doi：103969/j. issn. 0490-6756. 2019. 01. 003

一种部分非精确求解可分离凸优化问题的渐近点算法

陈小彪[1]，李耿华[2]，张玫玉[3]
(1. 太原工业学院理学系，太原 030008；2. 重庆大学数学与统计学院，重庆 401331)

摘　要：本文研究了一类具有可分离结构的凸优化问题,在经典的交替方向法的基础上得到了一种部分非精确的渐近点算法. 该方法分别求解凸优化问题的两个子问题,其中一个直接求解,另一个通过引入非精确项降低了求解的难度. 在合理的假设下,新算法的收敛性得到了证明. 数值实验表明新算法是有效的.

关键词：凸优化问题；结构型变分不等式；交替方向法；渐近点算法；预测-校正步法
中图分类号：O221. 2　　**文献标识码**：A　　**文章编号**：0490-6756(2019)01-0008-05

A partial inexact proximal point method for separable convex programming

CHEN Xiao-Biao[1], *LI Geng-Hua*[2], *ZHANG Mei-Yu*[1]
(1. Science Department, Taiyuan Institute of Technology, Taiyuan 030008, China;
2. College of Mathematics and Statistics, Chongqing University, Chongqing 401331, China)

Abstract: In this paper, a new method is proposed for solving a class of separable convex programming problem. The method is referred to as the partial inexact proximal point method. In the method, we take a fresh look at the alternating direction method of multipliers and two subproblems are solved independently. One is solved directly and the other is handled by bring in inexact minimization. Convergence of the method is proved under mild assumptions and its efficiency is also verified by numerical experiments

Keywords: Convex programming; Structured variational inequality; Alternating direction method; Proximal point method; Prediction-correction method
(2010 MSC 65K10, 90C25, 90C33)

1　Introduction

In this paper, we consider the following convex optimization problem:

$$\min \theta_1(x)+\theta_2(y), \text{ s. t. } Ax-y=0 \tag{1}$$

where $x\in X, y\in Y, X$ and Y are closed convex subset of $\mathbf{R}^n$ and $\mathbf{R}^m$ respectively, $A\in\mathbf{R}^{m\times n}$ is a given constant matrix. The functions θ_1 and θ_2 are closed proper convex functions on X and Y respectively. Let $f(x)=\partial\theta_1(x)$ and $g(y)=\partial\theta_2(y)$ be the subgradient of $\theta_1(x)$ and $\theta_2(y)$ respectively. By convexity of $\theta_1(x)$ and $\theta_2(y)$, we know that $f(x)$ and $g(y)$ are monotone with respect to X and Y respectively. Then problem(1)is equivalent to the following monotone structured variational inequality: find $(x,y)\in\Omega$, such that

$$\begin{cases}(x'-x)^T f(x)\geqslant 0,\\ (y'-y)^T g(y)\geqslant 0\end{cases} \tag{2}$$

$\forall (x',y')\in\Omega$, where

$$\Omega=\{(x,y|x\in X,y\in Y:Ax-y=0\} \tag{3}$$

收稿日期：2017-12-12
基金项目：太原工业学院青年基金(2015LQ16)
作者简介：陈小彪，男，主要从事最优化理论与算法的研究. E-mail：imchxb@126. com

By attaching a Lagrangian multiplier $\lambda \in \mathbf{R}^m$ to the linear $Ax-y$, problem (2)-(3) can be reformulated into the following equivalent form:

$$w\in W,\begin{cases}(x'-x)T(f(x)-A^T\lambda)\geqslant 0,\\(y'-y)T(g(y)+\lambda)\geqslant 0,\\(\lambda'-\lambda)T(Ax-y)\geqslant 0\end{cases}\qquad(4)$$

$\forall w'\in W$, where $w=(x,y,\lambda)$ and $W=X\times Y\times\mathbf{R}^m$.

In practical applications, there exist quite a few structured optimizations arising from the fields such as electronic engineering and computer science, including digital signal processing[1], signal enhancement[2], natural imaging processing[3], matrix processing[4] and traffic network analysis[5], etc. A classical method for solving (4) is alternating direction method, which was proposed originally in Ref. [6], and studied intensively in the literature, see e. g. Refs. [7～11]. Especially, Ye and Yuan[8] developed a variant of alternating direction method with an optimal stepsize. Given a couple of (y^k,λ^k), the new iterate of Ye-Yuan's algorithm is produced by

$$\begin{cases}(x'-\tilde{x}^k)T\{f(\tilde{x}^k)-A^T[\lambda^k-H(A\tilde{x}^k-y^k)]\}\geqslant 0,\\(y'-\tilde{y}^k)T\{g(\tilde{y}^k)+[\lambda^k-H(A\tilde{x}^k-\tilde{y}^k)]\}\geqslant 0,\\\tilde{\lambda}^k=\lambda^k-H(A\tilde{x}^k-\tilde{y}^k),\end{cases}$$

$\forall x'\in X$, $\forall y'\in Y$, and

$$\begin{cases}y^{k+1}=y^k-\gamma\alpha^*(y^k-\tilde{y}^k),\\\lambda^{k+1}=\lambda^k-\gamma\alpha^*(\lambda^k-\tilde{\lambda}^k),\end{cases}$$

where $\gamma\in(0,2)$ and α^* is the optimal stepsize.

However, the alternating direction method may fail since the subproblems are hard to be solved exactly in many practical applications. He and Liao *et al.*[12] suggested a method for solving subproblems (4) inexactly. This method is referred to as alternating projection based prediction correction methods (abbreviated as APBPCM), for a given triplet $w^k=(x^k,y^k,\lambda^k)$, the new iterate $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})$ of APBPCM is produced by the following two phases.

Prediction phase. Let H be a given positive matrix. The predictor $\tilde{w}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{\lambda}^k)$ is generated by the following procedure:

(i) Set

$$\tilde{x}^k=P_X\{x^k-\frac{1}{r_k}[f(x^k)-A^T(\lambda^k-H(Ax^k-y^k))]\},$$

where $r_k>0$ is a chosen parameter such that

$$\|\xi_x^k\|\leqslant vr_k\|x^k-\tilde{x}^k\|,$$

$$\xi_x^k=f(x^k)-f(\tilde{x}^k)+A^THA(x^k-\tilde{x}^k);$$

(ii) Set

$$\tilde{y}^k=P_Y\{y^k-\frac{1}{s_k}[g(y^k)+(\lambda^k-H(A\tilde{x}^k-y^k))]\},$$

where $s_k>0$ is a chosen parameter such that

$$\|\xi_y^k\|\leqslant vs_k\|y^k-\tilde{y}^k\|,$$

$$\xi_y^k=g(y^k)-g(\tilde{y}^k)+H(y^k-\tilde{y}^k);$$

(iii) Update $\tilde{\lambda}^k$ via $\tilde{\lambda}^k=\lambda^k-H(A\tilde{x}^k-\tilde{y}^k)$.

Correction phase.

$$w^{k+1}=w^k-\gamma\alpha d(w^k,\tilde{w}^k\xi^k),$$

where

$$d(w^k,\tilde{w}^k,\xi^k)=(w^k-\tilde{w}^k)-G_k^{-1}\xi^k,$$

and

$$G_k=\begin{bmatrix}r_kI&0&0\\0&s_kI+H&0\\0&0&H^{-1}\end{bmatrix},\xi^k=\begin{bmatrix}\xi_x^k\\\xi_y^k\\0\end{bmatrix}.$$

In recent years, the inexact alternating direction method have been studied intensively in the literature, see e. g. Refs. [13～18]. Accordingly, the following notations regarding the variables y and λ will simplify our analysis:

$$v=(y,\lambda);\ V=Y\times\mathbf{R}^m;v^k=(y^k,\lambda^k);$$

$$\tilde{v}^k=(\tilde{y}^k,\tilde{\lambda}^k),\forall k\in\mathbf{N};$$

$$V^*=\{(y^*,\lambda^*)|(x^*,y^*,\lambda^*)\in\Omega^*\}.$$

The rest of this paper is organized as follows. In Section 2, we summarize some properties and assumptions which will be useful for subsequent section. In Section 3, we present our new method. In Section 4, we analyze its convergence under some mild conditions. Preliminary numerical results are reported in Section 5. Finally, some concluding remarks are drawn in Section 6.

2 Preliminaries

In this section, we summarize some basic properties which play significant roles for further analysis.

Let $\|\cdot\|:=\|\cdot\|_2$ denote the Euclidean *norm*, $\| x \| = \sqrt{x^T x}$ for any $x \in \mathbf{R}^n$. For the nonempty closed convex set Ω, we denote by $P_\Omega(\cdot)$ the projection onto Ω under the Euclidean norm:

$$P_\Omega(x) = \operatorname{argmin}\{ \| x - y \| \mid y \in \Omega \}.$$

Then, we will summarize some important inequalities relate to the projection operator P_Ω, the proofs of these inequalities can be found in Ref. [19].

Lemma 2.1 Let $\Omega \subset \mathbf{R}^n$ be a nonempty, closed and convex set. Let $P_\Omega(\cdot)$ be defined as above. Then we have

$$(u - P_\Omega(u))^T (P_\Omega(u) - w) \geqslant 0,\ \forall u \in \mathbf{R}^n,\ w \in \Omega,$$

$$\| P_\Omega(u) - P_\Omega(v) \| \leqslant \| u - v \|,\ \forall u, v \in \Omega,$$

$$\| P_\Omega(u) - w \|^2 \leqslant \| u - w \|^2 - \| u - P_\Omega(u) \|^2,\ \forall u \in \mathbf{R}^n,\ \forall w \in \Omega.$$

Throughout, we make the following assumptions:

Assumption 1 It is easy to compute the projection onto the sets X and Y under the Euclidean norm;

Assumption 2 The mapping $f(x)$ and $g(y)$ are Lipschitz continuous on X and Y, respectively;

Assumption 3 The solution set of (2)-(3) is nonempty.

3 Method for solving (2)-(3)

For analysis convenience, we denote

$$M = \begin{bmatrix} s_k I + \beta I & I \\ \cdots\cdots\cdots & \\ I & \frac{1}{\beta} I \end{bmatrix},$$

$$M_1 = \begin{bmatrix} (1-\theta) s_k I + \beta I & I \\ I & \frac{1}{\beta} I \end{bmatrix},$$

where $\beta > 0, \theta \in (0,1)$. For a positive definite matrix $H \in \mathbf{R}^{m \times n}$, the operators $\lambda_m(H)$ denote the smallest eigenvalue of H, Let $\lambda_m(M_1) = \delta_1$.

Algorithm 3.1

Step 0. Let $\varepsilon > 0, w^0 = (x^0, y^0, \lambda^0) \in \mathbf{R}^{n_1} \times \mathbf{R}^{n_2} \times \mathbf{R}^{n_3}, \beta > 0, H = \beta I, \theta \in (0,1), \gamma \in [1,2)$, set $k = 0$.

Step 1. Find $\tilde{x}^k$ such that

$$\tilde{x}^k = P_X\{ \tilde{x}^k - \beta [f(\tilde{x}^k) - A^T (\lambda^k - H(A\tilde{x}^k - y^k))]\} \quad (5)$$

Step 2. Update $\tilde{\lambda}^k$ via

$$\tilde{\lambda}^k = \lambda^k - H(A\tilde{x}^k - y^k) \quad (6)$$

Step 3. Find $\tilde{y}^k$ such that

$$\tilde{y}^k = P_Y\{ y^k - \frac{1}{s_k} [g(y^k) + (\tilde{\lambda}^k - H(A\tilde{x}^k - y^k))]\} \quad (7)$$

where s_k is a proper parameter which satisfies

$$\| \xi_y^k \| \leqslant \theta s_k \| y^k - \tilde{y}^k \|,$$

$$\xi_y^k = g(y^k) - g(\tilde{y}^k) + \beta (y^k - \tilde{y}^k).$$

Step 4. Convergence verification: If

$$\max\{ \| y^k - \tilde{y}^k \|_\infty, \| \lambda^k - \tilde{\lambda}^k \|_\infty \} < \varepsilon,$$

then stop.

Step 5. The new iterate is produced by

$$v^{k+1} = v^k - \alpha d(v^k, \tilde{v}^k, \xi^k),$$

where

$$d(v^k, \tilde{v}^k, \xi^k) = M(v^k - \tilde{v}^k) - \xi^k,$$

$$\xi^k = (\xi_y^k, 0)^T,$$

and

$$\alpha = \gamma \alpha^*,\ \alpha^* = \frac{\varphi(v^k, \tilde{v}^k, \xi^k)}{\| d(v^k, \tilde{v}^k, \xi^k) \|^2},$$

$$\varphi(v^k, \tilde{v}^k, \xi^k) = \| v^k - \tilde{v}^k \|_M^2 - (v^k - \tilde{v}^k)^T \xi^K \quad (8)$$

4 Convergence

Lemma 4.1 Let $\varphi(v^k, \tilde{v}^k, \xi^k)$ be defined in (8), $\lambda_m(M_1) = \delta_1$. then for any k, we have

$$\varphi(v^k, \tilde{v}^k, \xi^k) \geqslant \delta_1 \| v^k - \tilde{v}^k \|^2.$$

Proof By using the Cauchy-Schwarz inequality and the definition of ξ^k, we have

$$-(v^k - \tilde{v}^k)^T \xi^k \geqslant -\theta s_k \| y^k - \tilde{y}^k \|^2.$$

By the definition of $\varphi(v^k, \tilde{v}^k, \xi^k)$, it is easy to get that

$$\varphi(v^k, \tilde{v}^k, \xi^k) \geqslant \| v^k - \tilde{v}^k \|_{M_1}^2 \geqslant \delta_1 \| v^k - \tilde{v}^k \|^2.$$

Lemma 4.2 Let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ be generated by Algorithm 3.1 from a given $v^k = (y^k, \lambda^k)$. Then for any $v^* = (y^*, \lambda^*) \in V^*$, we have

$$(v^k - v^*)^T d(v^k, \tilde{v}^k, \xi^k) \geqslant \varphi(v^k, \tilde{v}^k, \xi^k).$$

Proof The inequality (5)~(7) are equivalent to

$$(x - \tilde{x}^k)^T (f(\tilde{x}^k) - A^T \tilde{\lambda}^k) \geqslant 0 \quad (9)$$

$$(\lambda - \tilde{\lambda}^k)^T ((A\tilde{x}^k - \tilde{y}^k) + (\tilde{y}^k - y^k) + \beta^{-1} (\tilde{\lambda}^k - \lambda^k)) \geqslant 0 \quad (10)$$

$$(y - \tilde{y}^k)^T (g(\tilde{y}^k) + \tilde{\lambda}^k + (\beta + s_k)(\tilde{y}^k - y^k) +$$

$$(\tilde{\lambda}^k-\lambda^k)+\xi_y^k\geqslant 0 \qquad (11)$$

respectively.

Let $w^*=(x^*,y^*,\lambda^*)$ be the solution of (2)-(3) and $\tilde{x}^k\in X,\tilde{y}^k\in Y$, we have

$$(\tilde{x}^k-x^*)^T(f(x^*)-A^T\lambda^*)\geqslant 0 \qquad (12)$$

$$(\tilde{y}^k-y^*)^T(g(y^*)+\lambda^*)\geqslant 0 \qquad (13)$$

Adding (9) and (12) and using the monotone of $f(x)$, we have

$$(\tilde{\lambda}^k-\lambda^*)^TA(\tilde{x}^k-x^*)\geqslant 0 \qquad (14)$$

Adding (11) and (13) and using the monotone of $g(y)$, it follows that

$$(\tilde{y}^k-y^*)^T((\beta+s_k)(y^k-\tilde{y}^k)+(\lambda^k-\tilde{\lambda}^k)-\xi_y^k)-(\tilde{\lambda}^k-\lambda^*)^T(\tilde{y}^k-y^*)\geqslant 0 \qquad (15)$$

(10) is equivalent to

$$(\tilde{\lambda}^k-\lambda^*)^T(-(A\tilde{x}^k-\tilde{y}^k)+(y^k-\tilde{y}^k)+\beta^{-1}(\lambda^k-\tilde{\lambda}^k))\geqslant 0 \qquad (16)$$

Combining (14)~(16), and using $Ax^*-y^*=0$, we get

$$(\tilde{v}^k-v^*)^T(M(v^k-\tilde{v}^k)-\xi^k)\geqslant 0 \qquad (17)$$

By the definition of $d(v^k,\tilde{v}^k,\xi^k)$ and $\varphi(v^k,\tilde{v}^k,\xi^k)$, we get

$$(v^k-v^*)^Td(v^k,\tilde{v}^k,\xi^k)\geqslant\varphi(v^k,\tilde{v}^k,\xi^k).$$

For analysis convenience, we denote

$$\theta(\alpha)=\|v^k-v^*\|^2-\|v^{k+1}(\alpha)-v^*\|^2.$$

Lemma 4.3 Let $\tilde{w}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{\lambda}^k)$ be generated by Algorithm 3.1 from a given $v^k=(y^k,\lambda^k)$. Then for any $v^*=(y^*,\lambda^*)\in V^*$, we have

$$\theta(\alpha)\geqslant 2\alpha\varphi(v^k,\tilde{v}^k,\xi^k)-\alpha^2\|d(v^k,\tilde{v}^k,\xi^k)\|_M^2.$$

Proof Since

$$\theta(\alpha)=\|v^k-v^*\|^2-\|v^{k+1}(\alpha)-v^*\|^2=\|v^k-v^*\|^2-\|v^k-v^*-\alpha d(v^k,\tilde{v}^k,\xi^k)\|^2=2\alpha(v^k-v^*)^Td(v^k,\tilde{v}^k,\xi^k)-\alpha^2\|d(v^k,\tilde{v}^k,\xi^k)\|^2\geqslant 2\alpha\varphi(v^k,\tilde{v}^k,\xi^k)-\alpha^2\|d(v^k,\tilde{v}^k,\xi^k)\|^2,$$

We can choose the value of $\alpha=\alpha^*$ to maximize the lower bound of the $\theta(\alpha)$.

Theorem 4.4 Let $\tilde{w}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{\lambda}^k)$ be generated by Algorithm 3.1 from a given $v^k=(y^k,\lambda^k)$. Then for any $v^*=(y^*,\lambda^*)\in V^*$, we have

$$\|v^{k+1}-v^*\|^2\leqslant\|v^k-v^*\|^2-\gamma(2-\gamma)\alpha^*\delta_1\|v^k-\tilde{v}^k\|^2.$$

Proof we have

$$\|v^{k+1}-v^*\|^2\leqslant\|v^k-v^*\|^2-(2\gamma\alpha^*\varphi(v^k,\tilde{v}^k,\xi^k)-\gamma^2(\alpha^*)^2\|d(v^k,\tilde{v}^k,\xi^k)\|^2)\leqslant\|v^k-v^*\|^2-\gamma(2-\gamma)\alpha^*\varphi(v^k,\tilde{v}^k,\xi^{kk})\leqslant\|v^k-v^*\|^2-\gamma(2-\gamma)\alpha^*\delta_1\|v^k-\tilde{v}^k\|^2.$$

We see that v^k is Feje′r monotone. After a similar proof procedure as in Ref. [11], we can derive that the sequence $\{w^k\}$ generated by the proposed method converges to a solution of problem (2)-(3). The proof is end.

5 Numerical experiment

This section is devoted to test the efficiency of Algorithm 3.1 in comparison with IPSALM in Ref. [11] and APBPCM in Ref. [12]. We set $\varepsilon=10^{-5}$, $s_k=s=2.4$, $\beta=1$, $\gamma=1.3$.

We consider the following problem:

$$\min\{\frac{1}{2}\|X-C\|_F^2\,|\,X\in S_+^n\cap S_B\},$$

where

$$S_+^n=\{H\in\mathbf{R}^{n\times n}\,|\,H^T=H,H\geqslant 0\},$$

and

$$S_B=\{H\in H\in\mathbf{R}^{n\times n}\,|\,H^T=H,H_L\leqslant H\leqslant H_U\}.$$

Tab. 1 Numerical comparison of IPSALM, APBPCM and Algorithm 3.1

n	IPSALM		APBPCM		Algorithm 3.1	
	No. of iteration	Time (s)	No. of iteration	Time (s)	No. of iteration	Time (s)
100	81	0.81	74	0.61	59	0.35
200	105	5.49	109	4.79	73	2.55
300	124	17.28	121	13.64	79	8.02
400	133	43.02	136	34.26	95	20.43
500	148	89.41	151	72.20	97	38.68
800	192	412.25	173	311.12	111	179.72

6 Conclusions

In this paper, we take a fresh look at the alternating direction method of multipliers and propose a new partial inexact proximal point method based on the alternating direction method for monotone variational inequalities with separable structures. The proposed method handles one subproblem by an inexact proximal point method and solves the other one directly. The efficiency of the method is also verified by some numerical experiments.

References:

[1] Chen S, Saunders M A, Donoho D L. Atomic decomposition by basis pursuit [J]. Siam Rev, 2001, 43: 33.

[2] Candes E J, Tao T. Decoding by linear programming [J]. IEEE T Inf Theory, 2005, 51: 4203.

[3] Yang J, Zhang Y, Yin W. An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise [J]. Siam J Sci Comput, 2009, 31: 2842.

[4] Cai J F, Candes E J. A singular value thresholding algorithm for matrix completion [J]. Siam J Optimiz, 2008, 20: 1956.

[5] Boyd S, Vandenberghe L. Convex optimization [M]. Cambridge: Cambridge University Press, 2004.

[6] Glowinski R, Marrocco A. Sur l'approximation par éléments finis d'ordre un et la résolution par pénalisation-dualité d'une classe de problémes de Dirichlet non linéaires [J]. Revue Fr Autom Inform Rech Opér Anal Numér 1975, 2: 41.

[7] He B S, Liao L Z, Han D R. A new inexact alternating directionsmethod for monotone variational inequalities [J]. Math Program, 2002,92: 103.

[8] Ye C H, Yuan X M. A descent method for structured monotone variational inequalities [J]. Optim Method Softw, 2007, 22: 329.

[9] He B S, Tao M, Yuan X M. Alternating direction method with Gaussian back substitution for Separable convex programming [J]. SIAM [J] Optimiz, 2012, 22: 313.

[10] He B S. Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities [J]. Comput Optim Appl, 2009, 42: 195.

[11] Cai X J, Gu G Y, He B S. A proximal point algorithm revisit on the alternating direction method of multipliers [J]. Sci China Math, 2013, 56: 2179.

[12] He B S, Liao L Z, Yuan X M. Alternating projection based prediction-correction methods for structuredvariational inequalities [J]. Comput Math, 2006, 24: 693.

[13] Tao M, Yuan X M. An inexact parallel splitting augmented Lagrangian method for monotone variational inequalities with separable structures [J]. Comput Optim Appl, 2012, 52: 439.

[14] Chen Z M, Wan L, Yang Q Z. An inexact direction method for structured variational inequalities [J]. J Optimiz Theory Appl, 2014, 163: 439.

[15] Gu G Y, He B S, Yang J F. Inexact alternating direction-based contraction methods for separable linearly constrained convex optimization [J]. J Optimiz Theory Appl, 2013, 163: 105.

[16] Peng Z, Zhu W X. A partial inexact alternating direction method for structured variational inequalities [J]. Optimization, 2014, 63: 1043.

[17] Peng Z, Wu D H. An inexact proximal alternating directions method for structured variational inequalities [J]. Lect Notes Eng Comput Sci, 2010, 1: 1.

[18] Li H, Kou X P. A new inexact parallel splitting method for monotonevariational inequalities with separable structures [J]. J Sichuan Univ: Nat Sci Ed(四川大学学报:自然科学版), 2016, 53: 503.

[19] Bersekas D P, Tsitsiklis J N. Parallel and distributed computation [M]. Englewood Cliffs: Prentice-Hall, 1997.

引用本文格式:

中 文:陈小彪,李耿华,张玫玉. 一种部分非精确求解可分离凸优化问题的渐近点算法 [J]. 四川大学学报:自然科学版, 2019, 56: 8.

英 文:Chen X B, Li G H, Zhang M Y. A partial inexact proximal point method for separable convex programming [J]. J Sichuan Univ: Nat Sci Ed, 2019, 56: 8.