Hindawi
Complexity
Volume 2021, Article ID 6640781, 8 pages
https://doi.org/10.1155/2021/6640781

Research Article

An Infeasible Incremental Bundle Method for Nonsmooth Optimization Problem Based on CVaR Portfolio

Jia-Tong Li,[1] Jie Shen ,[2] and Na Xu[2]

[1]*College of Science, Northeast Forestry University, Harbin 150040, China*
[2]*School of Mathematics, Liaoning Normal University, Dalian 116029, China*

Correspondence should be addressed to Jie Shen; tt010725@163.com

Received 12 November 2020; Revised 30 November 2020; Accepted 28 December 2020; Published 18 January 2021

Academic Editor: Abdelalim Elsadany

For CVaR (conditional value-at-risk) portfolio nonsmooth optimization problem, we propose an infeasible incremental bundle method on the basis of the improvement function and the main idea of incremental method for solving convex finite min-max problems. The presented algorithm only employs the information of the objective function and one component function of constraint functions to form the approximate model for improvement function. By introducing the aggregate technique, we keep the information of previous iterate points that may be deleted from bundle to overcome the difficulty of numerical computation and storage. Our algorithm does not enforce the feasibility of iterate points and the monotonicity of objective function, and the global convergence of the algorithm is established under mild conditions. Compared with the available results, our method loosens the requirements of computing the whole constraint function, which makes the algorithm easier to implement.

1. Introduction

Optimization problems arise in the wide range of practical applications, and they have been successfully solved by utilizing various methods, especially by state of the art approaches [1–3]. For an actual engineering optimization problem [2], a new optimal mutation strategy based on the complementary advantages of five mutation strategies is designed to develop a novel improved differential evolution algorithm with the wavelet basis function; the proposed method can improve the search quality, accelerate convergence and avoid fall into local optimum and stagnation. Parametric analysis and optimization are conducted for a novel geothermal combined cooling and power system [3], and not only the combined system performs better than the separate system but also the n-nonane brings the lowest total product unit cost to the proposed system. Nonsmooth optimization (NSO) problems are in general difficult to solve. Lots of approaches are proposed to solve these problems [4–8]. Among others, bundle methods are considered as one of the most efficient and promising methods. Infeasible bundle methods [9,10] can be viewed as the unconstrained proximal-like bundle methods applied to improvement functions, and the main advantage superior to other methods is that it does not require the feasibility of the iterate points and the monotonicity of the objective function. Conditional value-at-risk (CVaR) is currently the main tool to measure financial risk when we face portfolio for selected risky assets, and the study of CVaR model usually brings about the following nonsmooth optimization problem:

$$\begin{aligned}&\min_{(x,\alpha)\in R^n\times R} \quad F_\theta(x,\alpha)=\alpha+\frac{1}{(1-\theta)J}\sum_{j=1}^{J}\left(-x^T y_j-\alpha\right)^+,\\&\text{s.t,}\quad -x^T m\leq -M,\sum_{i=1}^{n}x_i=1, x\geq 0,\end{aligned}\tag{1}$$

where $F_\theta(x,\alpha)$ is the approximate performance function which is convex and $\theta\in(0,1)$ is the probability level, $Y_{J\times n}=(y_{ji})$ is the scenario matrix, J is the number of scenarios, and its element y_{ji} denotes the yield rate of risky asset under scenario j. The vector $y_j=(y_{j1},y_{j2},\ldots,y_{jn})^T$ is the vector of yield rate of scenario, j. $(\cdot)^+=\max\{0,\cdot\}$. The vector $y\in R^n$

denotes the yield rate at the end of investment which is uncertain, and $m = E(y)$, the constant M is the given rate of expected return. We rewrite problem (1) in more general form:

$$\begin{aligned} &\min_{(x,\alpha)\in R^n\times R} \quad F_\theta(x,\alpha),\\ &\text{s.t} \quad c(x,\alpha)\le 0, \end{aligned} \tag{2}$$

where $c(x,\alpha) = \max\{c_j(x,\alpha)|, t jn \in q\tilde{J}h =_{\{1,2,\ldots,n+3\}}\}$ is the pointwise maximum of finite many convex functions, $c_1(x,\alpha) = -x^T e_1$, $c_2(x,\alpha) = -x^T e_2, \ldots,$ $c_n(x,\alpha) = -x^T e_n$, $c_{n+1}(x,\alpha) = -x^T m + M$, and $c_{n+2}(x,\alpha) = x^T e - 1$, $c_{n+3}(x,\alpha) = -x^T e + 1$ are called the component functions of c, and $e_i = (0,\ldots,1^{\text{ith}},\ldots,0)^T \in R^n, e = (1,1,\ldots,1)^T \in R^n$. Obviously, function $c\colon R^{n+1} \longrightarrow R$ is convex and nondifferentiable.

Problem (2) involves a special constraint function, namely, the max-valued function. Qing-Ye Zhang and Yan Gao [11] once transform the objective function in the CVaR model into a piecewise smooth function and present an algorithm by employing common proximal bundle methods, where the values of every component function need to be evaluated at each iteration. An incremental method for solving convex finite min-max problem provides us a new approach to deal with max-valued function, which extends the philosophy of the incremental approaches [12], and it does not need to evaluate the actual value of max-valued function. The idea has already been applied to linearly constrained min-max problems [13] and to inequality constraint min-max problems [14]. The algorithms provided by the above two papers are feasible and descent methods. But in some cases, the feasibility of iterate points is difficult to realize since we have to make a tradeoff between the search for feasibility and for the reduction of the objective function.

Motivated by the work [14, 15], we provide an iterate method which is strongly connected with the incremental technique [12] and the infeasible idea [9]. The algorithm we design does not require the evaluation of each component function of the constraint function. In other words, we employ the incomplete knowledge of the constraint function to construct a lower approximate model for improvement function associated with the original problem, which reduces the number of elements in bundle and decreases the difficulty of implementation of the algorithm. Based on the approximate model, a descent test is presented. Even though the model is a more rough approximation to improvement function, the proposed algorithm still possesses good convergence properties under mild conditions.

This paper is organized as follows: In Section 2, we introduce basic notations, concepts, and existing results which are the basis for the construction of approximate model and the design of overall algorithm. A new cutting-planes framework and the way to update the elements in bundle are also given in this section. Section 3 presents an infeasible incremental bundle algorithm for solving problem (2). The convergence analysis of the presented algorithm is discussed in Section 4, and the optimal solution to problem (2) is obtained under mild conditions.

2. Preliminaries and the Construction of Subproblem

We denote by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ the usual inner product and norm in R^n, respectively. The subdifferential of a convex function $f\colon R^n \longrightarrow R$ at x is defined by $\partial f(x) = \{p \in R^n | f(y) \ge f(x) + \langle p, y - x\rangle, \forall y \in R^n\}$. Let

$$D = \left\{(x,\alpha) \in R^{n+1} | c(x,\alpha) \le 0\right\} \tag{3}$$

be the feasible set of problem (2), and for sake of the notation simplicity, we leave out the confidence parameter θ in (2) and indicate $F_\theta(x,\alpha)$ by $F(x,\alpha)$.

Lemma 1. *Suppose that $f_i(x)$ $(i \in I)$ is the convex differentiable function and the index set I is finite. Then, the subdifferential of function $f(x) = \max\{f_i(x) | it \in nI\}$ is $\partial f(x) = co\{\nabla f_i(x) | it \in nIq(x)\}$, where "co" denotes the convex hull, $I(x) = \{i \in I | f_i t(x) n = qfh(x)\}$.*

Definition 1. For given $(y,\beta) \in R^{n+1}$, the following function,

$$H_{(y,\beta)}(x,\alpha) = \max\{F(x,\alpha) - F(y,\beta), c(x,\alpha)\}, \quad (x,\alpha) \in R^{n+1}, \tag{4}$$

is called the improvement function associated with problem (2).

The following lemmas present some useful properties of improvement function which play an important role in designing our algorithm.

Lemma 2. *[9] The subdifferential of improvement function $H_{(y,\beta)}(\cdot,\cdot)$ at point (x,α) is given by*

$$\partial H_{(y,\beta)}(x,\alpha) = \begin{cases} \partial F(x,\alpha), & \text{if } F(x,\alpha) - F(y,\beta) > c(x,\alpha),\\ \text{conv}\{\partial F(x,\alpha) \cup \partial c(x,\alpha)\}, & \text{if } F(x,\alpha) - F(y,\beta) = c(x,\alpha),\\ \partial c(x,\alpha), & \text{if } F(x,\alpha) - F(y,\beta) < c(x,\alpha). \end{cases} \tag{5}$$

Lemma 3. *[16] Suppose that the Slater constraint qualification is satisfied for problem (2), the following statements are equivalent:*

(i) $(\overline{x}, t\overline{\alpha})$ *is a solution to problem (2)*

(ii) $\min\left\{H_{(\overline{x},t\overline{\alpha})}(y,\beta) | (y,\beta) \in R^{n+1}\right\} = H_{(\overline{x},t\overline{\alpha})}(\overline{x}, t\overline{\alpha})$

(iii) $0 \in \partial H_{(\overline{x},t\overline{\alpha})}(\overline{x},t\overline{\alpha})$, *i.e.*, $0 \in \partial \phi(\overline{x},t\overline{\alpha})$, *where* $\phi(\cdot,\cdot) = H_{(\overline{x},t\overline{\alpha})}(\cdot,\cdot)$

Suppose that (x^k,α^k) is the current stability center (the current last serious step) and (y^i,β^i) is the trial point generated from previous iterations. We indicate $H_{(x^k,\alpha^k)}(\cdot,\cdot)$ by $H_k(\cdot,\cdot)$. Choose arbitrarily one component function c_j of c and define the values of the linearization functions of F and c_j at point (x^k,α^k):

$$
\begin{aligned}
s^k_{F_i} &= F(y^i,\beta^i) + \langle g^i_F, (x^k,\alpha^k) - (y^i,\beta^i)\rangle,\\
t^k_{c_{ji}} &= c_j(y^i,\beta^i) + \langle g^i_{c_j}, (x^k,\alpha^k) - (y^i,\beta^i)\rangle,
\end{aligned}
\tag{6}
$$

where $g^i_F \in \partial F(y^i,\beta^i)$, $g^i_{c_j} \in \partial c_j(y^i,\beta^i)$. Bundle B_l keeps memory of the information of previous iterations:

$$
\begin{aligned}
&B_l = B_l^1 \cup B_l^2 \cup \left\{(x^k,\alpha^k), F(x^k,\alpha^k), c(x^k,\alpha^k)\right\}, k = k(l), \text{ with}\\
&B_l^1 \subseteq \cup_{i<l}\left\{F(y^i,\beta^i), c_j(y^i,\beta^i),\quad s^k_{F_i}, t^k_{c_{ji}}, g^i_F \in \partial F(y^i,\beta^i), g^i_{c_j} \in \partial c_j(y^i,\beta^i)\right\},\\
&B_l^2 \subseteq \cup_{i<l}\left\{\left(\widehat{\varepsilon}^k_i, \widehat{g}^i \in \partial_{\widehat{\varepsilon}^k_i} H_k(x^k,\alpha^k)\right)\right\}, \text{ (if there is compression)},
\end{aligned}
\tag{7}
$$

where $k(l)$ denotes the index of the last serious step preceding the iteration l, when it is clear from the context, we do not specify the dependence of k on the current iteration index l. The element $(\widehat{\varepsilon}^k_i, \widehat{g}^i)$ in B_l^2 is the aggregate couple that is introduced when compression of bundle is implemented for controlling the size of bundle, and we will discuss aggregate technique later in detail.

Lemma 4. *Define*

$$
\begin{aligned}
e^k_i &= \begin{cases} F(x^k,\alpha^k) - s^k_{F_i} + c^+(x^k,\alpha^k), & \text{if } F(y^i,\beta^i) - F(x^k,\alpha^k) \geq c(y^i,\beta^i),\\ -t^k_{c_{ji}} + c^+(x^k,\alpha^k), & \text{if } F(y^i,\beta^i) - F(x^k,\alpha^k) < c(y^i,\beta^i),\end{cases}\\
g^i_{H_k} &= \begin{cases} g^i_F, & \text{if } F(y^i,\beta^i) - F(x^k,\alpha^k) \geq c(y^i,\beta^i),\\ g^i_{c_j}, & \text{if } F(y^i,\beta^i) - F(x^k,\alpha^k) < c(y^i,\beta^i),\end{cases}
\end{aligned}
\tag{8}
$$

where notations $s^k_{F_i}, t^k_{c_{ji}}, g^i_F, g^i_{c_j}, (y^i,\beta^i)$, and (x^k,α^k) are from (6) and (7). Then, we have $e^k_i \geq 0$ and $g^i_{H_k} \in \partial_{e^k_i} H_k(x^k,\alpha^k)$.

Proof. By definitions of $s^k_{F_i}, t^k_{c_{ji}}$ and $g^i_F \in F(y^i,\beta^i)$, $g^i_{c_j} \in \partial c_j(y^i,\beta^i)$, we have $F(x^k,\alpha^k) - s^k_{F_i} \geq 0$ and $c_j(x^k,\alpha^k) - t^k_{c_{ji}} \geq 0$. Since $c^+(x^k,\alpha^k) \geq 0$, $c^+(x^k,\alpha^k) \geq c_j(x^k,\alpha^k)$, it follows from (8) that $e^k_i \geq 0$. According to the definition of improvement function $H_k(\cdot,\cdot)$, we have

$$
\begin{aligned}
H_k(y,\beta) &\geq \max\begin{cases} F(y,\beta) - F(x^k,\alpha^k)\\ c(y,\beta)\end{cases} \geq \max\begin{cases} F(y,\beta) - F(x^k,\alpha^k)\\ c_j(y,\beta)\end{cases}\\
&= \max\begin{cases} -F(x^k,\alpha^k) + s^k_{F_i} + \langle g^i_F, (y,\beta) - (x^k,\alpha^k)\rangle\\ t^k_{c_{ji}} + \langle g^i_{c_j}, (y,\beta) - (x^k,\alpha^k)\rangle\end{cases}\\
&= c^+(x^k,\alpha^k) + \langle g^i_{H_k}, (y,\beta) - (x^k,\alpha^k)\rangle -\\
&\quad\begin{cases} F(x^k,\alpha^k) - s^k_{F_i} + c^+(x^k,\alpha^k), & \text{if } F(y^i,\beta^i) - F(x^k,\alpha^k) \geq c(y^i,\beta^i),\\ -t^k_{c_{ji}} + c^+(x^k,\alpha^k), & \text{if } F(y^i,\beta^i) - F(x^k,\alpha^k) < c(y^i,\beta^i).\end{cases}
\end{aligned}
\tag{9}
$$

Notice that $H_k(x^k,\alpha^k) = c^+(x^k,\alpha^k)$ and the above inequality implies $g^i_{H_k} \in \partial_{e^k_i} H_k(x^k,\alpha^k)$.

Now, we are in position to define the cutting-planes model for $H_k(\cdot,\cdot)$:

$$\psi_l(y,\beta) = c^+\left(x^k,\alpha^k\right) + \max\Big\{\max_{i\in B_l^1}\left\{-e_i^k + \langle g_{H_k}^i, (y,\beta) - \left(x^k,\alpha^k\right)\rangle\right\},$$
$$\max_{i\in B_l^2}\left\{-\hat{\varepsilon}_i^k + \langle \hat{g}^i, (y,\beta) - \left(x^k,\alpha^k\right)\rangle\right\}\Big\}. \tag{10}$$

By Lemma 3 and $\hat{g}^i \in \partial_{\hat{\varepsilon}_i^k} H_k(x^k,\alpha^k)$, ψ_l is a lower approximate model for H_k, that is,

$$\psi_l(y,\beta) \le H_k(y,\beta), \quad \forall (y,\beta) \in R^{n+1}. \tag{11}$$

Given μ_l, a positive proximal parameter, the next trial point (y^l,β^l) is generated by solving the following quadratic programming:

$$\min_{(y,\beta)\in R^{n+1}} \psi_l(y,\beta) + \frac{\mu_l}{2}\left\|(y,\beta) - \left(x^k,\alpha^k\right)\right\|^2. \tag{12}$$

Obviously, (y^l,β^l) is unique and

$$\left(y^l,\beta^l\right) = \left(x^k,\alpha^k\right) - \frac{1}{\mu_l}\hat{g}^l, \quad \text{where } \hat{g}^l \in \partial\psi_l\left(y^l,\beta^l\right). \tag{13}$$

In order to measure whether the new trial point (y^l,β^l) provides sufficient decrease of $H_k(\cdot,\cdot)$, we define the nominal decrease δ_l by

$$\delta_l = H_k\left(x^k,\alpha^k\right) - \psi_l\left(y^l,\beta^l\right) - \frac{\mu_l}{2}\left\|\left(y^l,\beta^l\right) - \left(x^k,\alpha^k\right)\right\|^2$$
$$= \hat{\varepsilon}_l^k + \frac{1}{2\mu_l}\left\|\hat{g}^l\right\|^2, \tag{14}$$

where

$$\hat{\varepsilon}_l^k = H_k\left(x^k,\alpha^k\right) - \psi_l\left(y^l,\beta^l\right) - \frac{1}{\mu_l}\left\|\hat{g}^l\right\|^2 (\ge 0). \tag{15}$$

Since (y^l,β^l) is the solution to problem (12) and ψ_l is a lower approximation to H_k, we have

$$H_k\left(x^k,\alpha^k\right) \ge \psi_l\left(x^k,\alpha^k\right) \ge \psi_l\left(y^l,\beta^l\right) + \frac{\mu_l}{2}\left\|\left(y^l,\beta^l\right) - \left(x^k,\alpha^k\right)\right\|^2. \tag{16}$$

It follows that $\delta_l \ge 0$. Now, we present the descent test. Let $m \in (0,1)$ be a given parameter, when (y^l,β^l) satisfies:

$$H_k\left(y^l,\beta^l\right) \le c_j^+\left(x^k,\alpha^k\right) - m\delta_l, \tag{17}$$

then we declare a serious step: $(x^{k+1},\alpha^{k+1}) = (y^l,\beta^l)$; otherwise, a null step is declared: $(x^{k+1},\alpha^{k+1}) = (x^k,\alpha^k)$. Notice that here we only employ the information of one component function c_j of c to measure the "property" of the new generated trial point (y^l,β^l), unlike [9]; therein, they use the whole information of c.

Obviously, for any $(y,\beta) \in R^{n+1}$,

$$H_k(y,\beta) \ge \psi_l(y,\beta) \ge \psi_l\left(y^l,\beta^l\right) + \langle \hat{g}^l, (y,\beta) - \left(y^l,\beta^l\right)\rangle$$
$$= H_k\left(x^k,\alpha^k\right) - \langle \hat{g}^l, (y,\beta) - \left(x^k,\alpha^k\right)\rangle - \hat{\varepsilon}_l^k, \tag{18}$$

Hence, $\hat{g}^l \in \partial_{\hat{\varepsilon}_l^k} H(x^k,\alpha^k)$. If $\hat{\varepsilon}_l^k$ and $\|\hat{g}^l\|$ are very small, it follows that the approximate optimality condition is satisfied and the algorithm stops (see Step 3 in Algorithm 1).

As iterations go along, the number of elements in bundle B_l increases and the burden of computation and storage increases simultaneously. When the size of bundle becomes too big, it is necessary to compress and to clean the model. We have to discard some elements from bundle B_l^1 and append the aggregate couple $(\hat{\varepsilon}_l^k,\hat{g}^l)$ into bundle B_l^2 since the aggregate couple summarizes the information of previous iterate points [17]. For this purpose, we introduce the aggregate function:

$$l_{k,l}(y,\beta) = H_k\left(x^k,\alpha^k\right) + \langle \hat{g}^l, (y,\beta) - \left(x^k,\alpha^k\right)\rangle - \hat{\varepsilon}_l^k, \quad (y,\beta) \in R^{n+1}. \tag{19}$$

It has the following equivalent expression:

$$l_{k,l}(y,\beta) = \psi_l\left(y^l,\beta^l\right) + \langle \hat{g}^l, (y,\beta) - \left(y^l,\beta^l\right)\rangle, \quad (y,\beta) \in R^{n+1}. \tag{20}$$

For each k, the following conditions guarantee the bundle technique applied to function $H_k$either produces a serious step after a finite number of null steps, or the current serious step (x^k,α^k) is the minimum of H_k[18]:

$$\begin{cases} (a) \quad \psi_l(y,\beta) \le H_k(y,\beta), & \text{for all } l \ge 1; \\ (b) \quad l_{k(l),l}(y,\beta) \le \psi_{l+1}(y,\beta), & \left(y^l,\beta^l\right) \text{ is a null step;} \\ (c) \quad H_k\left(y^l,\beta^l\right) + \langle g_{H_k}^l, (y,\beta) - \left(y^l,\beta^l\right)\rangle, & \left(y^l,\beta^l\right) \text{ is a null step.} \end{cases} \tag{21}$$

Conditions (b) and (c) mean that the bundle B_l should contain both the aggregate information and the last generated information.

For model function ψ_l defined by (10) to satisfy conditions (a)-(c) when passing to the next iteration $l+1$, we should consider two cases of the lth iteration being a null step or a serious step. When there is a null step, the updates of the elements in bundle and the model do not present any problems since the improvement function is fixed between consecutive serious steps. Now suppose that the lth iteration

Input: $\varepsilon_1 > 0, \varepsilon_2 > 0, \gamma > 0, \mu_{\max} > \mu_{\min} > 0, \overline{m} \in (0,1)$, and $|B_{\max}| \geq 2$.
Step 1: given an initial proximal parameter $\mu_0 > 0$. Choose $(x^0,\alpha^0) \in R^{n+1}$ and one component function c_j of c, set $(y^0,\beta^0) = (x^0,\alpha^0)$, and compute $(F(y^0,\beta^0), c_j(y^0,\beta^0), g_F^0 \in \partial F(y^0,\beta^0), g_{c_j}^0 \in \partial c_j(y^0,\beta^0))$. Set $k = 0, l = 1$, $s_{F_0}^0 = F(y^0,\beta^0), t_{c_{j0}}^0 = c_j(y^0,\beta^0)$.
Define $B_1^1 = \{(s_{F_0}^0, t_{c_{j0}}^0, F(y^0,\beta^0), c_j(y^0,\beta^0), g_F^0, g_{c_j}^0)\}$, $B_1^2 = \varnothing$.
Step 2: find the solution (y^l,β^l) to problem (12). Compute $\widehat{g}^l = \mu_l((x^k,\alpha^k) - (y^l,\beta^l)), \widehat{\varepsilon}_l^k = H_k(x^k,\alpha^k) - \psi_l(y^l,\beta^l) - (1/\mu_l)\|\widehat{g}^l\|^2, \delta_l = \widehat{\varepsilon}_l^k + (1/2\mu_l)\|\widehat{g}^l\|^2$, $(F(y^l,\beta^l), c_j(y^l,\beta^l), g_F^l, g_{c_j}^l)$, and $(s_{F_l}^k, t_{c_{jl}}^k)$ using (6) written i with l.
Step 3: if $\widehat{\varepsilon}_l^k \leq \varepsilon_1$ and $\|\widehat{g}^l\|^2 \leq \varepsilon_2$, then stop, (x^k,α^k) is an approximate solution to problem (2).
Step 4: compute $H_k(y^l,\beta^l)$. If $H_k(y^l,\beta^l) \leq c_j^+(x^k,\alpha^k) - \overline{m}\delta_l$, then set $(x^{k+1},\alpha^{k+1}) = (y^l,\beta^l)$ (serious step) and $\mu_{l+1} = \max\{(\mu_l/\gamma), \mu_{\min}\}$. Otherwise, set $(x^{k+1},\alpha^{k+1}) = (x^k,\alpha^k)$ (null step) and $\mu_{l+1} = \min\{\gamma\mu_l, \mu_{\max}\}$.
Step 5: set $B_{l+1}^1 = B_l^1, B_{l+1}^2 = B_l^2$, if $|B_{l+1}| = |B_{\max}|$, then delete at least two elements from B_{l+1}, and append $(\widehat{\varepsilon}_l^k, \widehat{g}^l)$ to B_{l+1}^2. Append $(F(y^l,\beta^l), c_j(y^l,\beta^l), g_F^l, g_{c_j}^l, s_{F_l}^k, t_{c_{jl}}^k)$ to B_{l+1}^1.
Step 6: if a serious step is taken, choose $\overline{j} \in \widetilde{J}(x^{k+1},\alpha^{k+1}) = \{j \in \widetilde{J} | c_j t(x^{k+1},\alpha^{k+1}) n = qch(x^{k+1},\alpha^{k+1})\}$ such that the function $c_{\overline{j}}$ is the component function of c we choose for constructing the cutting-planes model for $H_{k+1}(\cdot,\cdot)$. Replace component function c_j by $c_{\overline{j}}$ and compute $t_{c_{ji}}^k$ again. Update $s_{F_i}^{k+1}, t_{c_{ji}}^{k+1}$ for $i \in B_{l+1}^1$ by (22) and update $\widehat{\varepsilon}_i^{k+1}$ for $i \in B_{l+1}^2$ by (23).
Set $k = k + 1$.
Step 7: set $l = l + 1$, go to Step 2.
End of Algorithm 1.

ALGORITHM 1: Infeasible incremental bundle method for CVaR portfolio.

produces a serious step, i.e., $(x^{k+1},\alpha^{k+1}) = (y^l,\beta^l)$, it means that for the next iteration we have to work with the new function $H_{k+1}(\cdot,\cdot) = H_{(x^{k+1},\alpha^{k+1})}(\cdot,\cdot)$. Since serious step can be infeasible and the monotonicity of the objective function is not enforced, we can only make sure $H_k(x^{k+1},\alpha^{k+1}) < H_k(x^k,\alpha^k)$, it is possible that $F(x^{k+1},\alpha^{k+1}) > F(x^k,\alpha^k)$, and in that case, we have $H_{k+1}(\cdot,\cdot) \leq H_k(\cdot,\cdot)$, and, as a consequence, the cutting-planes model for $H_k(\cdot,\cdot)$ may not be a lower approximation for $H_{k+1}(\cdot,\cdot)$. Thus, the old model must be revised and adjusted to ensure that the conditions $(a - c)$ are satisfied for the new function $H_{k+1}(\cdot,\cdot)$.

Next lemma provides one approach to make sure that all elements in bundle correspond to approximate subgradients of new improvement function H_{k+1} at (x^{k+1},α^{k+1}); hence, the model ψ_{l+1} is still a lower approximation function for H_{k+1} after adjusting the bundle. For elements in B_l^1, we only need to update the values of the linearization functions of F and c_j at (x^{k+1},α^{k+1}). For elements in B_l^2, we have to invoke the information of F and c in order to make sure the aggregate subgradient $\widehat{g}^i$ also corresponds to some approximate subgradient of H_{k+1} at (x^{k+1},α^{k+1}).

Lemma 5. *Suppose that the trial point (y^l,β^l) generated by (12) is a serious step, i.e., $(x^{k+1},\alpha^{k+1}) = (y^l,\beta^l)$. Then, the following statements hold:*

(i) For each $i \in B_l^1$,

$$\begin{aligned} s_{F_i}^{k+1} &= s_{F_i}^k + \langle g_F^i, (x^{k+1},\alpha^{k+1}) - (x^k,\alpha^k)\rangle, \\ t_{c_{ji}}^{k+1} &= t_{c_{ji}}^k + \langle g_{c_j}^i, (x^{k+1},\alpha^{k+1}) - (x^k,\alpha^k)\rangle. \end{aligned} \tag{22}$$

Furthermore, $g_{H_{k+1}}^i \in \partial_{e_i^{k+1}} H_{k+1}(x^{k+1},\alpha^{k+1})$, where $e_i^{k+1} \geq 0$ and $g_{H_{k+1}}^i$ are defined in (8) written with k replaced by $k+1$.

(ii) For each $i \in B_l^2$, define

$$\begin{aligned} \widehat{\varepsilon}_i^{k+1} &= \widehat{\varepsilon}_i^k + c^+(x^{k+1},\alpha^{k+1}) - c^+(x^k,\alpha^k) \\ &\quad + \left(F(x^{k+1},\alpha^{k+1}) - F(x^k,\alpha^k)\right)^+ \\ &\quad + \langle \widehat{g}^i, (x^k,\alpha^k) - (x^{k+1},\alpha^{k+1})\rangle. \end{aligned} \tag{23}$$

Then, $\widehat{\varepsilon}_i^{k+1} \geq 0$ and $\widehat{g}^i \in \partial_{\widehat{\varepsilon}_i^{k+1}} H_{k+1}(x^{k+1},\alpha^{k+1})$.

Proof. For each $i \in B_l^1$, we have, for all $(y,\beta) \in R^{n+1}$,

$$\begin{aligned} F(y,\beta) &\geq F(y^i,\beta^i) + \langle g_F^i, (y,\beta) - (y^i,\beta^i)\rangle \\ &= F(x^k,\alpha^k) + \langle g_F^i, (y,\beta) - (x^k,\alpha^k)\rangle - \left(F(x^k,\alpha^k) - s_{F_i}^k\right) \\ &\geq F(x^{k+1},\alpha^{k+1}) + \langle g_F^i, (y,\beta) - (x^{k+1},\alpha^{k+1})\rangle \\ &\quad - \left(F(x^{k+1},\alpha^{k+1}) - \left(s_{F_i}^k + \langle g_F^i, (x^{k+1},\alpha^{k+1}) - (x^k,\alpha^k)\rangle\right)\right). \end{aligned} \tag{24}$$

It follows from (22) and $g_F^i \in_{F(x^k,\alpha^k)-s_{F_i}^k} F(x^{k+1},\alpha^{k+1})$ that

$$\begin{aligned} F\left(x^{k+1},\alpha^{k+1}\right) &\ge F\left(x^k,\alpha^k\right) + \langle g_F^i, \left(x^{k+1},\alpha^{k+1}\right) - \left(x^k,\alpha^k\right)\rangle - \left(F\left(x^k,\alpha^k\right) - s_{F_i}^k\right) \\ &= s_{F_i}^k + \langle g_F^i, \left(x^{k+1},\alpha^{k+1}\right) - \left(x^k,\alpha^k\right)\rangle = s_{F_i}^{k+1}. \end{aligned} \tag{25}$$

Therefore, $F(x^{k+1},\alpha^{k+1}) - s_{F_i}^{k+1} \ge 0$ and $g_F^i \in_{F(x^{k+1},\alpha^{k+1})-s_{F_i}^{k+1}} F(x^{k+1},\alpha^{k+1})$. Similarly,

we have $g_{c_j}^i \in_{c_j(x^{k+1},\alpha^{k+1})-t_{c_{ji}}^{k+1}} c_j(x^{k+1},\alpha^{k+1})$. The remaining proof can be finished by imitating the proof of Lemma 2.3 in [9], where the quantities $(l,k,s_{F_i}^k,t_{c_{ji}}^k)$ are replaced by $(l+1,k+1,s_{F_i}^{k+1},t_{c_{ji}}^{k+1})$, respectively.

For each $i \in B_l^2$, the update pattern is just the one in [9], so we omit the proof.

To sum up, no matter whether the lth iteration produces a null step or a serious step, as long as,

$$\begin{aligned} &B_{l+1}^1 \subseteq \cup_{i<l+1}\left\{\left(F\left(y^i,\beta^i\right), c_j\left(y^i,\beta^i\right), s_{F_i}^{k+1}, t_{c_{ji}}^{k+1}, g_F^i, g_{c_j}^i\right)\right\}, \\ &B_{l+2}^2 \subseteq \cup_{i<l+1}\left\{\left(\widehat{\varepsilon}_i^{k+1}, \widehat{g}^i\right)\right\}, \end{aligned} \tag{26}$$

the model

$$\begin{aligned} \psi_{l+1}(y,\beta) = c^+\left(x^{k+1},\alpha^{k+1}\right) + \max\Big\{&\max_{i\in B_{l+1}^1}\big\{-e_i^{k+1} + \\ &\langle g_{H_{k+1}}^i, (y,\beta) - \left(x^{k+1},\alpha^{k+1}\right)\rangle\big\}, \\ &\max_{i\in B_{l+1}^2}\big\{-\widehat{\varepsilon}_i^{k+1} + \langle \widehat{g}^i, (y,\beta) - \left(x^{k+1},\alpha^{k+1}\right)\rangle\big\}\Big\} \end{aligned} \tag{27}$$

satisfies condition (a) in (21) written with l replaced by $l+1$ and k replaced by $k+1$, and the point (x^{k+1},α^{k+1}) indicates the $(k+1)$th stability center which may be (x^k,α^k) if a null step is executed. Furthermore, if $(\widehat{\varepsilon}_l^k,\widehat{g}^l) \subseteq B_{l+1}^2$, then ψ_{l+1} satisfies condition (b) in (21), and if $(F(y^l,\beta^l), c_j(y^l,\beta^l), s_{F_l}^{k+1}, t_{c_{jl}}^{k+1}, g_F^l, g_{c_j}^l) \subseteq B_{l+1}^1$, then ψ_{l+1} satisfies condition (c) in (21).

3. Infeasible Incremental Bundle Algorithm for Problem (2)

Remark 1. If a serious step is declared, we have

$$F\left(x^{k+1},\alpha^{k+1}\right) - F\left(x^k,\alpha^k\right) \le c_j^+\left(x^k,\alpha^k\right) - \overline{m}\delta_l, \tag{28}$$

$$c\left(x^{k+1},\alpha^{k+1}\right) \le c_j^+\left(x^k,\alpha^k\right) - \overline{m}\delta_l. \tag{29}$$

4. Convergence Analysis

In this part, we discuss the convergence of Algorithm 1 by borrowing the main idea from [9]. From now on we assume that Algorithm 1 produces an infinite sequence of iterate points. As usual in the convergence analysis of bundle methods, we consider two cases: the number of serious steps is infinite and the number of serious steps is finite, and the last serious step is followed by infinitely many null steps. Let $L_s = \left\{l \mid (y^l,\beta^l)t \text{ is a serious step}\right\}$ be the set which collects the indices of serious steps in the sequence $\left\{(y^l,\beta^l)\right\}$.

Proposition 1. *For any iteration index $k_0 \ge 0$ of serious step, it holds that*

$$\left(x^k,\alpha^k\right) \in \left\{(x,\alpha) \in R^{n+1} \mid c(x,\alpha) \le c^+\left(x^{k_0},\alpha^{k_0}\right)\right\}, \quad \forall k \ge k_0. \tag{30}$$

In particular, if $(x^{k_1},\alpha^{k_1}) \in D$ for some $k_1 \ge 0$, then $(x^k,\alpha^k) \in D$, for all $k \ge k_1$.

Proof. In Algorithm 1, the descent test is designed as follows: $H_k(y^l,\beta^l) \le c_j^+(x^k,\alpha^k) - \overline{m}\delta_l$. Since function c is the pointwise maximum of finite convex functions c_j $(j \in \widetilde{J})$, it follows that $c^+(x^k,\alpha^k) \ge c_j^+(x^k,\alpha^k)$. If a serious step is declared, we have $H_k(y^l,\beta^l) \le c^+(x^k,\alpha^k) - \overline{m}\delta_l$. Next, we adopt the techniques similar to [9], and the desired conclusions can be obtained. □

Proposition 2. *Suppose F is bounded below on D and Algorithm 1 generates an infinite number of serious steps. Then $\left\{\widehat{\varepsilon}_l^k\right\}_{l\in L_s} \longrightarrow 0$ and $\left\{\widehat{g}^l\right\}_{l\in L_s} \longrightarrow 0$.*

Proof. It follows from (29) and $c(x^k,\alpha^k) = \max\left\{c_j(x^k,\alpha^k) \mid j \in \widetilde{J}\right\}$ that if a serious step is declared,

$$c\left(x^{k+1},\alpha^{k+1}\right) \le c_j^+\left(x^k,\alpha^k\right) - \overline{m}\delta_{l(k+1)} \le c\left(x^k,\alpha^k\right) - \overline{m}\delta_{l(k+1)}. \tag{31}$$

Hence, the sequence $\{c(x^k,\alpha^k)\}$ is decreasing. Notice the update pattern of proximal parameter sequence $\{\mu_l\}$ in Algorithm 1, we have $\mu_l \le \mu_{\max}$ for all l, i.e., the condition $\mu_l \le \overline{\mu}$ for some $\overline{\mu} > 0$ holds. Following the thinking of [6], we can prove $\left\{\widehat{\varepsilon}_l^k\right\}_{l\in L_s} \longrightarrow 0$ and $\left\{\widehat{g}^l\right\}_{l\in L_s} \longrightarrow 0$.

Next proposition discusses the conditions which ensure the boundedness of sequence of serious steps $\{(x^k,\alpha^k)\}$.

Proposition 3. *Suppose that problem (2) has a solution $(\overline{x},t\overline{\alpha})$ and Algorithm 1 produces an infinite number of serious steps. If the feasible set D is bounded or there exists some iteration index k_1 such that $F(\overline{x},t\overline{\alpha}) \le F(x^k,\alpha^k) + c_j^+(x^k,\alpha^k)$ for all $k \ge k_1$ and for some $j \in \widetilde{J}$, then the sequence $\{(x^k,\alpha^k)\}$ is bounded.*

Proof. For the first case that the feasible set D is bounded, since D is the level set of function c, the convexity of function c implies that all the level sets of function c are bounded. Then, the boundedness of $\{(x^k, \alpha^k)\}$ follows from the conclusion of Proposition 1. For the second case, since $\widehat{g}^l \in \partial_{\widehat{\varepsilon}_l^k} H_k(x^k, \alpha^k)$ and $\delta_l = \widehat{\varepsilon}_l^k + (1/2\mu_l)\|\widehat{g}^l\|^2$, we have

$$\begin{aligned}\left\|\left(x^{k+1}, \alpha^{k+1}\right) - (\overline{x}, \overline{\alpha})\right\|^2 &= \left\|\left(x^k, \alpha^k\right) - (\overline{x}, \overline{\alpha})\right\|^2 - \frac{2}{\mu_l}\langle \widehat{g}^l, \left(x^k, \alpha^k\right) - (\overline{x}, \overline{\alpha})\rangle + \frac{1}{\mu_l^2}\left\|\widehat{g}^l\right\|^2 \\ &\le \left\|\left(x^k, \alpha^k\right) - (\overline{x}, \overline{\alpha})\right\|^2 + \frac{2}{\mu_l}\left(H_k(\overline{x}, \overline{\alpha}) - H_k\left(x^k, \alpha^k\right) + \delta_l\right).\end{aligned} \tag{32}$$

Observe that $F(\overline{x}, \overline{\alpha}) \le F(x^k, \alpha^k) + c_j^+(x^k, \alpha^k) \le F(x^k, \alpha^k) + c^+(x^k, \alpha^k)$ for all $k \ge k_1$ and $c_j(\overline{x}, \overline{\alpha}) \le 0$, $c^+(x^k, \alpha^k) \ge 0$, we have

$$H_k(\overline{x}, \overline{\alpha}) - H_k\left(x^k, \alpha^k\right) = \max\left\{F(\overline{x}, \overline{\alpha}) - F\left(x^k, \alpha^k\right), c(\overline{x}, \overline{\alpha})\right\} - c^+\left(x^k, \alpha^k\right) \le 0. \tag{33}$$

By combining (32) and (33), it follows that

$$\left\|\left(x^{k+1}, \alpha^{k+1}\right) - (\overline{x}, \overline{\alpha})\right\|^2 \le \left\|\left(x^k, \alpha^k\right) - (\overline{x}, \overline{\alpha})\right\|^2 + \frac{2}{\mu_l}\delta_l, \quad \forall k \ge k_l. \tag{34}$$

By [19], $\sum_{l \in L_s} \delta_l < +\infty$, and the conclusion we have just proved in Proposition 2, the sequence $\{\|(x^{k+1}, \alpha^{k+1}) - (\overline{x}, \overline{\alpha})\|\}$ is convergent, and hence the sequence $\{(x^k, \alpha^k)\}$ is bounded. □

Theorem 1. *Suppose that problem (2) satisfies the Slater constraint qualification and its solution set is nonempty. Assume that Algorithm 1 generates an infinite number of serious steps, which is bounded. Then, all the accumulation points of the sequence $\{(x^k, \alpha^k)\}$are solutions to problem (2). And, if there exists some iteration index k_1 such that $F(\overline{x}, \overline{\alpha}) \le F(x^k, \alpha^k) + c_j^+(x^k, \alpha^k)$for all $k \ge k_1$ and for some $j \in \widetilde{J}$, the whole sequence $\{(x^k, \alpha^k)\}$ converges to a solution to problem (2).*

Proof. Since the proof is very similar to the one in [9], we omit it.

Now we consider the second case that the Algorithm 1 generates finite serious steps followed by infinitely many null steps, i.e., there exists an index last $= \max\{l | lt \in nL_s\}$, and the corresponding last serious step index is denoted by k_{last}. Notice that the improvement function is fixed for $k \ge k_{\text{last}}$, that is, $H_k(\cdot, \cdot) = H_{k_{\text{last}}}(\cdot, \cdot)$, $\forall k \ge k_{\text{last}}$.

Theorem 2. *Suppose that problem (2) satisfies the Slater constraint qualification and Algorithm 1 produces a finite number of serious steps, then $(x^{k_{last}}, \alpha^{k_{last}})$ is a solution to problem (2).*

Proof. We consider the case $l \ge l_{\text{last}}$ and denote $H(\cdot, \cdot) = H_{k_{\text{last}}}(\cdot, \cdot)$. By imitating the proof of Theorem 4.5 in [9], we obtain

$$H(\overline{y}, \overline{\beta}) + \frac{\mu_{\max}}{2}\left\|(\overline{y}, \overline{\beta}) - \left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right)\right\|^2 \le H\left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right), \tag{35}$$

$$\psi_{l_i}\left(y^{l_i}, \beta^{l_i}\right) \longrightarrow H(\overline{y}, \overline{\beta}), \quad i \longrightarrow +\infty, \tag{36}$$

where $(\overline{y}, \overline{\beta})$ is an accumulation point of (y^l, β^l), i.e., $(y^{l_i}, \beta^{l_i}) \longrightarrow (\overline{y}, \overline{\beta})$ as $i \longrightarrow +\infty$. Since we assume that k_{last} is the last index for serious step, the descent test is not satisfied for $l \ge$ last:

$$\begin{aligned}H\left(y^l, \beta^l\right) - c_j^+\left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right) &> -\overline{m}\delta_l \\ &\ge -\overline{m}\left(H\left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right) - \psi_l\left(y^l, \beta^l\right)\right).\end{aligned} \tag{37}$$

From Step 6 in Algorithm 1, we know if a serious step is declared, we choose $\overline{j}$ from $\widetilde{J}(x^{k+1}, \alpha^{k+1}) = \left\{j \in \overline{J} | c_j(x^{k+1}, \alpha^{k+1}) = c(x^{k+1}, \alpha^{k+1})\right\}$ such that the function $c_{\overline{j}}$ is the component function of c we chose for constructing the cutting-planes model for $H_{k+1}(\cdot, \cdot)$. Therefore, according to (37),

$$H\left(y^l, \beta^l\right) - H\left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right) \ge -\overline{m}\left(H\left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right) - \psi_l\left(y^l, \beta^l\right)\right). \tag{38}$$

Taking the limits along the specified subsequence as $i \longrightarrow +\infty$, by using (36), we obtain $0 \le (1 - \overline{m})(H(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}) - H(y^l, \beta^l))$. Since $\overline{m} \in (0, 1)$, we have $H(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}) \le H(y^l, \beta^l)$. Taking into account (36), we have $(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}) = (\overline{y}, \overline{\beta})$. It follows from (35) that $(\overline{y}, \overline{\beta})$ is the optimal solution to the following problem:

$$\min_{(y,\beta)\in R^{n+1}} H(y,\beta) + \frac{\mu_{\max}}{2}\left\|(y,\beta) - \left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right)\right\|^2. \quad (39)$$

According to the optimality condition, $0 \in \partial H(y^l, \beta^l) = \partial H(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}})$, where $H(\cdot,\cdot) = H_{k_{\text{last}}}(\cdot,\cdot)$. By Lemma 3, $(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}})$ is a solution to problem (2). □

5. Conclusions

We present an infeasible incremental bundle method for the CVaR portfolio nonsmooth optimization problem; the algorithm is easier to implement since it only employs the information of the objective function and one component function of constraint functions. The algorithm does not enforce the feasibility of iterate points and the monotonicity of objective function, but the global convergence is established under mild conditions.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported by the National Nature Science Foundation of China (grant nos. 61877032 and 11601061), the Foundation of Educational Committee of Liaoning Province (LQ2019019), and the PhD Research Startup Foundation of Liaoning Normal University (203070091909).

References

[1] Y. Song, D. Wu, W. Deng et al., "Multi-population parallel co-evolutionary differential evolution for parameter optimization," in *Energy Conversion and Management*, Elsevier, Amsterdam, Netherlands, 2020.

[2] W. Deng, J. Xu, Y. Song, and H. Zhao, "Differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem," *Applied Soft Computing*, vol. 23, Article ID 106724, 2020.

[3] C. Wu, X. Xu, Q. Li et al., "Performance assessment and optimization of a novel geothermal combined cooling and power system integrating an organic flash cycle with an ammonia-water absorption refrigeration cycle," *Energy Conversion and Management*, vol. 227, no. 2021, Article ID 113562, 2020.

[4] W. Hare, C. Sagastizábal, and M. Solodov, "A proximal bundle method for nonsmooth nonconvex functions with inexact information," *Computational Optimization and Applications*, vol. 63, no. 1, pp. 1–28, 2016.

[5] J. Shen, Z.-Q. Xia, and L.-P. Pang, "A proximal bundle method with inexact data for convex nondifferentiable minimization," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 66, no. 9, pp. 2016–2027, 2007.

[6] J. Shen, X.-Q. Liu, F.-F. Guo, and S.-X. Wang, "An approximate redistributed proximal bundle method with inexact data for minimizing nonsmooth nonconvex functions," *Mathematical Problems in Engineering*, vol. 2015, Article ID 215310, 9 pages, 2015.

[7] K. C. Kiwiel, "Proximity control in bundle methods for convex nondifferentiable minimization," *Mathematical Programming*, vol. 46, no. 1–3, pp. 105–122, 1990.

[8] J. Shen, J. Lv, J. Lv, F.-F. Gao, and R. Zhao, "A new proximal Chebychev center cutting plane algorithm for nonsmooth optimization and its convergence," *Journal of Industrial & Management Optimization*, vol. 14, no. 3, pp. 1143–1155, 2018.

[9] S. Claudia and S. Mikhail, "An infeasible bundle method for nonsmooth convex constrained optimization without a penalty function or a filter," *SIAM. Journal on Optimization*, vol. 16, pp. 146–169, 2005.

[10] C. Lemaréchal, A. Nemirovskii, and Y. Nesterov, "New variants of bundle methods," *Mathematical Programming*, vol. 69, no. 1–3, pp. 111–147, 1995.

[11] Q.-Y. Zhang and Y. Gao, "A nonsmooth optimization method based on CVaR portfolio optimization problem," *Chinese Journal of Management Science*, vol. 25, no. 10, pp. 11–19, 2017.

[12] A. Nedic and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM Journal on Optimization*, vol. 12, no. 1, pp. 109–138, 2001.

[13] C. Tang, H. Chen, and J. Jian, "An improved partial bundle method for linearly constrained minimax problems," *Statistics, Optimization & Information Computing*, vol. 4, pp. 84–98, 2016.

[14] J. B. Jian, C. M. Tang, and F. Tang, "A feasible descent bundle method for inequality constrained Minimax problems (in Chinese)," *Science China Mathematics*, vol. 45, pp. 2001–2024, 2015.

[15] M. Gaudioso, G. Giallombardo, and G. Miglionico, "An incremental method for solving convex finite Min-Max problems," *Mathematics of Operations Research*, vol. 31, no. 1, pp. 173–187, 2006.

[16] K. C. Kiwiel, "Methods of descent for nondifferentiable optimization," in *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, Germany, 1985.

[17] J. Frédéric Bonnans, J. C. Gilbert, C. Lemaréchal et al., *Bundle Methods, the Quest for Descent, Numerical Optimization*, Springer, Berlin Germany, 2006.

[18] R. Correa and C. Lemaréchal, "Convergence of some algorithms for convex minimization," *Mathematical Programming*, vol. 62, no. 1–3, pp. 261–275, 1993.

[19] B. T. Polyak, *Introduction to Optimization*, Optimization Software, Inc., Publications Division, New York, NY, USA, 1987.