

Research Article

An Approximate Proximal Bundle Method to Minimize a Class of Maximum Eigenvalue Functions

Wei Wang,¹ Lingling Zhang,¹ Miao Chen,¹ and Sida Lin²

¹ School of Mathematics, Liaoning Normal University, Liaoning, Dalian 116029, China

² School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China

Correspondence should be addressed to Lingling Zhang; zllz00@sina.com

Received 26 February 2014; Revised 24 June 2014; Accepted 25 June 2014; Published 17 July 2014

Academic Editor: Zhihua Zhang

Copyright © 2014 Wei Wang et al. 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 present an approximate nonsmooth algorithm to solve a minimization problem, in which the objective function is the sum of a maximum eigenvalue function of matrices and a convex function. The essential idea to solve the optimization problem in this paper is similar to the thought of proximal bundle method, but the difference is that we choose approximate subgradient and function value to construct approximate cutting-plane model to solve the above mentioned problem. An important advantage of the approximate cutting-plane model for objective function is that it is more stable than cutting-plane model. In addition, the approximate proximal bundle method algorithm can be given. Furthermore, the sequences generated by the algorithm converge to the optimal solution of the original problem.

1. Introduction

Bundle method is one of the efficient and promising methods for solving nonsmooth optimization problems. For example, a class of maximum eigenvalue function can be minimized by bundle method [1], bundle-filter method can be used to deal with nonsmooth convex constrained optimization problem [2], and penalized bundle method was proposed to solve nonsmooth optimization problem by Bonnans et al. [3]. Recently, a minimization problem for a class of constrained maximum eigenvalue function has been solved in [4] with the help of penalized bundle method. However, when constructing the cutting-plane model for the objective function, the formula $\partial F(y) = \beta^* \partial \lambda_{\max}(A(y)) + \partial g(y)$ shows that the subdifferential of $\lambda_{\max}(A(y))$ is involved in [4]. Note that $\partial \lambda_{\max}(A(y))$ is the face of C_n exposed by $A(y)$, where $C_n := \{V \in S_n; V \succeq 0, \text{tr } V = 1\}$. So $\partial \lambda_{\max}(A(y))$ changes drastically when the multiplicity of the $\lambda_{\max}(A(y))$ is changed [5]. Thus $\partial F(y)$ is unstable and it leads to the instability of cutting-plane model in [4]. In this paper, to avoid this drawback, we try to give a more stable approximate cutting-plane model for the objective function.

In view of a class of nonsmooth optimization problem of the form

$$\min_{y \in R^n} f(y) + g(y), \quad (\text{P})$$

where $f(y) := (\lambda_{\max} \circ A)(y)$ is the composite function of λ_{\max} and an affine mapping A , specifically

$$R^n \ni y \mapsto A(y) := A_0 + \beta y \quad (1)$$

is affine, $A_0 \in S^n$ and β is a linear operator from R^n to S^n ; $g(y)$ is a nonsmooth convex function. We modify the elements in bundle and give an approximate proximal bundle method algorithm for (P). Our algorithm is established on the basic assumption that at least one approximate function value and one approximate subgradient at each point are available. And suppose $\text{ri dom } f(y) \cap \text{ri dom } g(y) \neq \emptyset$.

The initial motivation for our present work lies in the following facts. The whole idea of so-called bundle methods can be concentrated on constructing a good approximation for the objective function. Reference [4] has solved this class of optimization problem with the help of cutting-plane model and penalized bundle method. However, the cutting-plane

model in [4] is unstable. Our motivation is to construct a more stable approximate model for the objective function. We try to use the approximate subdifferential of the objective function $F(y) := f(y) + g(y) = (\lambda_{\max} \circ A)(y) + g(y)$. Note that in the equation $\partial_\varepsilon F(y) = \beta^* \partial_\varepsilon \lambda_{\max}(A(y)) + \partial_\varepsilon g(y)$, obviously, $\partial_\varepsilon \lambda_{\max}(A(y))$ is involved. And $\partial_\varepsilon \lambda_{\max}(A(y))$ is no longer a face of C_n ; it is the intersection of C_n with the half-space $\{V \in S_n : V \cdot A(y) \geq \lambda_{\max}(A(y)) - \varepsilon\}$. In particular, almost all matrices in $\partial_\varepsilon \lambda_{\max}(A(y))$ have rank n (for $\varepsilon > 0$), whereas the rank of matrices in $\partial \lambda_{\max}(A(y))$ is at most the multiplicity r , which is 1 for almost all $A(y)$ (so $\partial_\varepsilon \lambda_{\max}(A(y))$ is relative stable); this also gives an idea of the big gap existing between these two convex sets. Therefore we decide to introduce an enlarged subdifferential [6] of $\lambda_{\max}(A(y))$ which can be regarded as an outer approximation of $\partial \lambda_{\max}(A(y))$ and an inner approximation of $\partial_\varepsilon \lambda_{\max}(A(y))$. Through the enlarged subdifferential of $\lambda_{\max}(A(y))$, we can easily get an enlarged subdifferential of the objective function $F(y)$. Using the approximate subgradient in the enlarged subdifferential to construct the approximate cutting-plane model for the objective function, there are two advantages: on one hand it can avoid the instability of the cutting-plane model which is constructed in [4], on the other hand it can avoid too much elements in $\partial_\varepsilon \lambda_{\max}(A(y))$. After constructing a more stable approximate cutting-plane model, we propose our algorithm and prove the convergence of the algorithm.

The rest of this paper is organized as follows. Section 2 mainly contains the approximate cutting-plane model of the objective function. Here we firstly introduce an enlarged subdifferential which is an outer approximation of $\partial F(y)$, simultaneously, it is an inner approximation of $\partial_\varepsilon F(y)$. And then, using the approximate subgradient in the enlarged subdifferential to complete the constructing of an approximate cutting-plane model for the objective function. Section 3 gives the algorithm with respect to the approximate proximal bundle method and also provides a corresponding compression mechanism. Section 4 is devoted to convergence analysis of the algorithm mentioned in Section 3. Section 5 gives the conclusions.

In the paper, the standard norm and inner product are all in Hilbert space and are denoted by $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$.

2. The Approximate Model of the Objective Function

In this section, we will mainly give the approximate model for the objective function. It is known that [4] has given a kind of approximate model for the objective function $F(y)$:

$$\begin{aligned} \tilde{F}_k(y) &= \max_{i=1,2,\dots,k} \left\{ F(x^i) + \left\langle \beta^* Q(A(x^i)) ZQ(A(x^i))^T + s^i, y - x^i \right\rangle \right\}, \end{aligned} \quad (2)$$

where $Q(A(x^i))ZQ(A(x^i))^T \in \partial \lambda_{\max}(A(x^i))$ and $s^i \in \partial g(x^i)$.

It is known that $\partial \lambda_{\max}(A(y))$ is the face of C_n exposed by $A(y)$, where $C_n := \{V \in S_n : V \geq 0, \text{tr } V = 1\}$. If the multiplicity of $\lambda_{\max}(A(y))$ is changed, the subdifferential of $\lambda_{\max}(A(y))$ changes drastically. So it leads that the subdifferential of function $F(y)$ is unstable. Thus this causes the instability of $\tilde{F}_k(y)$. In order to construct a stable approximate model, we take the approximate subdifferential of function $F(y)$ into account. Firstly, we consider the approximate subdifferential of $\lambda_{\max}(A(y))$.

Definition 1. For all $y \in R^n$, $\varepsilon \geq 0$ and $A: R^n \rightarrow S^n$ is an affine mapping. One defines the following:

(1) the set of indices of ε -large eigenvalues:

$$\begin{aligned} I_\varepsilon(A(y)) &:= \{i \in 1, 2, \dots, n : \lambda_i(A(y)) > \lambda_{\max}(A(y)) - \varepsilon\}; \end{aligned} \quad (3)$$

(2) the ε -multiplicity of $\lambda_{\max}(A(y))$:

$$r_\varepsilon := \max \{i : i \in I_\varepsilon(A(y))\}; \quad (4)$$

(3) the ε -first eigenspace:

$$E_\varepsilon(A(y)) := \bigoplus_{i \in I_\varepsilon(A(y))} E_i(A(y)), \quad (5)$$

where $E_i(A(y))$ is the eigenspace of $A(y)$ associated with the its eigenspace $\lambda_i(A(y))$.

For the composite maximum eigenvalue function, its approximate subdifferential is

$$\partial_\varepsilon \lambda_{\max}(A(y)) = \{V \in C_n : V \cdot A(y) \geq \lambda_{\max}(A(y)) - \varepsilon\}, \quad (6)$$

where $C_n = \{V \in S^n : V \geq 0, \text{tr } V = 1\}$. From this one sees that $\partial_\varepsilon \lambda_{\max}(A(y))$ is the intersection of C_n with the half-space $\{V \in S^n : V \cdot A(y) \geq \lambda_{\max}(A(y)) - \varepsilon\}$, instead of being a face of C_n . In particular, almost all matrices in $\partial_\varepsilon \lambda_{\max}(A(y))$ have rank n (for $\varepsilon > 0$), whereas the rank of matrices in $\partial \lambda_{\max}(A(y))$ are at most the multiplicity r , which is 1 for almost all $A(y)$; this gives ideas that $\partial_\varepsilon \lambda_{\max}(A(y))$ is more stable than $\partial \lambda_{\max}(A(y))$ and there is a big gap existing between these two convex sets.

Introduce a compact convex set

$$\sigma_\varepsilon \lambda_{\max}(A(y)) = \{Q_\varepsilon(A(y)) ZQ_\varepsilon(A(y))^T : Z \in C_{r_\varepsilon}\}, \quad (7)$$

where $Q_\varepsilon(A(y))$ is a $n \times r_\varepsilon$ matrix whose columns form orthonormal basis of $E_\varepsilon(A(y))$.

Proposition 2. Let $A: R^n \rightarrow S^n$ be an affine mapping, then for all $y \in R^n$ and $\varepsilon \geq 0$ one has

$$\partial \lambda_{\max}(A(y)) \subset \sigma_\varepsilon \lambda_{\max}(A(y)) \subset \partial_\varepsilon \lambda_{\max}(A(y)). \quad (8)$$

Proof. The left inclusion can be derived from simple chain rules and (7). To see another inclusion, take $P \in \sigma_\varepsilon \lambda_{\max}(A(y))$. Then

$$P = Q_\varepsilon(A(y)) ZQ_\varepsilon(A(y))^T \quad \text{with } Z \in C_{r_\varepsilon}. \quad (9)$$

Since

$$\begin{aligned} Q_\varepsilon(A(y))^T T Q_\varepsilon(A(y)) \\ = \text{diag}(\lambda_{\max}(A(y)), \dots, \lambda_{r_\varepsilon}(A(y))) \geq \lambda_{r_\varepsilon} I_{r_\varepsilon} \end{aligned} \quad (10)$$

and $\text{tr } Z = 1$. Thus it follows that

$$\langle P, A(y) \rangle = \langle Z, Q_\varepsilon(A(y))^T T Q_\varepsilon(A(y)) \rangle \geq \lambda_{r_\varepsilon}. \quad (11)$$

Together with (3), we have $\langle P, A(y) \rangle \geq \lambda_{\max}(A(y)) - \varepsilon$; for $P \in C_n$ this means that, according to (6), $P \in \partial_\varepsilon \lambda_{\max}(A(y))$. Accordingly, the proof of right inclusion is completed. \square

Note that the compact convex set $\sigma_\varepsilon \lambda_{\max}(A(y))$ can be regarded as an outer approximation of $\partial \lambda_{\max}(A(y))$ and an inner approximation of $\partial_\varepsilon \lambda_{\max}(A(y))$. This set will avoid both the weakness of $\partial \lambda_{\max}(A(y))$ (unstable) and the drawback of $\partial_\varepsilon \lambda_{\max}(A(y))$ (too much elements). Applying the linear mapping β^* in (8). Then

$$\beta^* \partial \lambda_{\max}(A(y)) \subset \beta^* \sigma_\varepsilon \lambda_{\max}(A(y)) \subset \beta^* \partial_\varepsilon \lambda_{\max}(A(y)). \quad (12)$$

Set $\sigma_\varepsilon f(y) := \beta^* \sigma_\varepsilon \lambda_{\max}(A(y))$. Then we obtain

$$\partial \lambda_{\max} \circ A(y) \subset \sigma_\varepsilon \lambda_{\max} \circ A(y) \subset \partial_\varepsilon \lambda_{\max} \circ A(y). \quad (13)$$

This formula is equivalent to

$$\partial f(y) \subset \sigma_\varepsilon f(y) \subset \partial_\varepsilon f(y). \quad (14)$$

Add $\partial g(y)$ simultaneously, since $\text{ri dom } f(y) \cap \text{ri dom } g(y) \neq \emptyset$, we have

$$\begin{aligned} \partial F(y) &= \partial f(y) + \partial g(y) \subset \sigma_\varepsilon f(y) \\ &+ \partial g(y) \subset \partial_\varepsilon f(y) + \partial_\varepsilon g(y) = \partial_\varepsilon F(y). \end{aligned} \quad (15)$$

Set $\sigma_\varepsilon F(y) := \sigma_\varepsilon f(y) + \partial g(y)$. Thus $\partial F(y) \subset \sigma_\varepsilon F(y) \subset \partial_\varepsilon F(y)$. Finally, $\sigma_\varepsilon F(y)$ is the enlarged subdifferential of the objective function $F(y)$ and it can be regarded as an outer approximation of $\partial F(y)$ and an inner approximation of $\partial_\varepsilon F(y)$. The following work will base on the enlarged subdifferential $\sigma_\varepsilon F(y)$.

Choose $\beta^* Q_\varepsilon(A(x^i)) Z Q_\varepsilon(A(x^i))^T \in \sigma_\varepsilon f(x^i)$ and $s^i \in \partial g(x^i)$, then

$$\beta^* Q_\varepsilon(A(x^i)) Z Q_\varepsilon(A(x^i))^T + s^i \in \sigma_\varepsilon F(x^i) \subset \partial_\varepsilon F(x^i). \quad (16)$$

Next at each point x^i compute an approximate function value $F_\varepsilon(x^i)$ satisfying

$$F(x^i) - \varepsilon \leq F_\varepsilon(x^i) \leq F(x^i). \quad (17)$$

Thus, using these approximate information of $F(y)$, the approximate model $\bar{F}_k(y)$ becomes

$$\begin{aligned} \bar{F}_k(y) \\ = \max_{i=1, \dots, k} \left\{ F_\varepsilon(x^i) + \left\langle \beta^* Q_\varepsilon(A(x^i)) Z Q_\varepsilon(A(x^i))^T \right. \right. \\ \left. \left. + s^i, y - x^i \right\rangle \right\}. \end{aligned} \quad (18)$$

Let the aggregate linearization error at x^k be denoted by e^i :

$$\begin{aligned} e^i &:= F(x^k) - F_\varepsilon(x^i) \\ &- \left\langle \beta^* Q_\varepsilon(A(x^i)) Z Q_\varepsilon(A(x^i))^T + s^i, x^k - x^i \right\rangle. \end{aligned} \quad (19)$$

With the notation, we obtain the form of approximate model

$$\begin{aligned} \bar{F}_k(y) &= F(x^k) \\ &+ \max_{i=1, \dots, k} \left\{ -e^i + \left\langle \beta^* Q_\varepsilon(A(x^i)) Z Q_\varepsilon(A(x^i))^T + s^i, \right. \right. \\ &\quad \left. \left. y - x^k \right\rangle \right\}, \end{aligned} \quad (20)$$

where $Z \in C_{r_\varepsilon}$.

Observe that $\bar{F}_k(y)$ is a more stable approximate model than $\tilde{F}_k(y)$. In the following part, based on the approximate model $\bar{F}_k(y)$, we will give the approximate proximal bundle method algorithm.

3. The Approximate Proximal Bundle Method Algorithm

Algorithm 3 (the approximate proximal bundle method).

Step 0. Let $\theta \geq 0$, $m \in (0, 1)$ be given parameters. Choose x^1 , we can obtain an approximate function value $F_\varepsilon(x^1)$ and approximate subgradient $\beta^* Q_\varepsilon(A(x^1)) Z Q_\varepsilon(A(x^1))^T + s^1 \in \sigma_\varepsilon F(x^1) \subset \partial_\varepsilon F(x^1)$. Then construct the approximate model and set $k = 1$, $\delta_1 = \infty$.

Step 1. If $\delta_k \leq \theta$, stop.

Step 2. Solve the quadratic program:

$$\min_{y \in R^n} \bar{F}_k(y) + \frac{1}{2} \eta_k \|y - x^k\|^2. \quad (P_1)$$

Define δ_{k+1} as the nominal decrease:

$$\delta_{k+1} := F(x^k) - \bar{F}_k(y^{k+1}) - \frac{1}{2} \eta_k \|y^{k+1} - x^k\|^2. \quad (21)$$

Step 3. Call the black box again with $y = y^{k+1}$, if

$$F(x^k) - F(y^{k+1}) \geq m \delta_{k+1}. \quad (22)$$

Then set $x^{k+1} = y^{k+1}$, otherwise set $x^{k+1} = x^k$. Corresponding, the former one is said to be descent step and the latter one is called null step.

Step 4. Append y^{k+1} to the bundle model and construct \bar{F}_{k+1} . Change k to $k + 1$ and go to Step 1.

Remark 4. For a parameter $\eta_k > 0$, then the candidate point y^{k+1} can be obtained through solving the dual problem of (P_1) .

Theorem 5. If y^{k+1} is the unique solution to (P_1) and suppose that $\eta_k > 0$, then $y^{k+1} = x^k - (1/\eta_k) \sum_{i=1}^{np_k} \bar{\alpha}_i (\beta^* Q_\varepsilon(A(x^i)))^T \times ZQ_\varepsilon(A(x^i))^T + s^i$ and $\bar{\alpha} = (\bar{\alpha}_1, \dots, \bar{\alpha}_{np_k})$ is a solution to

$$\begin{aligned} \min & \frac{1}{2\eta_k} \left\| \sum_{i=1}^{np_k} \alpha_i \left(\beta^* Q_\varepsilon(A(x^i)) ZQ_\varepsilon(A(x^i))^T \right. \right. \\ & \left. \left. + s^i \right) \right\|^2 + \sum_{i=1}^{np_k} \alpha_i e^i \\ \text{s.t. } & \alpha \in \Delta_k = \left\{ \alpha_i \in [0, 1], \sum_{i=1}^{np_k} \alpha_i = 1, i = 1, \dots, np_k \right\}. \end{aligned} \quad (D_1)$$

In addition, one can also have the following relations:

- (i) $\sum_{i=1}^{np_k} \bar{\alpha}_i (\beta^* Q_\varepsilon(A(x^i)) ZQ_\varepsilon(A(x^i))^T + s^i) \in \partial_\varepsilon \bar{F}_k(y^{k+1});$
- (ii) $\delta_{k+1} = \epsilon_k + (1/2\eta_k) \left\| \sum_{i=1}^{np_k} \bar{\alpha}_i (\beta^* Q_\varepsilon(A(x^i)) ZQ_\varepsilon(A(x^i))^T + s^i) \right\|^2$, where $\epsilon_k = \sum_{i=1}^{np_k} \bar{\alpha}_i e^i$;
- (iii) $\sum_{i=1}^{np_k} \bar{\alpha}_i (\beta^* Q_\varepsilon(A(x^i)) ZQ_\varepsilon(A(x^i))^T + s^i) \in \partial_{\epsilon_k} F(x^k).$

Proof. Turn (P_1) into a quadratic programming problem with an extra scalar variable r :

$$\begin{aligned} \min_{(y,r) \in R^n \times R^r} & r + \frac{1}{2} \eta_k \|y - x^k\|^2 \\ \text{s.t. } & F(x^k) - e^i \\ & + \left\langle \beta^* Q_\varepsilon(A(x^i)) ZQ_\varepsilon(A(x^i))^T + s^i, y - x^k \right\rangle \\ & \leq r, \quad i = 1, 2, \dots, np_k. \end{aligned} \quad (P_2)$$

The corresponding lagrangian is, for $\alpha \in R_+^{np_k}$,

$$\begin{aligned} L(y, \alpha, r) \\ = r + \frac{1}{2} \eta_k \|y - x^k\|^2 \\ + \sum_{i=1}^{np_k} \alpha_i \left(F(x^k) - e^i \right. \\ \left. + \left\langle \beta^* Q_\varepsilon(A(x^i)) ZQ_\varepsilon(A(x^i))^T + s^i, \right. \right. \\ \left. \left. y - x^k \right\rangle - r \right). \end{aligned} \quad (23)$$

By strong convexity, the dual problem of (P_1) is equivalent to problem (D_1) .

And $y^{k+1} = x^k - (1/\eta_k) \sum_{i=1}^{np_k} \bar{\alpha}_i (\beta^* Q_\varepsilon(A(x^i)) ZQ_\varepsilon(A(x^i))^T + s^i)$ is the solution to the quadratic programming problem (P_1) , then (i) holds.

To see (ii) and (iii), note that since there is no duality gap, the primal optimal value in (P_1) equals the dual optimal value in (D_1) . Thus the term (ii) holds. The relation $F(y) \geq \bar{F}_k(y) \geq \bar{F}_k(y^{k+1})$ implies that the term (iii) holds. \square

From Theorem 5 we can ensure the candidate point y^{k+1} which appears in Step 2 can be obtained.

Remark 6. As iterations conduct, the elements in the bundle become more and more. When the size of the bundle becomes too big, it is necessary to compress bundle. So, at Step 4, one should append the compression subalgorithm.

When the current size of the bundle is bigger than the maximal size, that is, $np_k \geq np_{\max}$, Step 4 turns out to be as follows:

Step 4'. Let $n_{\text{act}} := \{i \leq np_k : \bar{\alpha}_i \geq 0\}$ be the active indices.

If $n_{\text{act}} \leq np_{\max} - 1$, then keep active couples and delete all inactive couples from the bundle. Set $n_{\text{left}} = n_{\text{act}}$ and define $np_{k+1} = n_{\text{left}} + 1$. Then append new element

$$\left(\beta^* Q_\varepsilon(A(x^{np_{k+1}})) ZQ_\varepsilon(A(x^{np_{k+1}}))^T + s^{np_{k+1}}, e^{np_{k+1}} \right) \quad (24)$$

to the bundle and construct model \bar{F}_{k+1} . Let $k = k + 1$, and go to Step 1. Note that $e^{np_{k+1}}$ is in new element:

(i) when it is a descent step, $e^{np_{k+1}} = 0$,

(ii) when it is a null step,

$$\begin{aligned} e^{np_{k+1}} &= F(x^k) - F_\varepsilon(y^{k+1}) \\ &\quad - \left\langle \beta^* Q_\varepsilon(A(x^{np_{k+1}})) \right. \\ &\quad \left. \times ZQ_\varepsilon(A(x^{np_{k+1}}))^T + s^{np_{k+1}}, x^k - y^{k+1} \right\rangle. \end{aligned} \quad (25)$$

If $n_{\text{act}} > np_{\max} - 1$, then delete two or more couples of elements. In addition, $n_{\text{left}} \leq np_{\max} - 2$. Define $np_{k+1} = np_{\text{left}} + 2$ and then append new element

$$\left(\beta^* Q_\varepsilon(A(x^{np_{k+1}})) ZQ_\varepsilon(A(x^{np_{k+1}}))^T + s^{np_{k+1}}, e^{np_{k+1}} \right) \quad (26)$$

to the bundle and construct model \bar{F}_{k+1} . Let $k = k + 1$, and go to Step 1. Note that $e^{np_{k+1}}$ is in new element:

(i) when it is a descent step, $e^{np_{k+1}} = 0$,

(ii) when it is a null step,

$$\begin{aligned} e^{np_{k+1}} &= F(x^k) - F_\varepsilon(y^{k+1}) \\ &\quad - \left\langle \beta^* Q_\varepsilon(A(x^{np_{k+1}})) ZQ_\varepsilon(A(x^{np_{k+1}}))^T \right. \\ &\quad \left. + s^{np_{k+1}}, x^k - y^{k+1} \right\rangle. \end{aligned} \quad (27)$$

Remark 7. Under certain circumstance with $np_k > np_{\max}$, if the remaining couples are still too many after discarding all inactive couples from the bundle, one synthesizes indispensable information of active elements in bundle.

Simultaneously the corresponding affine function is called aggregate linearization and is denoted by $F_\alpha(y)$

$$F_\alpha(y) := F(x^k) - \epsilon_k + \left\langle \sum_{i=1}^{np_k} \bar{\alpha}_i \left(\beta^* Q_\epsilon(A(x^i)) ZQ_\epsilon(A(x^i))^T + s^i \right), y - x^k \right\rangle. \quad (28)$$

For the aggregate linearization $F_\alpha(y)$, it holds that

$$F_\alpha(y) = \bar{F}_k(y^{k+1}) + \left\langle \sum_{i=1}^{np_k} \bar{\alpha}_i \left(\beta^* Q_\epsilon(A(x^i)) ZQ_\epsilon(A(x^i))^T + s^i \right), y - y^{k+1} \right\rangle; \quad (29)$$

$$\bar{F}_k(y) \geq F_\alpha(y) - \epsilon, \quad \forall y \in \mathbb{R}^n.$$

When the maximum capacity is reached, for instance, if $k = np_{\max}$, then assume that one discards the elements x^1, x^2, \dots, x^t ($t < k$) from the bundle and appends the aggregate couple. The resulting model will be

$$\bar{F}_{k+1}(y) = \max \left\{ \max_{t+1 \leq i \leq k+1} \{F_\epsilon(x^i) + \langle m^i, y - x^i \rangle\}, F_\alpha(y) \right\}, \quad (30)$$

where $m^i = \beta^* Q_\epsilon(A(x^i)) ZQ_\epsilon(A(x^i))^T + s^i$. Note that for all k and $y \in \mathbb{R}^n$, in any case, one can have

$$F_\alpha(y) \leq \bar{F}_{k+1}(y) \leq F(y);$$

$$\bar{F}_{k+1}(y) \geq F(x^{k+1}) + \left\langle \beta^* Q_\epsilon(A(x^{k+1})) ZQ_\epsilon(A(x^{k+1}))^T + s^{k+1}, y - y^{k+1} \right\rangle. \quad (31)$$

4. Convergence Analysis

To show convergence of the algorithm, we have to refer the stopping tolerance θ . So we consider two situations, that is, $\theta > 0$ and $\theta = 0$.

Firstly, when the parameter $\theta > 0$, use the notation K_s to denote the set of indices k in which a new descent step is done.

Theorem 8. Consider Algorithm 3 and use the notation F^* to denote $\lim_{k \in K_s} F(x^k)$. Assume that the algorithm never stops ($k \rightarrow \infty$) as well as $F^* > -\infty$. Then

$$0 \leq \sum_{k \in K_s} \delta_k \leq \frac{F(x^1) - F^*}{m}. \quad (32)$$

Proof. Since $\theta > 0$ and the algorithm never stops, the nominal decrease must satisfy $\delta_k > 0$ for all $k \in K_s$. Note that K_s is a descent index set, we have $x^{k+1} = y^{k+1}$ and $F(x^k) - F(x^{k+1}) = F(x^k) - F(y^{k+1})$. Let k' be the index following k in K_s . Between k and k' the algorithm makes null steps only without moving the stability center $x^{k+1} = y^{k+j}$ for all $j = 2, \dots, k' - k$. The descent test at k' gives

$$F(x^{k+1}) - F(x^{k'+1}) \geq m\delta_{k'+1}. \quad (33)$$

Thus for all $k'' \in K_s$, it holds that when we let $k'' \rightarrow \infty$,

$$m \sum_{k \in K_s}^{k''} \leq \sum_{k \in K_s}^{k''} F(x^k) - F(x^{k+1}) = F(x^1) - F(x^{k''+1}) = F(x^1) - F^*. \quad (34)$$

Removing m to the right side of the inequation, we obtain the desired result. \square

When θ is taken strictly positive, by Theorem 8, there is an index \bar{k} for which $\delta_{\bar{k}} \leq \theta$ if (P) has minimizers. By Theorem 5(ii), both ϵ_k and

$$\frac{1}{\eta_{k_{\text{last}}}} \left\| \sum_{i=1}^{np_k} \bar{\alpha}_i \left(\beta^* Q_\epsilon(A(x^{\bar{k}})) ZQ_\epsilon(A(x^{\bar{k}}))^T + s^{\bar{k}} \right) \right\|^2 \quad (35)$$

are all small. Therefore, $x^{\bar{k}}$ is the minimizer.

Secondly, when the stopping tolerance $\theta = 0$, the algorithm either stops by having found a solution to (P) or it never stops. In this case, there are two possibilities for the sequence of descent steps $\{x^k\}_{k \in K_s}$. One is that it has infinitely many elements. Another is that there is an iteration \bar{k} where a last descent step is done, that is, $x^k = x^{\bar{k}}$ for all $k \geq \bar{k}$. We proof these two cases separately.

Case 1. There are infinitely many elements in K_s .

Theorem 9. Suppose that the algorithm generates infinitely many descent steps x^k for all $k \in K_s$. One has

- (i) if (P) has an empty solution, then $\{f(x^k)\} \searrow 0$;
- (ii) if (P) has minimizers and $\sum_{k \in K_s} (1/\eta_k) = +\infty$ with $k \in K_s$ and $k \rightarrow \infty$, the sequence $\{x^k\}$ is minimizing for (P) and converges to a minimizer of (P).

Proof. (i) Note that the algorithm loops forever and $\theta = 0$ for all $k \in K_s$, thus δ_{k+1} holds. Then the infinite sequence of objective values $\{f(x^k)\}$ is strictly decreasing. If (P) has no

solution, the sequence $\{f(x^k)\}$ is close to $-\infty$. Therefore the proof of (i) is completed.

(ii) To show that $\{x^k\}$ is a minimizing sequence. Suppose for contradiction purposes that there exists $\bar{x} \in R^n$ and $t > 0$ such that $F(\bar{x}) \leq F(x^k) - t$ for all $k \in K_s$. By Theorem 5(iii), we have

$$\left\langle x - x^k, \sum_{i=1}^{np_k} \bar{\alpha}_i \left(\beta^* Q_\varepsilon(A(x^i)) ZQ_\varepsilon(A(x^i))^T + s^i \right) \right\rangle \leq F(x) - F(x^k) + \epsilon_k. \quad (36)$$

By $x^{k+1} = x^k - (1/\eta_k) \sum_{i=1}^{np_k} \bar{\alpha}_i (\beta^* Q_\varepsilon(A(x^i)) ZQ_\varepsilon(A(x^i))^T + s^i)$ and Theorem 5(ii), hence

$$\begin{aligned} \|x^{k+1} - x\|^2 &= \left\| x^k - \frac{1}{\eta_k} \sum_{i=1}^{np_k} \bar{\alpha}_i \left(\beta^* Q_\varepsilon(A(x^i)) ZQ_\varepsilon(A(x^i))^T + s^i \right) - x \right\|^2 \\ &\leq \|x^k - x\|^2 + \frac{2}{\eta_k} (F(x) - F(x^k) + \epsilon_k) + \frac{2}{\eta_k} (\delta_{k+1} - \epsilon_k) \\ &= \|x^k - x\|^2 + \frac{2}{\eta_k} (F(x) - F(x^k) + \delta_{k+1}). \end{aligned} \quad (37)$$

Write the relation (37) for $x = \bar{x}$, we obtain

$$\begin{aligned} 0 &\leq \|x^{k+1} - \bar{x}\|^2 \\ &\leq \|x^k - \bar{x}\|^2 + \frac{2}{\eta_k} (F(\bar{x}) - F(x^k) + \delta_{k+1}) \\ &\leq \|x^k - \bar{x}\|^2 + \frac{2}{\eta_k} (-t + \delta_{k+1}). \end{aligned} \quad (38)$$

Since $\{\delta_k\} \rightarrow 0$ with $k \rightarrow \infty$, there exists k_t such that $\delta_{k+1} < t/2$ for all $k \geq k_t$. Thus the relation (38) becomes

$$\begin{aligned} 0 &\leq \|x^{k+1} - \bar{x}\|^2 \\ &\leq \|x^k - \bar{x}\|^2 - \frac{t}{\eta_k} \quad \forall k \in K_s, k \geq k_t. \end{aligned} \quad (39)$$

Summing the inequations over $k_t \leq k \in K_s$ yields

$$0 \leq \|x^{k_t} - \bar{x}\|^2 - t \sum_{k \geq k_t, k \in K_s} \frac{1}{\eta_k}. \quad (40)$$

Letting $k \rightarrow \infty$, we obtain a contradiction for the divergence assumption $\sum (1/\eta_k) = +\infty$. Therefore, the sequence $\{x^k\}$ is a minimizing sequence.

To see that $\{x^k\}$ converges to a minimizer of (P), take in (37) $x = \bar{x}$, a solution to (P), and sum over $k \in K_s$. It holds that $F(\bar{x}) \leq F(x^k)$ for all k . We see that

$$\begin{aligned} \|x^k - \bar{x}\|^2 &\leq \|x^1 - \bar{x}\|^2 + \frac{2}{\eta_{\min}} \sum_{k \in K_s} \delta_k \\ &\leq \|x^1 - \bar{x}\|^2 + \frac{2(F(x^1) - F^*)}{\eta_{\min} m}. \end{aligned} \quad (41)$$

Then the sequence $\{x^k\}$ is bounded. Take a subsequence $\{x^{k^i}\}_{k^i \in K_s}$ converging to \bar{x} as $i \rightarrow \infty$. Given $u > 0$, take i big enough such that

$$\|x^{k^i} - \bar{x}\|^2 \leq \frac{u}{2}, \quad \sum_{k \in K_s, k \geq k^i} \delta_{k+1} \leq \frac{1}{2} u \eta_k. \quad (42)$$

Writing (37) with $x = \bar{x}$ and summing it over $k \in K_s$ from k^i to an arbitrary $\tilde{k} > k^i$ yield

$$\begin{aligned} 0 &\leq \|x^{\tilde{k}+1} - \bar{x}\|^2 \\ &\leq \|x^{k^i} - \bar{x}\|^2 + \frac{1}{\eta_k} \sum_{k \in K_s, k \geq k^i} \delta_{k+1} \leq u, \end{aligned} \quad (43)$$

which implies that $\{x^k\}$ converges to a minimizer of (P). \square

Case 2. There are finitely many descent steps. It means that the last descent step $x^{\tilde{k}}$ is followed by an infinite number of null step, that is, $x^k = x^{\tilde{k}}$ for all $k \geq \tilde{k}$.

Theorem 10. Assume that Algorithm 3 generates a last descent step $x^{\tilde{k}}$ followed by infinitely many null steps. Then sequence $\{x^k\}$ converges to $x^{\tilde{k}}$ and $x^{\tilde{k}}$ minimizes the function F .

Proof. Let y^{k+1} be the solution to (P_1) , we have

$$\bar{F}_k(y^{k+1}) + \frac{1}{2} \eta_k \|y^{k+1} - x^{\tilde{k}}\|^2 \leq \bar{F}_k(x^{\tilde{k}}) \leq F(x^{\tilde{k}}). \quad (44)$$

Set

$$\begin{aligned} M_k(y) &:= \bar{F}_k(y^{k+1}) + \frac{1}{2} \eta_k \|y - x^{\tilde{k}}\|^2 \\ &\quad + \frac{1}{2} \eta_k \|y^{k+1} - y\|^2. \end{aligned} \quad (45)$$

Then

$$M_k(y^{k+1}) \leq F(x^{\tilde{k}}) \quad \forall k \leq \tilde{k}. \quad (46)$$

By (30), we have

$$M_{k+1}(y^{k+2}) \geq M_k(y^{k+1}) + \frac{1}{2} \eta_k \|y^{k+2} - y^{k+1}\|^2; \quad (47)$$

$$\begin{aligned} M_k(y^{k+1}) &+ \frac{1}{2} \eta_k \|y^{k+1} - x^{\tilde{k}}\|^2 \\ &= \bar{F}_k(y^{k+1}) + \eta_k \|y^{k+1} - x^{\tilde{k}}\|^2 \\ &\leq F_\alpha(x^{\tilde{k}}) \leq F(x^{\tilde{k}}). \end{aligned} \quad (48)$$

Thus, the sequence $\{y^{k+1}\}$ is bounded. By the relation (47), we have

$$\{y^{k+1} - y^k\} \rightarrow 0 \quad \text{as } k \rightarrow \infty. \quad (49)$$

It is obvious from the convexity that

$$\bar{F}_k(y^{k+1}) - F(y^k) \leq F(y^{k+1}) - Fy^k \leq L\|y^{k+1} - y^k\|^2. \quad (50)$$

So, it holds that

$$\bar{F}_k(y^{k+1}) - F(y^k) \rightarrow 0. \quad (51)$$

From the bounded sequence $\{y^k\}$ extract a subsequence $\{y^{k^i}\} \rightarrow \bar{y}$ as $i \rightarrow \infty$, for $\{y^{k+1} - y^k\} \rightarrow 0$ and $\{y^{k^i+1} - \bar{y}\} \rightarrow 0$; thus

$$F(y^{k^i+1}) - \bar{F}_{k^i}(y^{k^i+1}) \rightarrow 0 \quad \text{as } i \rightarrow \infty. \quad (52)$$

Then

$$\bar{F}_{k^i}(y^{k^i+1}) \rightarrow F(\bar{y}). \quad (53)$$

When $k \geq \tilde{k}$, the descent test is never satisfied, that is, $F(y^{k+1}) - F(x^{\tilde{k}}) > -m\delta_{k+1}$, adding δ_{k+1} to the both sides of the inequation and using the definition of δ_{k+1} in (21), we can obtain

$$0 \leq (1 - m)\delta_{k+1} \leq F(y^{k+1}) - \bar{F}_{k^i}(y^{k^i+1}). \quad (54)$$

Let $i \rightarrow \infty$, then $\delta_{k+1} \rightarrow 0$. In addition, as $i \rightarrow \infty$

$$\epsilon_k \rightarrow 0,$$

$$\frac{1}{2\eta_k} \left\| \sum_{i=1}^{np_k} \bar{\alpha}_i (\beta^* Q_\epsilon(A(x^i)) ZQ_\epsilon(A(x^i))^T + s^i) \right\|^2 \rightarrow 0. \quad (55)$$

By Theorem 5(iii), it holds that

$$\begin{aligned} F(y) &\geq F(x^{\tilde{k}}) \\ &+ \left\langle \sum_{i=1}^{np_k} \bar{\alpha}_i (\beta^* Q_\epsilon(A(x^i)) ZQ_\epsilon(A(x^i))^T + s^i), \right. \\ &\quad \left. y - x^{\tilde{k}} \right\rangle - \epsilon_{k^i}. \end{aligned} \quad (56)$$

Passing to the limit as $i \rightarrow \infty$ the inequation shows that $\{x^{\tilde{k}}\}$ minimizes the function F .

This is to show that $\bar{y} = x^{\tilde{k}}$. Since

$$\begin{aligned} F(\bar{y}) &\geq F(x^{\tilde{k}}) \\ &\geq \bar{F}_{k^i}(x^{\tilde{k}}) \\ &\geq \bar{F}_{k^i}(y^{k^i+1}) + \frac{1}{2}\eta_{k^i+1}\|y^{k^i+1} - x^{\tilde{k}}\|^2 \\ &\geq \bar{F}_{k^i}(y^{k^i+1}) + \frac{1}{2}\eta_{\tilde{k}}\|y^{k^i+1} - x^{\tilde{k}}\|^2, \end{aligned} \quad (57)$$

passing to the limit as $i \rightarrow \infty$, we obtain

$$F(y) \geq \bar{F}(y) + \frac{1}{2}\eta_{\tilde{k}}\|\bar{y} - x^{\tilde{k}}\|^2. \quad (58)$$

The inequality holds if and only if $\bar{y} = x^{\tilde{k}}$. Thus $x^{\tilde{k}}$ is the minimizer. \square

5. Conclusions

In this paper, we have given an enlarged subdifferential and constructed a more stable approximate cutting-plane model for the objective function. Then we have proposed an approximate proximal bundle algorithm and shown how to use this algorithm to minimize a class of maximum eigenvalue functions. Finally, it may be possible to extend this method to an even larger class of constrained maximum eigenvalue function.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank the referees for the valuable discussion and recommendation for improving the paper. This work is supported by the National Natural Science Foundation of China under Project no.11171138.

References

- [1] C. Helmberg and F. Oustry, "Bundle methods to minimize the maximum eigenvalue function," in *Handbook of Semidefinite Programming*, vol. 27 of *International Series in Operations Research & Management Science*, pp. 307–337, Kluwer Academic, Boston, Mass, USA, 2000.
- [2] C. Sagastizabal and M. Solodov, "An infeasible bundle method for nonsmooth convex constrained optimization without a penalty function or a filter," *SIAM Journal on Optimization*, vol. 16, no. 1, pp. 146–169, 2005.
- [3] J. F. Bonnans, J. C. a. Gilbert, and C. A. Sagastizabal, *Numerical Optimization*, Springer, Berlin, Germany, 2006.
- [4] W. Wang, M. Chen, and L. L. Zhang, "Solving a minimization problem for a class of constrained maximum eigenvalue function," *International Journal of Pure and Applied Mathematics*, vol. 91, no. 3, pp. 291–303, 2014.
- [5] C. Lemarechal and F. Oustry, *Nonsmooth Algorithms to Solve Semidefinite Programs*, Society for Industrial and Applied Mathematics, Philadelphia, Pa, USA, 2000.
- [6] F. Oustry, "A second-order bundle method to minimize the maximum eigenvalue function," *Mathematical Programming*, vol. 89, no. 1, pp. 1–33, 2000.