

Research Article

On the Application of Iterative Methods of Nondifferentiable Optimization to Some Problems of Approximation Theory

Stefan M. Stefanov

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

Correspondence should be addressed to Stefan M. Stefanov; stefm@swu.bg

Received 30 September 2014; Accepted 13 November 2014; Published 27 November 2014

Academic Editor: Peng-Yeng Yin

Copyright © 2014 Stefan M. Stefanov. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We consider the data fitting problem, that is, the problem of approximating a function of several variables, given by tabulated data, and the corresponding problem for inconsistent (overdetermined) systems of linear algebraic equations. Such problems, connected with measurement of physical quantities, arise, for example, in physics, engineering, and so forth. A traditional approach for solving these two problems is the discrete least squares data fitting method, which is based on discrete ℓ_2 -norm. In this paper, an alternative approach is proposed: with each of these problems, we associate a nondifferentiable (nonsmooth) unconstrained minimization problem with an objective function, based on discrete ℓ_1 - and/or ℓ_{∞} -norm, respectively; that is, these two norms are used as proximity criteria. In other words, the problems under consideration are solved by minimizing the residual using these two norms. Respective subgradients are calculated, and a subgradient method is used for solving these two problems. The emphasis is on implementation of the proposed approach. Some computational results, obtained by an appropriate iterative method, are given at the end of the paper. These results are compared with the results, obtained by the iterative gradient method for the corresponding "differentiable" discrete least squares problems, that is, approximation problems based on discrete ℓ_2 -norm.

1. Introduction: Statement of Problems under Consideration

1.1. Problem Number 1. Let $f : \mathbb{R}^p \to \mathbb{R}$ be a real-valued function in p real variables and let the following tabulated data be given:

$$\frac{\mathbf{x} \in \mathbb{R}^{p} | \mathbf{x}_{1} \mathbf{x}_{2} \cdots \mathbf{x}_{m}}{f(\mathbf{x}) | f(\mathbf{x}_{1}) f(\mathbf{x}_{2}) \cdots f(\mathbf{x}_{m})}.$$
(1)

Find a generalized polynomial $P_n(\mathbf{x}) = \sum_{j=0}^n a_j \varphi_j(\mathbf{x})$, based on the system of linearly independent functions $\{\varphi_j(\mathbf{x})\}_{j=0}^n$, that is, a polynomial of generalized degree *n*, which approximates function $f(\mathbf{x})$ with respect to some distance (norm). Depending on the distance (norm) used, $P_n(\mathbf{x})$ is an optimal solution to various problems. In this paper, we discuss the approximation with respect to weighted discrete ℓ_1 -norm

$$\|f\|_{\ell_1} = \sum_{i=1}^{m} w_i |f(\mathbf{x}_i)|$$
 (2)

and weighted discrete ℓ_{∞} -norm

$$\|f\|_{\ell_{\infty}} = \sup_{1 \le i \le m} w_i |f(\mathbf{x}_i)|, \qquad (3)$$

and, only for comparison, a weighted discrete ℓ_2 -norm ("discrete least squares norm")

$$\|f\|_{\ell_2} = \left(\sum_{i=1}^m w_i f^2(\mathbf{x}_i)\right)^{1/2},\tag{4}$$

where $w_i > 0$, $i = 1, \ldots, m$, are weights.

In order to ensure uniqueness of the solution to problems under consideration, it is known that the following condition must be satisfied: $m \ge n + 1$. This requirement means that we must have at least n + 1 values of f, where n + 1 is the number of the unknown coefficients a_0, a_1, \ldots, a_n of the generalized polynomial $P_n(\mathbf{x})$. Thus, the polynomial $P_n(\mathbf{x})$ of best approximation to function $f(\mathbf{x})$ with respect to ℓ_1 -norm is an optimal solution to the minimization problem

$$\min_{\mathbf{a}} \left\{ \Phi_{1}\left(\mathbf{a}\right) \stackrel{\text{def}}{=} \left\| P_{n} - f \right\|_{\ell_{1}} = \sum_{i=1}^{m} w_{i} \left| \sum_{j=0}^{n} a_{j} \varphi_{j}\left(\mathbf{x}_{i}\right) - f\left(\mathbf{x}_{i}\right) \right| \right\},$$
(5)

and the best approximation to $f(\mathbf{x})$ with respect to ℓ_{∞} -norm is an optimal solution to the minimization problem

$$\min_{\mathbf{a}} \left\{ \Phi_{2}\left(\mathbf{a}\right)^{\text{def}} \| P_{n} - f \|_{\ell_{\infty}} = \max_{1 \le i \le m} w_{i} \left| \sum_{j=0}^{n} a_{j} \varphi_{j}\left(\mathbf{x}_{i}\right) - f\left(\mathbf{x}_{i}\right) \right| \right\}.$$
(6)

The corresponding discrete least squares data fitting problem, which can be associated with (1), is

$$\min_{\mathbf{a}} \left\{ \Phi_{3}\left(\mathbf{a}\right) \stackrel{\text{def}}{=} \left\| P_{n} - f \right\|_{\ell_{2}} = \left(\sum_{i=1}^{m} w_{i} \left(\sum_{j=0}^{n} a_{j} \varphi_{j}\left(\mathbf{x}_{i}\right) - f\left(\mathbf{x}_{i}\right) \right)^{2} \right)^{1/2} \right\}.$$
(7)

Here, $\mathbf{a} = (a_0, a_1, \dots, a_n) \in \mathbb{R}^{n+1}$. When the tabulated data is given by the same confidence (reliability) for all $i = 1, \dots, m$, then w_i are chosen to be equal to 1 for all $i = 1, \dots, m$.

Recall that the system of functions $\{\varphi_0, ..., \varphi_n\}$ is said to be *linearly independent* if whenever

$$c_0\varphi_0\left(\mathbf{x}\right) + c_1\varphi_1\left(\mathbf{x}\right) + \dots + c_n\varphi_n\left(\mathbf{x}\right) = 0, \qquad (8)$$

then $c_0 = c_1 = \cdots = c_n = 0$. Otherwise, the set of functions is said to be *linearly dependent*.

For the problems under consideration, the system $\{\varphi_j(\mathbf{x})\}_{j=0}^n$ of linearly independent functions can be chosen as follows:

$$\varphi_j(\mathbf{x}) = \sum_{k=1}^p x_k^j, \quad j = 0, 1, \dots, n;$$
 (9)

that is,

٢

$$\varphi_{0} (\mathbf{x}) = 1 + 1 + \dots + 1 = p,$$

$$\varphi_{1} (\mathbf{x}) = x_{1} + x_{2} + \dots + x_{p},$$

$$\varphi_{2} (\mathbf{x}) = x_{1}^{2} + x_{2}^{2} + \dots + x_{p}^{2},$$

$$\vdots$$

(10)

$$\varphi_n(\mathbf{x}) = x_1^n + x_2^n + \dots + x_p^n.$$

It is proved that functions $\{\varphi_j(\mathbf{x})\}_{j=0}^n$, defined by (10), are linearly independent (Theorem 1, Section 2.1).

1.2. Problem Number 2. Given an inconsistent (overdetermined) system of linear algebraic equations,

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = 1, \dots, m; \ m > n.$$
(11)

This system does not have a solution in the general case when m > n and all the equations are linearly independent.

We can associate the following minimization problems with (11):

$$\min\left\{\Phi_{4}\left(\mathbf{x}\right)\stackrel{\text{def}}{=}\max_{1\leq i\leq m}w_{i}\left|\sum_{j=1}^{n}a_{ij}x_{j}-b_{i}\right|\right\}$$
(12)

or

$$\min\left\{\Phi_{5}\left(\mathbf{x}\right)\stackrel{\text{def}}{=}\sum_{i=1}^{m}w_{i}\left|\sum_{j=1}^{n}a_{ij}x_{j}-b_{i}\right|\right\},\qquad(13)$$

where $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$. The corresponding discrete least squares data fitting problem, which can be associated with (11), is

$$\min\left\{\Phi_{6}\left(\mathbf{x}\right) \stackrel{\text{def}}{=} \left(\sum_{i=1}^{m} w_{i} \left(\sum_{j=1}^{n} a_{ij} x_{j} - b_{i}\right)^{2}\right)^{1/2}\right\}.$$
 (14)

Problem (12) is a special case of the problem of best Chebyshev approximation, based on ℓ_{∞} -norm (3), and problem (13) is based on ℓ_1 -norm.

Approximations with respect to ℓ_1 -norm are known as ℓ_1 -(or *absolute deviation*) *approximations*, and approximations with respect to ℓ_{∞} -norm, as *Chebyshev*, or *minimax*, or *uniform*, *approximations*.

1.3. Bibliographical Notes and Organization of the Paper. Problems like (1) and (11), connected with measurement of physical quantities, arise, for example, in physics, engineering, and so forth. The weights w_i , i = 1, ..., m, mean the reliability, with which each value (measurement, empirical datum) $f(\mathbf{x}_i)$ at \mathbf{x}_i for Problem Number 1, or each equation for Problem Number 2, can be accepted.

Problems, discussed in this paper and related to them, are considered in [1–32] and so forth.

The ℓ_1 -approximation is considered, for example, in papers of Barrodale and Roberts [2, 3] and Coleman and Li [13], and the ℓ_1 -solution to overdetermined linear systems is discussed in Bartels et al. [4]. ℓ_p -approximations are considered, for example, in papers of Calamai and Conn [9], Fischer [15], Li [19], Merle and Späth [21], Watson [27], Wolfe [28], and so forth. A global quadratically convergent method for linear ℓ_{∞} problems is suggested in the paper of Coleman and Li [12].

Papers of Andersen [1], Calamai and Conn [8], Overton [22], and Xue and Ye [29] consider minimization of sum of Euclidean norms.

Books of Clarke [11] and Demyanov and Vasiliev [14] are devoted to nondifferentiable optimization and book of

Korneichuk [17] is devoted to optimization problems of the approximation theory.

Numerical methods for best Chebyshev approximation are suggested, for example, in the book of Remez [23].

A subgradient algorithm for certain minimax and minisum problems is suggested in the paper of Chatelon et al. [10].

Least squares approach is discussed by Bertsekas [6], Björck [7], Lawson and Hanson [18], and so forth.

A quasi-Newton approach to nonsmooth convex optimization problems in machine learning is considered in Yu et al. [31]. Nonsmooth optimization methods in the problems of constructing a linear classifier are proposed in Zhuravlev et al. [32].

Polynomial algorithms for projecting a point onto a region defined by a linear constraint and box constraints in \mathbb{R}^n are proposed in Stefanov [25], and well-posedness and primal-dual analysis of some convex separable optimization problems is considered in Stefanov [26].

Rest of the paper is organized as follows. In Section 2, some results for calculation of subgradients of particular types of functions are formulated and proved, and solvability of the problems under consideration is analyzed. In Section 3, the iterative subgradient method for solving nondifferentiable unconstrained optimization problems is formulated and its convergence is proved. In Section 4, results of computational experiments are presented. In Section 5, *Conclusions*, the proposed approach and the obtained computational results are discussed. In the appendix, some known propositions, used in the paper, are formulated without proofs, and only for comparison purposes, the iterative gradient method for solving differentiable unconstrained optimization problems is presented and convergence theorem is formulated.

2. Preliminaries

2.1. Theoretical Matters. Some known results, called propositions, which are used in subsequent sections, are recalled without proofs in Appendix A.1 at the end of the paper.

We prove below some results, which guarantee solvability of considered problems (Theorem 1, combined with Proposition A.11 of the appendix) and which are used for calculating subgradients in Section 3.2 (Theorems 2 and 3).

Theorem 1 (linear independence of a system of multivariate functions). If $\varphi_j(\mathbf{x})$ is a polynomial of degree j for each j = 0, 1, ..., n, then the set of functions $\{\varphi_0, ..., \varphi_n\}$ is linearly independent.

Proof. Let c_0, \ldots, c_n be real numbers such that

$$P(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{j=0}^{n} c_{j} \varphi_{j}(\mathbf{x}) = 0.$$
(15)

Since the generalized polynomial $P(\mathbf{x})$ of degree *n* vanishes, the coefficients of x_k^n , k = 1, ..., p, are equal to zero.

Since $c_n \varphi_n(\mathbf{x})$ is the only term in $P(\mathbf{x})$, containing x_k^n , then c_n must be equal to zero. Therefore

$$P(\mathbf{x}) = \sum_{j=0}^{n-1} c_j \varphi_j(\mathbf{x}).$$
(16)

In this representation of $P(\mathbf{x})$, the only term that contains powers of x_k^{n-1} is $c_{n-1}\varphi_{n-1}(\mathbf{x})$. Hence, we must have $c_{n-1} = 0$, and

$$P(\mathbf{x}) = \sum_{j=0}^{n-2} c_j \varphi_j(\mathbf{x}).$$
(17)

Continuing in this way, we obtain that the remaining coefficients $c_{n-2}, \ldots, c_1, c_0$ are also equal to zero. Therefore, the functions $\{\varphi_0, \ldots, \varphi_n\}$ are linearly independent by definition.

The following two theorems give the rules for calculating subgradients for some types of functions.

Theorem 2 (subgradient of a sum of univariate convex functions). Let $f(x_1, ..., x_n) = \sum_{j=1}^n f_j(x_j)$, $f_j(x_j)$ be a convex function of x_j for each j = 1, ..., n. Then $\hat{\mathbf{f}}(\bar{\mathbf{x}}) = (\hat{f}_{x_1}(\bar{x}_1), ..., \hat{f}_{x_n}(\bar{x}_n))$, where

$$\widehat{f}_{x_j}\left(\overline{x}_j\right) = \lambda_j f_j^+\left(\overline{x}_j\right) + \left(1 - \lambda_j\right) f_j^-\left(\overline{x}_j\right),$$

$$0 \le \lambda_j \le 1, \quad j = 1, \dots, n,$$
(18)

 $f_j^+(\overline{x}_j) = \lim_{\epsilon \to 0} f_j'(\overline{x}_j + \epsilon), f_j^-(\overline{x}_j) = \lim_{\epsilon \to 0} f_j'(\overline{x}_j - \epsilon)$ are the derivatives of f_j on the right and on the left at \overline{x}_j , respectively, and $\widehat{\mathbf{f}}(\overline{\mathbf{x}})$ denotes the subgradient of function f at point $\overline{\mathbf{x}}$.

Proof. Since convex functions have derivatives on the right and on the left at each interior feasible point, then we can assume that $f_i^+(\overline{x}_i)$ and $f_i^-(\overline{x}_i)$ exist.

According to Proposition A.8 of the appendix, about the vectors $\mathbf{f}^+(\mathbf{x})$ and $\mathbf{f}^-(\mathbf{x})$ of the right and the left derivatives of f at \mathbf{x} , respectively, we have

$$f(\mathbf{y}) - f(\mathbf{x}) \ge \langle \mathbf{f}^{\pm}(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle; \qquad (19)$$

that is, $\mathbf{f}^{\pm}(\mathbf{x}) \in \partial f(\mathbf{x})$ by the definition of subgradient.

Since the subdifferential $\partial f(\mathbf{x})$ of a convex function $f(\mathbf{x})$ is a nonempty, convex, and compact set, and since $f_j^+(\overline{x}_j), f_j^-(\overline{x}_j) \in \partial f(\overline{x}_j)$ according to the above discussion, then

$$\lambda_{j}f_{j}^{+}\left(\overline{x}_{j}\right)+\left(1-\lambda_{j}\right)f_{j}^{-}\left(\overline{x}_{j}\right)\in\partial f_{j}\left(\overline{x}_{j}\right), \quad j=1,\ldots,n.$$
(20)

Therefore

$$f(\mathbf{x}) - f(\overline{\mathbf{x}}) \equiv \sum_{j=1}^{n} f_j(x_j) - \sum_{j=1}^{n} f_j(\overline{x}_j)$$
$$\equiv \sum_{j=1}^{n} \left[f_j(x_j) - f_j(\overline{x}_j) \right]$$

$$\geq \sum_{j=1}^{n} \left[\lambda_{j} f_{j}^{+} \left(\overline{x}_{j} \right) + \left(1 - \lambda_{j} \right) f_{j}^{-} \left(\overline{x}_{j} \right) \right] \left(x_{j} - \overline{x}_{j} \right)$$
$$\equiv \left\langle \widehat{\mathbf{f}} \left(\overline{\mathbf{x}} \right), \mathbf{x} - \overline{\mathbf{x}} \right\rangle; \tag{21}$$

that is, $\hat{f}(\overline{\mathbf{x}}) = (\hat{f}_{x_1}(\overline{x}_1), \dots, \hat{f}_{x_n}(\overline{x}_n))$ with $\hat{f}_{x_j}(\overline{x}_j)$, defined above, is a subgradient of $f(x_1, \dots, x_n)$ at $\overline{\mathbf{x}}$ by definition. \Box

Theorem 3 (subgradient of a function in two variables). Let f(x, y) be a convex function of x for each y, let there exist a y(x) such that

$$f(x) \stackrel{\text{def}}{=} \max_{y \in Y} f(x, y) = f(x, y(x))$$
(22)

and let the subgradient $\hat{f}_x(x, y)$ of f(x, y) with respect to x be known for each y. Then

$$\widehat{f}_{x}(x) = \widehat{f}_{x}(x, y)\big|_{y=y(x)}.$$
(23)

Proof. Since f(x, y) is a convex function of x for each y and since y(x) is an optimal solution to problem $\max_{y \in Y} f(x, y)$, then

$$f(z) - f(x) \equiv f(z, y(z)) - f(x, y(x))$$

$$\geq f(z, y(x)) - f(x, y(x)) \qquad (24)$$

$$\geq \left\langle \widehat{f}_x(x, y(x)), z - x \right\rangle.$$

Therefore $\hat{f}_x(x) = \hat{f}_x(x, y)|_{y=y(x)}$ according to definition of subgradient.

2.2. Some Properties of Objective Functions and Solvability of the Problems under Consideration. Functions $\sum_{j=0}^{n} a_j \varphi_j(\mathbf{x}_i) - f(\mathbf{x}_i), i = 1, ..., m$, are linear functions of $a_j, j = 0, 1, ..., n$; the "absolute value" function $|f| = \max\{f, -f\}$ is convex (Proposition A.4 of the appendix) when f is a linear function (and therefore, both f and -f are convex). Hence, $\Phi_1(\mathbf{a})$ is a convex function of \mathbf{a} as a linear combination with nonnegative coefficients $w_i, i = 1, ..., m$, of convex functions (Proposition A.3 of the appendix).

Using the same reasoning, we obtain that $\Phi_5(\mathbf{x})$ is a convex function of \mathbf{x} .

According to Proposition A.4 of the appendix, $\Phi_2(\mathbf{a})$ is a convex function of \mathbf{a} as maximum of the convex functions $w_i | \sum_{j=0}^n a_j \varphi_j(\mathbf{x}_i) - f(\mathbf{x}_i) |$, where $w_i > 0$, because functions $\sum_{j=0}^n a_j \varphi_j(\mathbf{x}_i) - f(\mathbf{x}_i)$ are both convex and concave as linear functions of a_j , i = 1, ..., m; j = 0, ..., n. Also, $\Phi_4(\mathbf{x})$ is a convex function of x_j , j = 1, ..., n, because of similar reasons.

Function $\Phi_3^2(\mathbf{a})$ is a strictly convex function of \mathbf{a} as a linear combination with nonnegative coefficients w_i , i = 1, ..., m, of the quadratic functions $(\sum_{j=0}^n a_j \varphi_j(\mathbf{x}_i) - f(\mathbf{x}_i))^2$, i = 1, ..., m, which, as it is known, are strictly convex.

Similarly, $\Phi_6^2(\mathbf{x})$ is a strictly convex function of \mathbf{x} .

Functions $\Phi_1(\mathbf{a})$, $\Phi_2(\mathbf{a})$, $\Phi_4(\mathbf{x})$, and $\Phi_5(\mathbf{x})$ are nondifferentiable (nonsmooth) whereas functions $\Phi_3(\mathbf{a})$ and $\Phi_6(\mathbf{x})$ are differentiable.

Functions $\Phi_1(\mathbf{a})$, $\Phi_2(\mathbf{a})$, $\Phi_3(\mathbf{a})$, $\Phi_4(\mathbf{x})$, $\Phi_5(\mathbf{x})$, and $\Phi_6(\mathbf{x})$ are *separable functions*; that is, these functions can be expressed as the sums of single-variable (univariate) functions, which follows from definitions of these six functions.

2.2.1. On Problems Associated with Problem Number 1 (1). Since (5) is a minimization problem, $\Phi_1(\mathbf{a})$ is a continuous (and, therefore, both lower and upper semicontinuous) function, bounded from below from 0 as a sum of nonnegative terms, and $\Phi_1(\mathbf{a}) \rightarrow +\infty$ as $\|\mathbf{a}\| \rightarrow \infty$, then problem (5) has an optimal solution according to Corollary A.2 of the appendix with $X = \mathbb{R}^n$.

Using the same reasoning, we can conclude that problems (6) and (7) are also solvable.

Since min $\Phi_3(\mathbf{a})$ and min $\Phi_3^2(\mathbf{a})$ are attained at the same point (vector) $\mathbf{a} = (a_0, a_1, \dots, a_n)$, we can consider problem

$$\min_{\mathbf{a}} \left\{ F_3\left(\mathbf{a}\right) \stackrel{\text{def}}{=} \Phi_3^2\left(\mathbf{a}\right) = \sum_{i=1}^m w_i \left(\sum_{j=0}^n a_j \varphi_j\left(\mathbf{x}_i\right) - f\left(\mathbf{x}_i\right)\right)^2 \right\}$$
(25)

instead of problem (7). Since $\Phi_3^2(\mathbf{a})$ is a strictly convex function, problem (25) has a unique solution (Proposition A.9 of the appendix).

Existence of solutions to these problems can also be proved by using some general results.

As it is known, ℓ_1 , ℓ_2 , and ℓ_{∞} are normed linear spaces; they are Banach spaces with the norms (2), (4), and (3), respectively; ℓ_1 , ℓ_2 are separable spaces and ℓ_{∞} is not a separable space (see, e.g., [16, 30], etc.).

Linear independence of $\{\varphi_j(\mathbf{x})\}_{j=0}^n$, proved in Theorem I, guarantees the existence of an element of best approximation for problems (5), (6), and (7) (Proposition A.11 of the appendix).

Furthermore, since ℓ_p , 1 , are strictly convexspaces, then problem (7) (and problem (25)) has a unique $solution (Proposition A.12 of the appendix), and since <math>\ell_1$ and ℓ_{∞} are not strictly convex spaces, in the general case we cannot conclude uniqueness of the optimal solution to problems (5) and (6).

The n + 1-tuple $\mathbf{a} = (a_0, a_1, \dots, a_n) \in \mathbb{R}^{n+1}$, which is obtained as an optimal solution to problem (5) (problem (6), problem (25), resp.), gives coefficients of the generalized polynomial $P_n(\mathbf{x})$ of best approximation for $f(\mathbf{x})$ (1), $\mathbf{x} \in \mathbb{R}^p$, with respect to ℓ_1 -norm (ℓ_{∞} -norm, ℓ_2 -norm, resp.).

When p = 1, that is, when f(x) is a single-variable (univariate) function, the generalized polynomial $P_n(\mathbf{x}) = \sum_{j=0}^{n} a_j \varphi_j(\mathbf{x})$ becomes an algebraic polynomial of degree *n*: $P_n(x) = \sum_{j=0}^{n} a_j x^j$ and problem (7) (or equivalently (25)) with p = 1 is the well-known discrete least squares data fitting problem.

2.2.2. On Problems Associated with Problem Number 2 (11). Solvability of problems (12), (13), and (14) follows from Corollary A.2 of the appendix: using that $\Phi_4(\mathbf{x}) \ge 0$, $\Phi_5(\mathbf{x}) \ge 0$ are continuous functions, a_{ij} , i = 1, ..., m, j = 1, ..., n, and

 b_i , i = 1, ..., m, are coefficients given by (11), then it follows that $\Phi_4(\mathbf{x}) \to +\infty, \Phi_5(\mathbf{x}) \to +\infty$ when $\|\mathbf{x}\| \to \infty$.

In addition, using the same reasoning, the following problem

$$\min \Phi_6^2(\mathbf{x}) \tag{26}$$

has also an optimal solution and it is unique (Proposition A.9 of the appendix) because $\Phi_6^2(\mathbf{x})$ is a strictly convex function. Existence and uniqueness of the optimal solution to problem (2) can also be proved by using an approach, similar to the alternative approach for problem (25).

Propositions A.7 and A.10 of the appendix imply that \mathbf{a}^* is an optimal solution to problem (6) if and only if

$$\mathbf{0}_{n+1} \in \partial \Phi_2\left(\mathbf{a}^*\right) = \operatorname{co}\left\{\partial F_{i2}\left(\mathbf{a}\right) : i \in I\left(\mathbf{a}\right)\right\}, \qquad (27)$$

where

$$F_{i2} \left(\mathbf{a} \right) = w_i \left| \sum_{j=0}^n a_j \varphi_j \left(\mathbf{x}_i \right) - f \left(\mathbf{x}_i \right) \right|,$$

$$I \left(\mathbf{a} \right) = \left\{ i \in \{1, \dots, m\} : F_{i2} \left(\mathbf{a} \right) = \Phi_2 \left(\mathbf{a} \right) \right\}.$$
(28)

Similarly, \mathbf{x}^* is an optimal solution to problem (12) if and only if

$$\mathbf{0}_{n} \in \partial \Phi_{4}\left(\mathbf{x}^{*}\right) = \operatorname{co}\left\{\partial F_{i4}\left(\mathbf{x}\right) : i \in I\left(\mathbf{x}\right)\right\}, \qquad (29)$$

where

$$F_{i4}\left(\mathbf{x}\right) = w_{i} \left| \sum_{j=1}^{n} a_{ij} x_{j} - b_{i} \right|, \qquad (30)$$

$$I(\mathbf{x}) = \{i \in \{1, ..., m\} : F_{i4}(\mathbf{x}) = \Phi_4(\mathbf{x})\},\$$

and "co X" denotes the convex hull (convex envelope) of X.

3. Iterative Methods for Solving Problems under Consideration

3.1. The Subgradient Method. Since $\Phi_1(\mathbf{a})$, $\Phi_2(\mathbf{a})$ and $\Phi_4(\mathbf{x})$, $\Phi_5(\mathbf{x})$ are nondifferentiable convex functions, we use the so-called subgradient (generalized gradient) method for solving problems (5), (6), (12), and (13).

Let $f(\mathbf{x})$ be a convex proper function defined on \mathbb{R}^n .

The subgradient method for solving problem

$$\min f\left(\mathbf{x}\right) \tag{31}$$

can be defined as

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \rho_k \gamma_k \widehat{\mathbf{f}}(\mathbf{x}_k), \quad k = 0, 1, \dots,$$
(32)

where $x_0 \in \mathbb{R}^n$ is an arbitrary initial guess (initial approximation); ρ_k is a step size, such that $\rho_k \to +0$ as $k \to \infty$, $\sum_{k=0}^{\infty} \rho_k = +\infty$, $\sum_{k=0}^{\infty} \rho_k^2 < +\infty$; γ_k is a norming multiplier; usually $\gamma_k = 1/\|\widehat{\mathbf{f}}(\mathbf{x}_k)\|$ or $\gamma_k = 1$; $\widehat{\mathbf{f}}(\mathbf{x}_k)$ is a subgradient of f at \mathbf{x}_k .

The following theorem guarantees convergence of the subgradient method (32).

Theorem 4 (convergence of the subgradient method). If $\rho_k \to 0$ when $k \to \infty$, $\rho_k \ge 0$, $\sum_{k=0}^{\infty} \rho_k = +\infty$, $\gamma_k = 1$ for all k and $\|\widehat{\mathbf{f}}(\mathbf{x}_k)\| < C = const$ for all \mathbf{x}_k , then there exists a subsequence $\{f(\mathbf{x}_{k_s})\}$ of the sequence $\{f(\mathbf{x}_k)\}$ such that $\lim_{s\to\infty} f(\mathbf{x}_{k_s}) = f(\mathbf{x}^*)$, where $\mathbf{x}^* \in M^*$, $M^* \stackrel{\text{def}}{=} \{\mathbf{x}^* \in \mathbb{R}^n : f(\mathbf{x}^*) = \inf_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})\}$.

Proof. By the assumptions of Theorem 4 we have that

$$\begin{aligned} \left\| \mathbf{x}^{*} - \mathbf{x}_{k+1} \right\|^{2} &= \left\| \mathbf{x}^{*} - \mathbf{x}_{k} + \rho_{k} \widehat{\mathbf{f}} \left(\mathbf{x}_{k} \right) \right\|^{2} \\ &= \left\| \mathbf{x}^{*} - \mathbf{x}_{k} \right\|^{2} + 2\rho_{k} \left\langle \widehat{\mathbf{f}} \left(\mathbf{x}_{k} \right), \mathbf{x}^{*} - \mathbf{x}_{k} \right\rangle \\ &+ \rho_{k}^{2} \left\| \widehat{\mathbf{f}} \left(\mathbf{x}_{k} \right) \right\|^{2} \\ &\leq \left\| \mathbf{x}^{*} - \mathbf{x}_{k} \right\|^{2} + 2\rho_{k} \left\langle \widehat{\mathbf{f}} \left(\mathbf{x}_{k} \right), \mathbf{x}^{*} - \mathbf{x}_{k} \right\rangle + C^{2} \rho_{k}^{2}. \end{aligned}$$
(33)

Choose some $\delta > 0$. For every k = 0, 1, ..., there are two possible cases:

(1)
$$2\left\langle \widehat{\mathbf{f}}\left(\mathbf{x}_{k}\right),\mathbf{x}^{*}-\mathbf{x}_{k}\right\rangle +C^{2}\rho_{k}\leq-\delta,$$
 (34)

(2)
$$2\left\langle \widehat{\mathbf{f}}\left(\mathbf{x}_{k}\right),\mathbf{x}^{*}-\mathbf{x}_{k}\right\rangle +C^{2}\rho_{k}>-\delta.$$
 (35)

It turns out that there exists a positive integer N such that (35) is satisfied for $k \ge N$. Assume, on the contrary, that (34) is satisfied and $k \ge N$. Then from (33) it follows that

$$\|\mathbf{x}^{*} - \mathbf{x}_{k+1}\|^{2} \leq \|\mathbf{x}^{*} - \mathbf{x}_{k}\|^{2} - \delta\rho_{k}$$

$$\leq (\|\mathbf{x}^{*} - \mathbf{x}_{k-1}\|^{2} - \delta\rho_{k-1}) - \delta\rho_{k}$$

$$\leq \cdots \leq \|\mathbf{x}^{*} - \mathbf{x}_{N}\|^{2} - \delta\sum_{s=N}^{k}\rho_{s}.$$
(36)

The right-hand side of (36) tends to $-\infty$ when $k \to \infty$ because $\sum_{k=0}^{\infty} \rho_k = \infty$ by the assumption, which contradicts $\|\mathbf{x}^* - \mathbf{x}_{k+1}\|^2 \ge 0$. Therefore, there exist sufficiently large numbers k_s , s = 1, 2, ..., such that

$$2\left\langle \widehat{\mathbf{f}}\left(\mathbf{x}_{k_{s}}\right),\mathbf{x}^{*}-\mathbf{x}_{k_{s}}\right\rangle +C^{2}\rho_{k_{s}}>-\delta,$$
(37)

which satisfy (35). Since $\rho_{k_s} \rightarrow 0$, for any $\varepsilon > 0$, a sequence $\{k_s\}$ and a number S_{ε} can be found such that

$$\left\langle \widehat{\mathbf{f}}\left(\mathbf{x}_{k_{s}}\right),\mathbf{x}^{*}-\mathbf{x}_{k_{s}}\right\rangle >-\varepsilon$$
 (38)

is satisfied for $s \ge S_{\varepsilon}$. Moreover, using the property of convex functions (Proposition A.8 of the appendix), we have $f(\mathbf{x}^*) - f(\mathbf{x}_{k_s}) \ge \langle \hat{\mathbf{f}}(\mathbf{x}_{k_s}), \mathbf{x}^* - \mathbf{x}_{k_s} \rangle > -\varepsilon$; that is, $f(\mathbf{x}_{k_s}) - f(\mathbf{x}^*) < \varepsilon$. However, $\mathbf{x}^* = \operatorname{argmin} f(\mathbf{x})$; therefore $f(\mathbf{x}_{k_s}) - f(\mathbf{x}^*) > 0$.

Both inequalities imply $\lim_{s\to\infty} f(\mathbf{x}_{k_s}) = f(\mathbf{x}^*)$.

The subgradient method (32) can be modified for the case of nondifferentiable *constrained* optimization as follows:

$$\mathbf{x}_{k+1} = \Pi_X \left(\mathbf{x}_k - \rho_k \gamma_k \widehat{\mathbf{f}} \left(\mathbf{x}_k \right) \right), \quad k = 0, 1, \dots, \quad (39)$$

where $\Pi_X(\mathbf{y})$ denotes the projection operation of \mathbf{y} onto the feasible region X. This modification is not considered here because the optimization problems, considered in this paper, are unconstrained.

3.2. Calculation of Subgradients. In order to apply the subgradient method for solving the problems under consideration, we have to calculate the corresponding subgradients.

Using that $\Phi_1(\mathbf{a})$, $\Phi_2(\mathbf{a})$, $\Phi_4(\mathbf{x})$, $\Phi_5(\mathbf{x})$ are convex separable functions and statements of Propositions A.5, A.6, and A.7 of the appendix and statements of Theorems 2 and 3, we can calculate corresponding subdifferentials (subgradient sets) at iteration *k* as follows, respectively:

$$\partial \Phi_{1}^{(k)}\left(\mathbf{a}_{k}\right) = \sum_{i=1}^{m} \partial \mathbf{f}_{i}^{(k)}\left(\mathbf{a}_{k}\right) \in \mathbb{R}^{n+1}, \tag{40}$$

where

$$\partial \mathbf{f}_{i}^{(k)}(\mathbf{a}_{k}) = \begin{cases} w_{i} \cdot (p, \varphi_{1}(\mathbf{x}_{i}), \dots, \varphi_{n}(\mathbf{x}_{i})), & \sum_{j=0}^{n} a_{j}^{(k)} \varphi_{j}(\mathbf{x}_{i}) \\ & -f(\mathbf{x}_{i}) > 0, \end{cases}$$

$$= \begin{cases} w_{i} \cdot (p, \varphi_{1}(\mathbf{x}_{i}), \dots, \varphi_{n}(\mathbf{x}_{i})) & \\ \cdot [-1, 1], & \sum_{j=0}^{n} a_{j}^{(k)} \varphi_{j}(\mathbf{x}_{i}) & \\ & -f(\mathbf{x}_{i}) = 0, \end{cases}$$

$$= w_{i} \cdot (p, \varphi_{1}(\mathbf{x}_{i}), \dots, \varphi_{n}(\mathbf{x}_{i})), & \sum_{j=0}^{n} a_{j}^{(k)} \varphi_{j}(\mathbf{x}_{i}) & \\ & -f(\mathbf{x}_{i}) < 0, \end{cases}$$

$$(41)$$

 $i = 1, \ldots, m$.

Let $\max_{1 \le i \le m} w_i | \sum_{j=0}^n a_j \varphi_j(\mathbf{x}_i) - f(\mathbf{x}_i)|$ be attained for $i = i(x) \in I(x)$. Then

$$\partial \Phi_{2}^{(k)} \left(\mathbf{a}_{k} \right) = \operatorname{co} \left\{ \left(\partial f_{0,i(x)}^{(k)} \left(\mathbf{a}_{k} \right), \partial f_{1,i(x)}^{(k)} \left(\mathbf{a}_{k} \right), \dots, \right. \\ \left. \partial f_{n,i(x)}^{(k)} \left(\mathbf{a}_{k} \right) \right) : i\left(x \right) \in I\left(x \right) \right\},$$

$$(42)$$

where

$$\partial f_{j,i(x)}^{(k)} \left(\mathbf{a}_{k} \right) = \begin{cases} w_{i(x)} \varphi_{j} \left(\mathbf{x}_{i(x)} \right), & \sum_{j=0}^{n} a_{j}^{(k)} \varphi_{j} \left(\mathbf{x}_{i(x)} \right) \\ & -f \left(\mathbf{x}_{i(x)} \right) > 0, \\ w_{i(x)} \varphi_{j} \left(\mathbf{x}_{i(x)} \right) \cdot [-1, 1], & \sum_{j=0}^{n} a_{j}^{(k)} \varphi_{j} \left(\mathbf{x}_{i(x)} \right) \\ & -f \left(\mathbf{x}_{i(x)} \right) = 0, \\ -w_{i(x)} \varphi_{j} \left(\mathbf{x}_{i(x)} \right), & \sum_{j=0}^{n} a_{j}^{(k)} \varphi_{j} \left(\mathbf{x}_{i(x)} \right) \\ & -f \left(\mathbf{x}_{i(x)} \right) < 0, \end{cases}$$
(43)

 $j=0,1,\ldots,n.$

Let $\max_{1 \le i \le m} w_i |\sum_{j=1}^n a_{ij} x_j - b_i|$ be attained for $i = i(x) \in I(x)$. Then

$$\partial \Phi_4^{(k)} \left(\mathbf{x}_k \right) = \operatorname{co} \left\{ \left(\partial f_{1,i(x)}^{(k)} \left(\mathbf{x}_k \right), \dots, \partial f_{n,i(x)}^{(k)} \left(\mathbf{x}_k \right) \right) : \\ i \left(x \right) \in I \left(x \right) \right\},$$
(44)

where

$$\partial f_{j,i(x)}^{(k)} \left(\mathbf{x}_{k} \right) = \begin{cases} w_{i(x)} a_{i(x)j}, & \sum_{j=1}^{n} a_{i(x)j} x_{j}^{(k)} - b_{i(x)} > 0, \\ w_{i(x)} a_{i(x)j} \cdot [-1,1], & \sum_{j=1}^{n} a_{i(x)j} x_{j}^{(k)} - b_{i(x)} = 0, \\ -w_{i(x)} a_{i(x)j}, & \sum_{j=1}^{n} a_{i(x)j} x_{j}^{(k)} - b_{i(x)} < 0, \end{cases}$$

$$(45)$$

 $j=1,\ldots,n,$

 $\partial \mathbf{f}^{(k)}(\mathbf{v})$

$$\partial \Phi_{5}^{(k)}\left(\mathbf{x}_{k}\right) = \sum_{i=1}^{m} \partial \mathbf{f}_{i}^{(k)}\left(\mathbf{x}_{k}\right) \in \mathbb{R}^{n}, \tag{46}$$

where

$$= \begin{cases} w_i \cdot (a_{i1}, a_{i2}, \dots, a_{in}), & \sum_{j=1}^n a_{ij} x_j^{(k)} - b_i > 0, \\ w_i \cdot (a_{i1}, a_{i2}, \dots, a_{in}) \cdot [-1, 1], & \sum_{j=1}^n a_{ij} x_j^{(k)} - b_i = 0, \\ -w_i \cdot (a_{i1}, a_{i2}, \dots, a_{in}), & \sum_{j=1}^n a_{ij} x_j^{(k)} - b_i < 0, \end{cases}$$

$$(47)$$

 $i=1,\ldots,m.$

Obviously, elements of $\partial \Phi_1^{(k)}(\mathbf{a}_k)$, $\partial \Phi_2^{(k)}(\mathbf{a}_k)$, $\partial \Phi_4^{(k)}(\mathbf{x}_k)$, $\partial \Phi_5^{(k)}(\mathbf{x}_k)$ depend on the sign of the corresponding expression from (41), (43), (45), and (47), respectively, and therefore they depend on the current values $a_j^{(k)}$, $j = 0, \ldots, n$; $a_j^{(k)}$, $j = 0, \ldots, n$; $x_j^{(k)}$, $j = 1, \ldots, n$; $x_j^{(k)}$, $j = 1, \ldots, n$, respectively. We can choose, for example, $\rho_k = c/k$, $k = 1, 2, \ldots; \rho_0 = 1$; c = const > 0. The requirements for the step size are satisfied for this choice of ρ_k .

4. Computational Experiments

In this section, we present results of some computational experiments, obtained by the subgradient method for problems (5), (6), (12), and (13). As it was pointed out, only for comparison, we give results obtained by the gradient method for solving the least squares problems (7) and (14). Each type of problems was run 30 times. Parameters and data were randomly generated. The computations were performed on an Intel Pentium Dual-Core CPU E5800 3.20 GHz/2.00 GB using RZTools interactive system. For both methods (32) and (A.15), two termination tests are used: an "accuracy" stopping criterion

$$\|\mathbf{x}_{k+1} - \mathbf{x}_{k}\| \equiv \left\{ \begin{array}{l} \left\| \rho_{k} \gamma_{k} \widehat{\mathbf{f}}^{(k)} \left(\mathbf{x}_{k} \right) \right\| \\ \left\| \rho_{k} \mathbf{f}' \left(\mathbf{x}_{k} \right) \right\| \end{array} \right\} < \varepsilon,$$
(48)

where $\varepsilon > 0$ is some given (or chosen) tolerance value, and an upper limit criterion on the number of iterations.

Example 1 (for problem (1)). Consider problem with m = 300, n = 5, p = 10, and $\varepsilon = 10^{-6}$. Results: see Table 1.

Example 2 (for problem (1)). Consider problem with m = 300, n = 5, p = 12, and $\varepsilon = 10^{-6}$. Results: see Table 2.

Example 3 (for problem (1)). Consider problem with m = 300, n = 4, p = 15, and $\varepsilon = 10^{-6}$. Results: see Table 3.

Example 4 (for problem (11)). Consider problem with m = 500, n = 20, and $\varepsilon = 10^{-6}$. Results: see Table 4.

Example 5 (for problem (11)). Consider problem with m = 500, n = 22, and $\varepsilon = 10^{-6}$. Results: see Table 5.

Example 6 (for problem (11)). Consider problem with m = 500, n = 28, and $\varepsilon = 10^{-6}$. Results: see Table 6.

Examples 7 and 8 below present results for simple particular problems of the forms (1) and (11), respectively.

Example 7 (problem (1)). Consider

Results: see Table 7.

Therefore, algebraic polynomials obtained by the two methods are

$$P_2^{(1)}(x) = -0.22x^2 + 2.21x + 1.27,$$
 (50)

$$P_2^{(2)}(x) = -0.975x^2 + 1.986x + 7.904,$$
 (51)

respectively.

Example 8 (Problem (11)). Consider the system of linear equations

 $\begin{aligned} x_1 + x_2 &= 1, \\ x_2 + x_3 &= 1, \\ x_1 + x_3 &= 1, \\ x_1 + x_2 + x_3 &= 1. \end{aligned}$

Results: see Table 8.

	TABLE 1	
By method (32)	By method (32)	By method (A.15)
for problem (5)	for problem (6)	for problem (7)
Iterations 101	Iterations 98	Iterations 97
Run time 0.00045 s	Run time 0.00055 s	Run time 0.00035 s
	TABLE 2	
By method (32)	By method (32)	By method (A.15)
for problem (5)	for problem (6)	for problem (7)
Iterations 103	Iterations 103	Iterations 96
Run time 0.00037 s	Run time 0.000375 s	Run time 0.00038 s
	TABLE 3	
By method (32)	By method (32)	By method (A.15)
for problem (5)	for problem (6)	for problem (7)
Iterations 100	Iterations 105	Iterations 82
Run time 0.00015 s	Run time 0.00017 s	Run time 0.00006 s
	TABLE 4	
By method (32)	By method (32)	By method (A.15)
for problem (12)	for problem (13)	for problem (14)
Iterations 100	Iterations 104	Iterations 108
Run time 0.00065 s	Run time 0.0018 s	Run time 0.0019 s
	Table 5	
By method (32)	By method (32)	By method (A.15)
for problem (12)	for problem (13)	for problem (14)
Iterations 108	Iterations 118	Iterations 111
Run time 0.0048 s	Run time 0.0051 s	Run time 0.0049 s
	Table 6	
By method (32)	By method (32)	By method (A.15)
for problem (12)	for problem (13)	for problem (14)
Iterations 102	Iterations 119	Iterations 101
Run time 0.00375 s	Run time 0.0039 s	Run time 0.0037 s
	Table 7	
By method (32)		By method (A.15)
for problem (5)		for problem (7)
$a_0^* = 1.27$		$a_0^* = 7.904$
$a_1^* = 2.21$		$a_1^* = 1.986$
$a_2^* = -0.22$		$a_2^* = -0.97$

5. Conclusions

 $\Phi_1(\mathbf{a}^*) = 25.4635$

Run time 0.00135 s

Iterations 101

Computational experiments presented above, as well as many other experiments, allow us to conclude that the subgradient method (32), applied for minimizing the nondifferentiable

 $\Phi_3(\mathbf{a}^*) = 12.9625$

Iterations 106

Run time 0.0019 s

	TABLE 8	
By method (32)	By method (32)	By method (A.15)
for problem (12)	for problem (13)	for problem (14)
$x_1^* = 0.3945$	$x_1^* = 0.4999$	$x_1^* = 0.4261$
$x_2^* = 0.4016$	$x_2^* = 0.4999$	$x_2^* = 0.4261$
$x_3^* = 0.3946$	$x_3^* = 0.4999$	$x_3^* = 0.4261$
$\Phi_4(\mathbf{x}^*) = 0.2107$	$\Phi_5(\mathbf{x}^*) = 0.500$	$\Phi_6(\mathbf{x}^*) = 0.3780$
Iterations 101	Iterations 84	Iterations 18
Run time 0.0011 s	Run time 0.0008 s	Run time 0.00165 s

functions $\Phi_1(\mathbf{a})$, $\Phi_2(\mathbf{a})$ and $\Phi_4(\mathbf{x})$, $\Phi_5(\mathbf{x})$, is computationally comparable with the gradient method (A.15), applied to the corresponding "differentiable" problems (25) and (26), based on ℓ_2 -norm, respectively. For some problems, the gradient method gives better results with respect to number of iterations and therefore with respect to run time. However, in many cases it is preferable to approximate with respect to either ℓ_1 -norm (2) or ℓ_{∞} -norm (3) instead of using the ℓ_2 approximation.

Appendix

A. Review of Some Results

A.1. Review of Some Known Results. In this section, some known results, called propositions, used in this paper, are recalled without proofs.

The following Weierstrass theorem and the corollary turn out to be useful concerning solvability of the problems under consideration.

Proposition A.1 (Weierstrass, e.g., [20, Theorem C.4.1]). A lower (upper) semicontinuous function f, defined on a compact set X in \mathbb{R}^n , is bounded from below (above) and attains in X the value

$$\overline{\alpha} = \inf_{\mathbf{x} \in X} f(\mathbf{x}) \qquad \left[\overline{\beta} = \sup_{\mathbf{x} \in X} f(\mathbf{x}) \right]. \tag{A.1}$$

Corollary A.2. Let $X \neq \emptyset$ be a closed set in \mathbb{R}^n and let function $f: X \to \mathbb{R}$ be lower (upper) semicontinuous in X and $\lim_{k\to\infty} f(\mathbf{x}_k) = +\infty$ ($\lim_{k\to\infty} f(\mathbf{x}_k) = -\infty$) for each sequence $\{\mathbf{x}_k\}_{k=1}^{\infty} \in X$ such that $\lim_{k\to\infty} \|\mathbf{x}_k\| = +\infty$. Then f attains on X the value

$$\overline{\alpha} = \inf_{\mathbf{x} \in X} f(\mathbf{x}) \qquad \left[\overline{\beta} = \sup_{\mathbf{x} \in X} f(\mathbf{x})\right]. \tag{A.2}$$

Proposition A.1 and Corollary A.2 mean that, under their assumptions, problem

$$\min_{\mathbf{x}\in X} f(\mathbf{x}) \qquad \left[\max_{\mathbf{x}\in X} f(\mathbf{x}) \right] \tag{A.3}$$

has a minimim [maximum] solution.

Since a continuous function is both lower and upper semicontinuous, then Proposition A.1 and Corollary A.2 are also valid for continuous functions. **Proposition A.3** (nonnegative linear combinations of convex and concave functions, [20, Theorem 4.1.6]). Let f_i , i = 1, ..., m, be numerical functions defined on $X \subseteq \mathbb{R}^n$. If f_i are convex (concave), then each linear combination with nonnegative coefficients of these functions

$$f(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{i=1}^{m} \alpha_i f_i(\mathbf{x}), \quad \alpha_i \ge 0, \ i = 1, \dots, m,$$
(A.4)

is convex (concave).

If f_i are convex (concave) and at least one of them is strictly convex (strictly concave) and corresponding α_i is positive, then the function $f(\mathbf{x})$ defined above is strictly convex (strictly concave).

Proposition A.4 (convexity of the supremum of a family of convex functions, [20, Theorem 4.1.13]). Let $f_i : X \to \mathbb{R}$, $i \in I$, be convex functions which are bounded from above on the convex set X in \mathbb{R}^n . Then the function

$$f(\mathbf{x}) \stackrel{\text{def}}{=} \sup_{i \in I} f_i(\mathbf{x})$$
(A.5)

is convex on X.

f is strictly convex if each f_i is strictly convex and I is finite.

Recall that a vector $\hat{\mathbf{f}}(\overline{\mathbf{x}})$ is said to be a *subgradient* or a *generalized gradient* of f at $\overline{\mathbf{x}}$ if

$$f(\mathbf{x}) - f(\overline{\mathbf{x}}) \ge \langle \widehat{\mathbf{f}}(\overline{\mathbf{x}}), \mathbf{x} - \overline{\mathbf{x}} \rangle,$$
 (A.6)

for any $\mathbf{x} \in \mathbb{R}^n$, where $\langle \mathbf{x}, \mathbf{y} \rangle$ denotes the inner (scalar) product of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

The set containing all subrgadients of $f(\mathbf{x})$ at $\overline{\mathbf{x}}$ is said to be a *subdifferential* $\partial f(\overline{\mathbf{x}})$ of f at $\overline{\mathbf{x}}$.

If $f(\mathbf{x})$ is differentiable at $\overline{\mathbf{x}}$, then $\widehat{\mathbf{f}}(\overline{\mathbf{x}}) = \mathbf{f}'(\overline{\mathbf{x}})$ and $\partial f(\overline{\mathbf{x}}) = {\mathbf{f}'(\overline{\mathbf{x}})}$ is a singleton, where $\mathbf{f}'(\overline{\mathbf{x}})$ is the gradient of $f(\mathbf{x})$ at $\overline{\mathbf{x}}$.

Proposition A.5 (subdifferential of a product of convex function with a positive real number). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then $\partial(\alpha f)(\mathbf{x}_0) = \alpha \partial f(\mathbf{x}_0)$ for each scalar $\alpha > 0$.

Proposition A.6 (subdifferential of a sum of convex functions, [24, Theorem 23.8]). Let $f(\mathbf{x}) = \sum_{i=1}^{m} f_i(\mathbf{x})$, where f_i are proper convex functions on \mathbb{R}^n , and the convex sets $ri(\text{dom } f_i)$, i = 1, ..., m, have a point in common, where "ri" stands for relative interior of a set and "dom" stands for effective domain of a function. Then $\partial f(\mathbf{x}) = \sum_{i=1}^{m} \partial f_i(\mathbf{x})$ for each $\mathbf{x} \in \mathbb{R}^n$.

Proposition A.7 (subdifferential of a maximum of convex functions, [14, Lemma 5.4]). Let $f_i(\mathbf{x})$, i = 1, ..., m, be convex functions on \mathbb{R}^n and

$$f(\mathbf{x}) = \max_{i=1,\dots,m} f_i(\mathbf{x}).$$
(A.7)

Then

$$\partial f(\mathbf{x}) = \operatorname{co}\left\{\partial f_i(\mathbf{x}) : i \in I(\mathbf{x})\right\},$$
 (A.8)

where

$$I(\mathbf{x}) = \{i \in \{1, \dots, m\} : f_i(\mathbf{x}) = f(\mathbf{x})\}$$
(A.9)

and co X denotes the convex hull of X.

Proposition A.8 (convexity, strict convexity, concavity, and strict concavity of differentiable multivariate functions, [20, Theorems 6.1.2 and 6.2.2]). Let f be a numerical differentiable function on an open convex set X in \mathbb{R}^n . f is convex on X if and only if

$$f(\mathbf{x}_{1}) - f(\mathbf{x}_{2}) \ge \left\langle \mathbf{f}'(\mathbf{x}_{2}), \mathbf{x}_{1} - \mathbf{x}_{2} \right\rangle, \qquad (A.10)$$

for each $\mathbf{x}_1, \mathbf{x}_2 \in X$. Similarly, f is concave on X if and only if, for each $\mathbf{x}_1, \mathbf{x}_2 \in X$,

$$f(\mathbf{x}_1) - f(\mathbf{x}_2) \le \langle \mathbf{f}'(\mathbf{x}_2), \mathbf{x}_1 - \mathbf{x}_2 \rangle.$$
 (A.11)

f is strictly convex (strictly concave) on *X* if and only if these inequalities are strict, respectively, for each $\mathbf{x}_1, \mathbf{x}_2 \in X$, $\mathbf{x}_1 \neq \mathbf{x}_2$.

Proposition A.9 (uniqueness of the optimal solution to a strictly convex program, [20, Theorem 5.2.2]). Let X be a convex set in \mathbb{R}^n , let f be a strictly convex numerical function on X, and let \mathbf{x}^* be a solution to the minimization problem $\min_{\mathbf{x} \in X} f(\mathbf{x})$. Then \mathbf{x}^* is the unique solution to this problem.

Proposition A.10 (Fermat's generalized rule). Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and let X be a nonempty convex set in \mathbb{R}^n . The point \mathbf{x}^* is an optimal solution to the minimization problem

$$\min_{\mathbf{x}\in X} f(\mathbf{x}) \tag{A.12}$$

if and only if there exists a subgradient $\hat{\mathbf{x}} \in \partial f(\mathbf{x}^*)$ *such that for each* $\mathbf{x} \in X$ *the following inequality holds true:*

$$\langle \hat{\mathbf{x}}, \mathbf{x} - \mathbf{x}^* \rangle \ge 0.$$
 (A.13)

In particular, if $X = \mathbb{R}^n$, \mathbf{x}^* is an optimal solution to the minimization problem $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ if and only if $\mathbf{0} \in \partial f(\mathbf{x}^*)$.

Proposition A.11 (existence of element of best approximation, e.g., [17, Propositions 1.3.1 and 1.3.2]). Let *L* be a linear subspace of the normed linear space *X* and let *L* be generated by the linearly independent elements $\{\varphi_j\}_{j=0}^n$ of *X*. Then for each $f \in X$ there exists an element of best approximation in *L*.

Proposition A.12 (uniqueness of the element of best approximation, e.g., [17, Proposition 1.3.3]). *If X is a strictly convex space, then the element of best approximation is unique.*

A.2. The Gradient Method for Differentiable Functions. In order to compare the results, obtained by the subgradient method for nonsmooth optimization for problems (5) [(6)] and (12) [(13)], with the corresponding results, obtained by methods for "differentiable" optimization for problems (25) and (26), respectively, consider the iterative gradient method

for solving the "differentiable" unconstrained minimization problem

$$\min f(\mathbf{x}), \qquad (A.14)$$

where $f: \mathbb{R}^n \to \mathbb{R}$.

The *gradient method* for solving problem (A.14) is defined through

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \rho_k \mathbf{f}'(\mathbf{x}_k), \quad k = 0, 1, \dots,$$
(A.15)

where $\mathbf{x}_0 \in \mathbb{R}^n$ is an arbitrary initial guess (initial approximation); $\rho_k \geq 0$) is a step size; $\mathbf{f}'(\mathbf{x}_k)$ is the unique gradient of the differentiable function $f(\mathbf{x})$ at \mathbf{x}_k .

We use, for example, a line search method for choosing the step size ρ_k . The gradient method with such a choice of step size is known as the *steepest descent method*. The value of ρ_k is an optimal solution to the following single-variable problem of ρ :

$$\min\left\{f\left(\mathbf{x}_{k}-\rho\mathbf{f}'\left(\mathbf{x}_{k}\right)\right)\right\},\qquad(A.16)$$

subject to $\rho \ge 0$; that is,

$$f\left(\mathbf{x}_{k}-\rho_{k}\mathbf{f}'\left(\mathbf{x}_{k}\right)\right)=\min_{\rho\geq0}f\left(\mathbf{x}_{k}-\rho\mathbf{f}'\left(\mathbf{x}_{k}\right)\right).$$
 (A.17)

An alternative way of choosing the step length ρ_k is the so-called *doubling method*. Set, for example, $\rho_0 = 1$. Choose $\rho_k = \rho_{k-1}$. If $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$, then $\rho_k := 2\rho_{k-1}$. If $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$ again, then this doubling continues until $f(\mathbf{x})$ stops to decrease. If $f(\mathbf{x}_{k+1}) \ge f(\mathbf{x}_k)$, then $\mathbf{x}_{k+1} := (1/2)\rho_{k-1}$. If $f(\mathbf{x}_k - (1/2)\rho_{k-1}\mathbf{f}'(\mathbf{x}_k)) < f(\mathbf{x}_k)$, then $\mathbf{x}_{k+1} := \mathbf{x}_k - (1/2)\rho_{k-1}\mathbf{f}'(\mathbf{x}_k)$; go to iteration k + 2. If $f(\mathbf{x}_k - (1/2)\rho_{k-1}\mathbf{f}'(\mathbf{x}_k)) \ge f(\mathbf{x}_k)$, then $\rho_k := (1/4)\rho_{k-1}$, and so on.

The gradient method (A.15) can be considered as a special case of subgradient method (32) (with $\gamma_k = 1$) when the function *f* to be minimized is differentiable.

Gradients of $\Phi_3^2(\mathbf{a})$ and $\Phi_6^2(\mathbf{x})$, respectively, at iteration *k* are

$$\left[\left(\Phi_{3}^{2}\right)'\right]^{(k)}\left(\mathbf{a}_{k}\right) = \left(f_{0}^{(k)}\left(\mathbf{a}_{k}\right), \dots, f_{n}^{(k)}\left(\mathbf{a}_{k}\right)\right), \qquad (A.18)$$

where

$$f_{l}^{(k)}(\mathbf{a}_{k}) = 2\sum_{i=1}^{m} w_{i} \left(\sum_{j=0}^{n} a_{j}^{(k)} \varphi_{j}(\mathbf{x}_{i}) - f(\mathbf{x}_{i})\right) \varphi_{l}(\mathbf{x}_{i}),$$
(A.19)
$$l = 0, 1, \dots, n,$$
$$\left[\left(\Phi_{6}^{2}\right)'\right]^{(k)}(\mathbf{x}_{k}) = \left(f_{1}^{(k)}(\mathbf{x}_{k}), \dots, f_{n}^{(k)}(\mathbf{x}_{k})\right),$$
(A.20)

where

$$f_{l}^{(k)}(\mathbf{x}_{k}) = 2\sum_{i=1}^{m} w_{i}a_{il}\left(\sum_{j=1}^{n} a_{ij}x_{j}^{(k)} - b_{i}\right), \quad l = 1, \dots, n.$$
(A.21)

$$m \left\| \mathbf{y} \right\|^{2} \le \left\langle f'' \left(\mathbf{x} \right) \mathbf{y}, \mathbf{y} \right\rangle \le M \left\| \mathbf{y} \right\|^{2}$$
(A.22)

for any $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$, and let sequence $\{\mathbf{x}_k\}$ be generated by the steepest descent method (method (A.15) with ρ_k determined by a line search method).

Then f has a unique minimum solution $\mathbf{x}^* \in \mathbb{R}^n$ and for each $\mathbf{x} \in \mathbb{R}^n$ the following inequality holds true:

$$\left\|\mathbf{f}'(\mathbf{x})\right\|^2 \ge m\left(1 + \frac{m}{M}\right)\left(f\left(\mathbf{x}\right) - f\left(\mathbf{x}^*\right)\right). \tag{A.23}$$

Further, there exist constants q and $C: 0 \le q \le 1$, C > 0 *such that*

$$f(\mathbf{x}_{k+1}) - f(\mathbf{x}^{*}) \le q^{k} (f(\mathbf{x}_{0}) - f(\mathbf{x}^{*})),$$

$$\|\mathbf{x}_{k+1} - \mathbf{x}^{*}\| \le Cq^{k/2}, \quad k = 0, 1, \dots.$$
(A.24)

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

References

- K. D. Andersen, "An efficient Newton barrier method for minimizing a sum of Euclidean norms," *SIAM Journal on Optimization*, vol. 6, no. 1, pp. 74–95, 1996.
- [2] I. Barrodale and F. Roberts, "An improved algorithm for discrete *ℓ*₁ linear approximation," *SIAM Journal on Numerical Analysis*, vol. 10, no. 5, pp. 839–848, 1973.
- [3] I. Barrodale and F. Roberts, "An efficient algorithm for discrete l₁ linear approximation with constraints," *SIAM Journal on Numerical Analysis*, vol. 15, no. 3, pp. 603–611, 1978.
- [4] R. H. Bartels, A. R. Conn, and J. W. Sinclair, "Minimization techniques for piecewise differentiable functions: the l₁ solution to an overdetermined linear system," *SIAM Journal on Numerical Analysis*, vol. 15, no. 2, pp. 224–241, 1978.
- [5] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming. Theory and Algorithms*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1993.
- [6] D. P. Bertsekas, "A new class of incremental gradient methods for least squares problems," *SIAM Journal on Optimization*, vol. 7, no. 4, pp. 913–926, 1997.
- [7] Å. Björck, Numerical Methods for Least Squares Problems, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, USA, 1996.
- [8] P. H. Calamai and A. R. Conn, "A stable algorithm for solving the multifacility location problem involving Euclidean distances," *Society for Industrial and Applied Mathematics*, vol. 1, no. 4, pp. 512–526, 1980.
- [9] P. H. Calamai and A. R. Conn, "A projected Newton method for *l_p* norm location problems," *Mathematical Programming*, vol. 38, no. 1, pp. 75–109, 1987.
- [10] J. A. Chatelon, D. W. Hearn, and T. J. Lowe, "A subgradient algorithm for certain minimax and minisum problems," *Mathematical Programming*, vol. 15, no. 2, pp. 130–145, 1978.

- [11] F. Clarke, Optimization and Nonsmooth Analysis, vol. 5 of Classics in Applied Mathematics, SIAM, Philadelphia, Pa, USA, 1990.
- [12] T. F. Coleman and Y. Li, "A global and quadratically convergent method for linear l_{∞} problems," *SIAM Journal on Numerical Analysis*, vol. 29, no. 4, pp. 1166–1186, 1992.
- [13] T. F. Coleman and Y. Li, "A globally and quadratically convergent affine scaling method for linear *ℓ*₁ problems," *Mathematical Programming*, vol. 56, no. 1–3, pp. 189–222, 1992.
- [14] V. F. Demyanov and L. V. Vasiliev, Nondifferentiable Optimization, Springer, Berlin, Germany, 1985.
- [15] J. Fischer, "An algorithm for discrete linear l_p approximation," *Numerische Mathematik*, vol. 38, no. 1, pp. 129–139, 1981.
- [16] L. Kantorovich and G. Akilov, *Functional Analysis*, Nauka, Moscow, Russia, 1983, (Russian).
- [17] N. P. Korneichuk, Extremum Problems of Approximation Theory, Nauka, Moscow, Russia, 1976 (Russian).
- [18] C. L. Lawson and R. J. Hanson, Solving Least Squares Problems, SIAM, Philadelphia, Pa, USA, 1995.
- [19] Y. Li, "A globally convergent method for *l_p* problems," *SIAM Journal on Optimization*, vol. 3, no. 3, pp. 609–629, 1993.
- [20] O. L. Mangasarian, Nonlinear Programming, vol. 10 of Classics in Applied Mathematics, SIAM, Philadelphia, PA, USA, 1994.
- [21] G. Merle and H. Späth, "Computational experiences with discrete l_p-approximation," *Computing*, vol. 12, no. 4, pp. 315– 321, 1974.
- [22] M. L. Overton, "A quadratically convergent method for minimizing a sum of Euclidean norms," *Mathematical Programming*, vol. 27, no. 1, pp. 34–63, 1983.
- [23] E. Remez, Fundamentals of Numerical Methods for Chebyshev Approximation, Naukova Dumka, Kiev, Ukraine, 1969 (Russian).
- [24] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, USA, 1997.
- [25] S. M. Stefanov, "Polynomial algorithms for projecting a point onto a region defined by a linear constraint and box constraints in Rⁿ," *Journal of Applied Mathematics*, vol. 2004, no. 5, pp. 409– 431, 2004.
- [26] S. M. Stefanov, "Well-posedness and primal-dual analysis of some convex separable optimization problems," *Advances in Operations Research*, vol. 2013, Article ID 279030, 10 pages, 2013.
- [27] G. A. Watson, "On two methods for discrete ℓ_p approximation," *Computing*, vol. 18, no. 3, pp. 263–266, 1977.
- [28] J. M. Wolfe, "On the convergence of an algorithm for discrete *l_p* approximation," *Numerische Mathematik*, vol. 32, no. 4, pp. 439–459, 1979.
- [29] G. Xue and Y. Ye, "An efficient algorithm for minimizing a sum of Euclidean norms with applications," *SIAM Journal on Optimization*, vol. 7, no. 4, pp. 1017–1036, 1997.
- [30] K. Yosida, Functional Analysis, Classics in Mathematics, Springer, Berlin, Germany, 1995.
- [31] J. Yu, S. V. N. Vishwanathan, S. Günter, and N. N. Schraudolph, "A quasi-Newton approach to nonsmooth convex optimization problems in machine learning," *The Journal of Machine Learning Research*, vol. 11, pp. 1145–1200, 2010.
- [32] Y. I. Zhuravlev, Y. Laptin, A. Vinogradov, N. Zhurbenko, and A. Likhovid, "Non smooth optimization methods in the problems of constructing a linear classifier," *International Journal Information Models and Analyses*, vol. 1, pp. 103–111, 2012.



The Scientific World Journal





Decision Sciences







Journal of Probability and Statistics



Hindawi Submit your manuscripts at





International Journal of Differential Equations





International Journal of Combinatorics





Mathematical Problems in Engineering



Abstract and Applied Analysis



Discrete Dynamics in Nature and Society







Journal of Function Spaces



International Journal of Stochastic Analysis



Journal of Optimization