constraint or linear equality constraint," *Journal of Applied Mathematics and Decision Sciences*, vol. 2006, Article ID 89307, 19 pages, 2006.

[28] S. M. Stefanov, "Minimization of a convex linear-fractional separable function subject to a convex inequality constraint or linear equality constraint and bounds on the variables," *Applied Mathematics Research eXpress*, vol. 2006, no. 4, Article ID 36581, 24 pages, 2006.

[29] S. M. Stefanov, "Solution of some convex separable resource allocation and production planning problems with bounds on the variables," *Journal of Interdisciplinary Mathematics*, vol. 13, no. 5, pp. 541–569, 2010.