SCIENCE CHINA
Mathematics

• ARTICLES •

October 2013 Vol. 56 No. 10: 2179–2186
doi: 10.1007/s11425-013-4683-0

A proximal point algorithm revisit on the alternating direction method of multipliers

CAI XingJu[1], GU GuoYong[1], HE BingSheng[1] & YUAN XiaoMing[2,*]

[1]*Department of Mathematics, Nanjing University, Nanjing* 210093*, China;*
[2]*Department of Mathematics, Hong Kong Baptist University, Hong Kong, China*
Email: caixingju@njnu.edu.cn, ggu@nju.edu.cn, hebma@nju.edu.cn, xmyuan@hkbu.edu.hk

Received September 16, 2012; accepted January 19, 2013; published online August 26, 2013

Abstract The alternating direction method of multipliers (ADMM) is a benchmark for solving convex programming problems with separable objective functions and linear constraints. In the literature it has been illustrated as an application of the proximal point algorithm (PPA) to the dual problem of the model under consideration. This paper shows that ADMM can also be regarded as an application of PPA to the primal model with a customized choice of the proximal parameter. This primal illustration of ADMM is thus complemental to its dual illustration in the literature. This PPA revisit on ADMM from the primal perspective also enables us to recover the generalized ADMM proposed by Eckstein and Bertsekas easily. A worst-case $O(1/t)$ convergence rate in ergodic sense is established for a slight extension of Eckstein and Bertsekas's generalized ADMM.

Keywords alternating direction method of multipliers, convergence rate, convex programming, proximal point algorithm

MSC(2010) 65K10, 90C25, 90C33

Citation: Cai X J, Gu G Y, He B S, et al. A proximal point algorithm revisit on the alternating direction method of multipliers. Sci China Math, 2013, 56: 2179–2186, doi: 10.1007/s11425-013-4683-0

1 Introduction

The augmented Lagrangian method (ALM) in [13,16] and the proximal point algorithm (PPA) in [15,18] are two fundamental methods in optimization literature, and they are closely related. Consider a convex minimization problem with linear constraints (the primal problem):

$$\min\{\theta(u) \mid Cu = b,\ u \in \mathcal{U}\}, \tag{1.1}$$

where $\theta(u) : \mathbb{R}^n \to \mathbb{R}$ is a closed convex function (not necessarily smooth), $C \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $\mathcal{U} \subseteq \mathbb{R}^n$ is a closed convex set. Then, as analyzed in [17], the ALM for solving (1.1) is essentially an application of PPA to the dual problem of (1.1).

The primal-dual relationship between ALM and PPA for (1.1) takes a splitting version of illustration when a separable form of (1.1) is considered. More specifically, we consider

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}, \tag{1.2}$$

where $\theta_1(x) : \mathbb{R}^{n_1} \to \mathbb{R}$ and $\theta_2(y) : \mathbb{R}^{n_2} \to \mathbb{R}$ are closed convex functions (not necessarily smooth), $A \in \mathbb{R}^{m\times n_1}$, $B \in \mathbb{R}^{m\times n_2}$, $b \in \mathbb{R}^m$, $\mathcal{X} \subseteq \mathbb{R}^{n_1}$ and $\mathcal{Y} \subseteq \mathbb{R}^{n_2}$ are closed convex sets; and $n_1 + n_2 = n$.

*Corresponding author

The solution set of (1.2) is assumed to be nonempty in our discussion. Then, it was elaborated in [5] that when the Douglas-Rachford splitting method in [4, 14] which is actually a special splitting form of PPA (see [5]) is applied to the dual problem of (1.2), the alternating direction method of multipliers (ADMM) in [8, 9] which is a splitting version of the ALM is recovered. Hence, for either the generic convex minimization model (1.1) or the separable form (1.2), PPA plays a crucial role in unifying several benchmark algorithms. On the other hand, we notice that PPA's central role has been highlighted mainly in the dual context of convex programming problems, while discussion on its direct application to the primal model (1.1) or (1.2) is less in the literature, see e.g., [7, 12].

Recently, ADMM becomes very popular and it has found many efficient applications in a variety of areas, see e.g., [3] and references cited therein. The main purpose of this paper is to show that ADMM can be regarded as an application of PPA to the primal problem (1.2) with a customized choice of the proximal parameter in metric form. This primal illustration is thus complemental to ADMM's dual illustration established insightfully in [5]. The primal illustration differs from the dual explanation in [5] in two aspects. First, our primal illustration simply focuses on customized choices of the proximal parameter when PPA is applied to solve the primal model (1.2), while the dual explanation in [5] treats the dual problem of (1.2) from operator splitting perspectives. Second, our analysis is based on a variational reformulation of (1.2), while the analysis in [5] is based on a fixed point reformulation of (1.2). Based on the primal illustration, the generalized ADMM proposed in [5] can also be recovered easily. Then, inspired by the convergence rate of the original ADMM established recently in [11], we establish a similar worst-case convergence rate for a slight extension of the generalized ADMM in [5]. This is a theoretical result absent in [5].

The rest of the paper is organized as follows. In Section 2, we review briefly some useful preliminaries. In Section 3, we discuss the primal application of PPA to (1.2) and elucidate how to choose a customized proximal parameter to recover ADMM. Eckstein and Bertsekas's generalized ADMM in [5] is also derived in primal context. Then, we establish a worst-case $O(1/t)$ convergence rate for a slight extension of Eckstein and Bertsekas's generalized ADMM in Section 4. Finally, some concluding remarks are given in Section 5.

2 Preliminaries

In this section, we review some preliminaries which are useful for further discussions. More specifically, we recall a variational reformulation of (1.2), the relaxed PPA in [10] for solving the variational reformulation, and the ADMM in [8, 9] for solving (1.2).

We first show that the model (1.2) can be characterized by a variational reformulation. Let $\lambda \in \mathbb{R}^m$ be the Lagrange multiplier associated with the linear constraint in (1.2). Then the first-order optimality condition of (1.2) is to find $w^* = (x^*, y^*, \lambda^*) \in \Omega$ such that the following inequalities

$$\begin{cases} \theta_1(x) - \theta_1(x^*) + (x - x^*)^{\mathrm{T}}(-A^{\mathrm{T}}\lambda^*) \geqslant 0, \\ \theta_2(y) - \theta_2(y^*) + (y - y^*)^{\mathrm{T}}(-B^{\mathrm{T}}\lambda^*) \geqslant 0, \qquad \forall (x, y, \lambda) \in \Omega, \\ (\lambda - \lambda^*)^{\mathrm{T}}(Ax^* + By^* - b) \geqslant 0, \end{cases} \tag{2.1}$$

are satisfied. Moreover, for $w = (x, y, \lambda) \in \Omega$, let $u = (x, y)$ be the sub-vector of w and let

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad F(w) = \begin{pmatrix} -A^{\mathrm{T}}\lambda \\ -B^{\mathrm{T}}\lambda \\ Ax + By - b \end{pmatrix}. \tag{2.2}$$

Then (2.1) can be rewritten as finding $w^* \in \Omega$ such that

$$\mathrm{VI}(\Omega, \theta, F) \qquad \theta(u) - \theta(u^*) + (w - w^*)^{\mathrm{T}} F(w^*) \geqslant 0, \quad \forall w \in \Omega. \tag{2.3}$$

As we shall show, with the variational reformulation (2.1) or (2.3), it is easier to expose our PPA revisit on ADMM. In fact, the variational reformulation (2.1) or (2.3) only serves as a theoretical tool in the coming analysis. Furthermore, we denote by Ω^* the set of such w^* that satisfies (2.1). Then, under the aforementioned nonempty assumption on the solution set of (1.2), Ω^* is also nonempty.

Now we review the application of the relaxed PPA in [10] to (2.3). Let $w^k \in \Omega$ and $G \in \mathbb{R}^{(n+m)\times(n+m)}$ be a symmetric positive definite matrix, then the original PPA in [15, 18] generates the new iterate w^{k+1} via solving the regularized variational problem: finding $w^{k+1} \in \Omega$, such that

$$\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^{\mathrm{T}}\{F(w^{k+1}) + G(w^{k+1} - w^k)\} \geqslant 0, \quad \forall w \in \Omega. \tag{2.4}$$

The relaxed PPA in [10], however, combines the PPA step (2.4) with a relaxation step. More precisely, letting the solution of (2.4) be denoted by $\tilde{w}^k$, then the relaxed PPA in [10] yields the new iterate via

$$w^{k+1} = w^k - \gamma(w^k - \tilde{w}^k), \tag{2.5}$$

where $\gamma \in (0, 2)$ is the relaxation factor. In particular, γ is called an under-relaxation when $\gamma \in (0, 1)$ or over-relaxation factor when $\gamma \in (1, 2)$; and the relaxed PPA (2.5) reduces to the original PPA (2.4) when $\gamma = 1$.

Finally, we recall ADMM proposed originally in [8, 9]. The iterative scheme of ADMM for solving (1.2) is

$$\begin{cases} x^{k+1} = \arg\min\left\{\theta_1(x) - (\lambda^k)^{\mathrm{T}}(Ax + By^k - b) + \dfrac{\beta}{2}\|Ax + By^k - b\|^2 \,\middle|\, x \in \mathcal{X}\right\}, \\ y^{k+1} = \arg\min\left\{\theta_2(y) - (\lambda^k)^{\mathrm{T}}(Ax^{k+1} + By - b) + \dfrac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \,\middle|\, y \in \mathcal{Y}\right\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{cases} \tag{2.6}$$

The ADMM scheme (2.6) is a splitting version of the ALM where the ALM subproblem is decomposed into two subproblems in Gauss-Seidel fashion at each iteration, and thus the variables x and y can be solved separably. Since the functions $\theta_1(x)$ and $\theta_2(y)$ often have specific properties for a particular application of (1.2), the decomposition treatment of ADMM makes it possible to exploit these particular properties separately. In fact, the decomposed subproblems in (2.6) are often simple enough to have closed-form solutions or can be easily solved up to high precisions.

3 A primal application of PPA to (1.2)

In this section, we elucidate how to derive the ADMM by applying PPA directly to the primal problem (1.2) with a customized proximal parameter.

First, notice that the PPA subproblem (2.4) is only of conceptual or theoretical interest, as it could be as difficult as the original variational problem (2.3) from the algorithmic implementation point of view. But, the specific choice of the proximal regularization matrix G enjoys favorable freedom and it enables us to choose its entries judiciously by considering the specific structure of (2.1). Here, we would emphasize that we do not require the positive definiteness of G. Instead, positive semi-definiteness of G is enough for our algorithmic design. With only a positive semi-definite regularization matrix, it sounds not rigorous to still call (2.4) PPA. But, by this slight abuse of name, we only try to make our notation easier.

Then, revisiting the iterative scheme (2.6), it is easy to observe that the variable x plays only an intermediate role and it is not involved in the execution of ADMM, e.g., see the elaboration in [3]. Therefore, the input for executing the iteration of (2.6) is only the sequence $\{(y^k, \lambda^k)\}$. These facts inspire us to generate x^{k+1} exactly as ADMM and then focus only on the proximal regularization on the variables (y, λ) in (2.4), i.e., the first n_1 rows of G are all zero. More concretely, x^{k+1} is generated via solving

$$x^{k+1} = \arg\min\left\{\theta_1(x) - (\lambda^k)^{\mathrm{T}}(Ax + By^k - b) + \frac{\beta}{2}\|Ax + By^k - b\|^2 \,\middle|\, x \in \mathcal{X}\right\}.$$

Accordingly, the above formula means that

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^{\mathrm{T}}\{-A^{\mathrm{T}}[\lambda^k - \beta(Ax^{k+1} + By^k - b)]\} \geqslant 0, \quad \forall x \in \mathcal{X}. \tag{3.1}$$

Moreover, recall the x-related portion in the variational reformulation (2.1). After the step (3.1), if we update the Lagrange multiplier via

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b), \tag{3.2}$$

then (3.1) can be written as

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^{\mathrm{T}}(-A^{\mathrm{T}}\lambda^{k+1}) \geqslant 0, \quad \forall x \in \mathcal{X},$$

which implies that the x-related portion in the first-order optimality condition (2.1) is satisfied. This suggests us to update the Lagrange multiplier via (3.2) right after the computation of x^{k+1} by (3.1). By casting the scheme (3.2) into the form (2.4), it is then clear how to choose the λ-corresponding part of the regularization matrix G. In fact, (3.2) can be rewritten as

$$(\lambda - \lambda^{k+1})^{\mathrm{T}}\left((Ax^{k+1} + By^{k+1} - b) - B(y^{k+1} - y^k) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\right) \geqslant 0, \quad \forall \lambda \in \mathbb{R}^m.$$

Therefore, the scheme (3.2) essentially implies that the last m rows of the regularization matrix G in (2.4) is taken as $(0, -B, \frac{1}{\beta}I_m)$, where I_m denotes the $m \times m$ identity matrix.

Now, the rest is to specify the middle n_2 rows for the regularization matrix G in (2.4), i.e., the regularization on the variable y. In fact, with the specified regularization on x and λ, it is natural to specify the middle n_2 rows of G as $(0, \beta B^{\mathrm{T}}B, -B^{\mathrm{T}})$ for the sake of ensuring the positive semi-definiteness of G. Then, with this choice, it follows from (2.4) that y^{k+1} should be generated by the following problem

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^{\mathrm{T}}(-B^{\mathrm{T}}\lambda^{k+1} + \beta B^{\mathrm{T}}B(y^{k+1} - y^k) - B^{\mathrm{T}}(\lambda^{k+1} - \lambda^k)) \geqslant 0, \quad \forall y \in \mathcal{Y}.$$

That is, y^{k+1} is generated via solving the following problem,

$$y^{k+1} = \arg\min\left\{\theta_2(y) - (\lambda^{k+1})^{\mathrm{T}}(Ax^{k+1} + By - b) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \,\middle|\, y \in \mathcal{Y}\right\}. \tag{3.3}$$

In summary, in (2.4) we propose to choose a customized proximal regularization matrix as

$$G = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta B^{\mathrm{T}}B & -B^{\mathrm{T}} \\ 0 & -B & \frac{1}{\beta}I_m \end{pmatrix} \tag{3.4}$$

and the resulting customized PPA for (1.2) has the following algorithmic framework:

$$\begin{cases} x^{k+1} = \arg\min\left\{\theta_1(x) - (\lambda^k)^{\mathrm{T}}(Ax + By^k - b) + \frac{\beta}{2}\|Ax + By^k - b\|^2 \,\middle|\, x \in \mathcal{X}\right\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \arg\min\left\{\theta_2(y) - (\lambda^{k+1})^{\mathrm{T}}(Ax^{k+1} + By - b) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \,\middle|\, y \in \mathcal{Y}\right\}. \end{cases} \tag{3.5}$$

Comparing with the ADMM scheme (2.6), the primal application of PPA (3.5) yields the new iterate in the order of $x \to \lambda \to y$. Thus, (3.5) is identical with (2.6) in cyclical sense.

As mentioned, the variable x is not involved in the execution of (2.6) or (3.5). The following notations regarding the variables y and λ will simplify the notation of our analysis:

$$v = (y, \lambda); \quad \mathcal{V} = \mathcal{Y} \times \mathbb{R}^m;$$

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

Since (3.5) is just an application of PPA, we can combine it with the relaxation step (2.5) and yield a relaxed ADMM scheme. The resulting algorithm is summarized in Algorithm 1 below. Note that this algorithm is slightly more general than the generalized ADMM proposed in [5] where only the special case $B = -I$ was discussed. More specifically, instead of $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ in (3.5), the output of (3.5) is labeled as $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ below. Since the intermediate variable x is not required by the iteration, the relaxation step (2.5) is only implemented for the variables y and λ. That is why we only relax v^k, instead of w^k, in the relaxation step (3.7).

Algorithm 1 (Generalized ADMM for (1.2))**.** Let $\beta > 0$ and $\gamma \in (0, 2)$. With the given iterate w^k, the new iterate w^{k+1} is generated as follows.

Step 1. Primal PPA step. Obtain $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ via

$$\begin{cases} \tilde{x}^k = \arg\min \left\{ \theta_1(x) - (\lambda^k)^{\mathrm{T}}(Ax + By^k - b) + \dfrac{\beta}{2}\|Ax + By^k - b\|^2 \,\middle|\, x \in \mathcal{X} \right\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \arg\min \left\{ \theta_2(y) - (\tilde{\lambda}^k)^{\mathrm{T}}(A\tilde{x}^k + By - b) + \dfrac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \,\middle|\, y \in \mathcal{Y} \right\}; \end{cases} \tag{3.6}$$

Step 2. Relaxation step.

$$v^{k+1} = v^k - \gamma(v^k - \tilde{v}^k). \tag{3.7}$$

Remark 3.1. As conjectured in [1,5], aggressive values of the relaxation factor γ are preferred empirically. We shall provide a reason to explain it in Theorem 4.3.

4 Convergence rate

The global convergence of Algorithm 1 can be established easily by following the standard analytic framework of contraction methods in [2]; we thus omit it. In this section, inspired by [11], we establish a worst-case $O(1/t)$ convergence rate in ergodic sense for Algorithm 1. That is, after t iterations of Algorithm 1, we can find an approximate solution of (1.2) with an accuracy of $O(1/t)$. The $O(1/t)$ convergence rate of the original PPA without relaxation has been shown in the literature; but our result is for the application of a relaxed PPA to (1.2).

First, we define

$$M = \begin{pmatrix} \beta B^{\mathrm{T}}B & -B^{\mathrm{T}} \\ -B & \frac{1}{\beta}I_m \end{pmatrix}, \tag{4.1}$$

which is a submatrix of G defined in (3.4) for its non-zero entries. Note M is only positive semi-definite without further assumption on the matrix B. For notational convenience, by a slight abuse, throughout we still use the notation

$$\|v^k - \tilde{v}^k\|_M := \sqrt{(v^k - \tilde{v}^k)^{\mathrm{T}} M (v^k - \tilde{v}^k)}.$$

Then, we recall a variational characterization to the solution set of (1.2). Since the proof can be found in [6,11], we omit the proof.

Theorem 4.1. *The solution set* Ω^* *of* $\mathrm{VI}(\Omega, \theta, F)$ (2.3) *is convex and it can be characterized as*

$$\Omega^* = \bigcap_{w \in \Omega} \{w^* \in \Omega \mid \theta(u) - \theta(u^*) + (w - w^*)^{\mathrm{T}} F(w) \geqslant 0\}. \tag{4.2}$$

Theorem 4.1 thus implies that in order to show a worst-case $O(1/t)$ convergence rate of Algorithm 1, we need to prove that after t iterations, we can find $\bar{w} \in \Omega$ such that

$$\theta(\bar{u}) - \theta(u) + (\bar{w} - w)^{\mathrm{T}} F(w) \leqslant \epsilon, \quad \forall w \in \Omega, \tag{4.3}$$

with $\epsilon = O(1/t)$. For this purpose, we prove a key inequality in the following lemma.

Lemma 4.2. *Let the sequences* $\{v^k\}$, $\{\tilde{v}^k\}$ *and* $\{\tilde{w}^k\}$ *be generated by Algorithm* 1 *and the matrix* M *be defined in* (4.1)*. Then, we have*

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^{\mathrm{T}}F(\tilde{w}^k) \geqslant \frac{1}{2\gamma}(\|v-v^{k+1}\|_M^2-\|v-v^k\|_M^2)+\left(1-\frac{\gamma}{2}\right)\|v^k-\tilde{v}^k\|_M^2, \quad \forall w\in\Omega.$$

Proof. First, using the notation M in (4.1) and taking the first-order optimality conditions of the subproblems in (3.6), we get

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^{\mathrm{T}}F(\tilde{w}^k)\geqslant (v-\tilde{v}^k)^{\mathrm{T}}M(v^k-\tilde{v}^k), \quad \forall w\in\Omega.$$

Since the relaxation step (3.7) can be rewritten as $(v^k-v^{k+1})=\gamma(v^k-\tilde{v}^k)$, we have

$$(v-\tilde{v}^k)^{\mathrm{T}}M(v^k-\tilde{v}^k)=\frac{1}{\gamma}(v-\tilde{v}^k)^{\mathrm{T}}M(v^k-v^{k+1}).$$

Thus, it suffices to show that

$$(v-\tilde{v}^k)^{\mathrm{T}}M(v^k-v^{k+1})=\frac{1}{2}(\|v-v^{k+1}\|_M^2-\|v-v^k\|_M^2)+\gamma\left(1-\frac{\gamma}{2}\right)\|v^k-\tilde{v}^k\|_M^2. \tag{4.4}$$

By setting $a=v,\ b=\tilde{v}^k,\ c=v^k$ and $d=v^{k+1}$ in the identity

$$(a-b)^{\mathrm{T}}M(c-d)=\frac{1}{2}(\|a-d\|_M^2-\|a-c\|_M^2)+\frac{1}{2}(\|c-b\|_M^2-\|d-b\|_M^2),$$

we derive that

$$(v-\tilde{v}^k)^{\mathrm{T}}M(v^k-v^{k+1})=\frac{1}{2}(\|v-v^{k+1}\|_M^2-\|v-v^k\|_M^2)+\frac{1}{2}(\|v^k-\tilde{v}^k\|_M^2-\|v^{k+1}-\tilde{v}^k\|_M^2).$$

On the other hand, we have

$$\begin{aligned}\|v^k-\tilde{v}^k\|_M^2-\|v^{k+1}-\tilde{v}^k\|_M^2&=\|v^k-\tilde{v}^k\|_M^2-\|(v^k-\tilde{v}^k)-(v^k-v^{k+1})\|_M^2\\&=\|v^k-\tilde{v}^k\|_M^2-\|(v^k-\tilde{v}^k)-\gamma(v^k-\tilde{v}^k)\|_M^2\\&=\gamma(2-\gamma)\|v^k-\tilde{v}^k\|_M^2.\end{aligned}$$

Combining the last two equations, we obtain (4.4). Hence, the lemma is proved. □

Now, we are ready to prove a worst-case $O(1/t)$ convergence rate in ergodic sense for Algorithm 1.

Theorem 4.3. *Let the sequences* $\{\tilde{w}^k\}$ *be generated by Algorithm* 1*. For an integer* $t>0$*, let*

$$\bar{w}_t:=\frac{1}{t+1}\sum_{k=0}^{t}\tilde{w}^k. \tag{4.5}$$

Then $\bar{w}_t\in\Omega$ *and*

$$\theta(\bar{u}_t)-\theta(u)+(\bar{w}_t-w)^{\mathrm{T}}F(w)\leqslant\frac{1}{2\gamma(t+1)}\|v-v^0\|_M^2, \quad \forall w\in\Omega. \tag{4.6}$$

Proof. First, for an integer $t>0$, we have $\tilde{w}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{\lambda}^k)\in\Omega$ for $k=0,1,\ldots,t$. Since $\frac{1}{t+1}\sum_{k=0}^{t}\tilde{w}^k$ can be viewed as a convex combination of $\tilde{w}^k$'s, we obtain $\bar{w}_t=\frac{1}{t+1}\sum_{k=0}^{t}\tilde{w}^k\in\Omega$.

Second, since $\gamma\in(0,2)$, it follows from Lemma 4.2 that

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^{\mathrm{T}}F(\tilde{w}^k)\geqslant\frac{1}{2\gamma}(\|v-v^{k+1}\|_M^2-\|v-v^k\|_M^2), \quad \forall w\in\Omega.$$

By combining the monotonicity of $F(\cdot)$ with the last inequality, we obtain

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^{\mathrm{T}} F(w) + \frac{1}{2\gamma}(\|v - v^k\|_M^2 - \|v - v^{k+1}\|_M^2) \geqslant 0, \quad \forall w \in \Omega.$$

Summing the above inequality over $k = 0, 1, \ldots, t$, we derive that

$$\left((t+1)\theta(u) - \sum_{k=0}^{t} \theta(\tilde{u}^k)\right) + \left((t+1)w - \sum_{k=0}^{t} \tilde{w}^k\right)^{\mathrm{T}} F(w) + \frac{1}{2\gamma}(\|v - v^0\|_M^2 - \|v - v^{t+1}\|_M^2) \geqslant 0, \quad \forall w \in \Omega.$$

By dropping the minus term, we have

$$\left((t+1)\theta(u) - \sum_{k=0}^{t} \theta(\tilde{u}^k)\right) + \left((t+1)w - \sum_{k=0}^{t} \tilde{w}^k\right)^{\mathrm{T}} F(w) + \frac{1}{2\gamma}\|v - v^0\|_M^2 \geqslant 0, \quad \forall w \in \Omega,$$

which is equivalent to

$$\left(\frac{1}{t+1}\sum_{k=0}^{t} \theta(\tilde{u}^k) - \theta(u)\right) + \left(\frac{1}{t+1}\sum_{k=0}^{t} \tilde{w}^k - w\right)^{\mathrm{T}} F(w) \leqslant \frac{1}{2\gamma(t+1)}\|v - v^0\|_M^2, \quad \forall w \in \Omega. \tag{4.7}$$

Since θ is convex and $\bar{u}_t := \frac{1}{t+1}\sum_{k=0}^{t} \tilde{u}^k$, we have

$$\theta(\bar{u}_t) \leqslant \frac{1}{t+1}\sum_{k=0}^{t} \theta(\tilde{u}^k).$$

Substituting it in (4.7) and using (4.5), we get (4.6) and the proof is completed. □

According to Theorem 4.3, for any given compact set $\mathcal{D} \subset \Omega$, let $d := \sup\{\|v - v^0\|_M^2 \mid w \in \mathcal{D}\}$. Then, after t iterations of Algorithm 1, the point $\bar{w}_t$ defined in (4.5) satisfies

$$\sup_{w \in \mathcal{D}} \{\theta(\bar{u}_t) - \theta(u) + (\bar{w}_t - w)^{\mathrm{T}} F(w)\} \leqslant \frac{d}{2\gamma(t+1)},$$

which means that $\bar{w}_t$ is an approximate solution of (2.1) with the accuracy $O(1/t)$. That is, a worst-case $O(1/t)$ convergence rate in ergodic sense is established for Algorithm 1.

Last, we would remark on the choice of the relaxation factor γ. According to (4.6), it is obvious that larger values of γ is more beneficial for accelerating the convergence of the proposed Algorithm 1. Considering the requirement $\gamma \in (0, 2)$, it is preferred to choose $\gamma \to 2$ empirically. Thus, we provide an intuitive understanding on the conjecture in [1] of choosing an over-relaxation factor in $(1, 2)$ empirically.

5 Conclusions

This paper takes a fresh look at the alternating direction method of multipliers (ADMM) for solving a convex minimization model with a separable objective function and linear constraints. It is shown that ADMM is a primal application of the proximal point algorithm (PPA) with a customized proximal parameter in metric form. We thus reach the conclusion that PPA can be equally effective for separable convex programming when it is applied to either the primal or the dual problem of the model under consideration. A worst-case $O(1/t)$ convergence rate in ergodic sense is established for a slight extension of the generalized ADMM proposed by Eckstein and Bertsekas. The PPA revisit provides a novel and simple way to understand ADMM. For future work, it is interesting to investigate the worst-case $O(1/t)$ convergence rate in non-ergodic sense for the generalized ADMM.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 11001124 and 91130007), the Doctoral Fund of Ministry of Eduction of China (Grant No. 20110091110004) and the General Research Fund from Hong Kong Research Grants Council (Grant No. HKBU 203712).

References

1 Bertsekas D P. Constrained Optimization and Lagrange Multiplier Methods. New York: Academic Press, 1982

2 Blum E, Oettli W. Mathematische Optimierung, Econometrics and Operations Research XX. Berlin: Springer-Verlag, 1975

3 Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn, 2010, 3: 1–122

4 Douglas J, Rachford H H. On the numerical solution of the heat conduction problem in 2 and 3 space variables. Trans Amer Math Soc, 1956, 82: 421–439

5 Eckstein J, Bertsekas D P. On the Douglas-Rachford splitting method and the proximal points algorithm for maximal monotone operators. Math Program, 1992, 55: 293–318

6 Facchinei F, Pang J S. Finite-Dimensional Variational Inequalities and Complementarity Problems, vol. I. Springer Ser Oper Res. Berlin: Springer-Verlag, 2003

7 Fukushima M. The primal Douglas-Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem. Math Program, 1996, 72: 1–15

8 Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput Math Appl, 1976, 2: 17–40

9 Glowinski R, Marrocco A. Sur l'approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problémes non linéaires. Anal Numer, 1975, 2: 41–76

10 Gol'shtein E G, Tret'yakov N V. Modified Lagrangian in convex programming and their generalizations. Math Program Study, 1979, 10: 86–97

11 He B S, Yuan X M. On the $O(1/n)$ convergence rate of Douglas-Rachford alternating direction method. SIAM J Numer Anal, 2012, 50: 700–709

12 He B S, Yuan X M, Zhang W X. A customized proximal point algorithm for convex minimization with linear constraints. Comput Optim Appl, submitted

13 Hestenes M R. Multiplier and gradient methods. J Optim Theory Appl, 1969, 4: 303–320

14 Lions P L, Mercier B. Splitting algorithms for the sum of two nonlinear operators. SIAM J Numer Anal, 1979, 16: 964–979

15 Martinet B. Regularision d'inéquations variationnelles par approximations successive. Revue Francaise d'Automatique et Informatique Recherche Opérationnelle, 1970, 126: 154–159

16 Powell M J D. A method for nonlinear constraints in minimization problems. In: Fletcher R, ed. Optimization. New York: Academic Press, 1969, 283–298

17 Rockafellar R T. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math Oper Res, 1976, 1: 97–116

18 Rockafellar R T. Monotone operators and the proximal point algorithm. SIAM J Control Optim, 1976, 14: 877–898

Information for authors

SCIENCE CHINA Mathematics, a peer-reviewed mathematical journal cosponsored by Chinese Academy of Sciences and National Natural Science Foundation of China, and published monthly in both print and electronic forms by Science China Press and Springer, is committed to publishing high-quality, original results in both basic and applied research.

Categories of articles:

Reviews summarize representative results and achievements in a particular topic or an area, comment on the current state of research, and advise on the research directions. The author's own opinion and related discussion are requested.

Articles report on important original results in all areas of mathematics.

Brief reports present short reports in a timely manner of the latest important results.

Authors are recommended to use the online submission services. To submit a manuscript, please visit http://mc03.manuscriptcentral.com/scmath, get an account, and follow the instructions to upload the text and image/table files.

Authors should also submit such accompanying materials as a short statement on the research background, area/subarea and significance of the work, a brief introduction to the first and corresponding authors including their mailing address, post code, telephone number, fax number, and email address. Authors may suggest several referees (please supply full names, addresses, phone, fax and email), and/or request the exclusion of specific reviewers.

All submissions will be reviewed by referees selected by the editorial board. The decision of acceptance or rejection of a manuscript is made by the editorial board based on the referees' reports. The entire review process may take 90 to 180 days, and the editorial office will inform the author of the decision as soon as the process is completed.

Authors should guarantee that their submitted manuscript has not been published before, and has not been submitted elsewhere for print or electronic publication consideration.

Submission of a manuscript is taken to imply that all the named authors are aware that they are listed as co-authors, and they have seen and agreed to the submitted version of the paper. No change in the order of listed authors can be made without an agreement signed by all the authors.

Once a manuscript is accepted, the authors should send a copyright transfer form signed by all authors to Science China Press. Authors of one published paper will be presented one sample copy. If offprints and more sample copies are required, please contact the managing editor and pay the extra fee. The full text in Chinese and in English opens freely to the readers in China at www.scichina.com, and the full text in English is available to overseas readers at www. springerlink.com.

Subscription information

ISSN print edition: 1674-7283
ISSN electronic edition: 1869-1862

Subscription rates:
For information on subscription rates please contact:
Customer Service
China: sales@scichina.org
North and South America:
journals-ny@springer.com
Outside North and South America:
subscriptions@springer.com

Orders and inquiries:
China
Science China Press
16 Donghuangchenggen North Street,
Beijing 100717, China
Tel: 86-10-64034559 or 86-10-64034134
Fax: 86-10-64016350
Email: sales@scichina.org
North and South America
Springer New York, Inc.
Journal Fulfillment
P.O. Box 2485
Secaucus, NJ 07096, USA
Tel: 1-800-SPRINGER or 1-201-348-4033
Fax: 1-201-348-4505
Email: journals-ny@springer.com
Outside North and South America
Springer Distribution Center
Customer Service Journals
Haberstr. 7, 69126
Heidelberg, Germany
Tel: 49-6221-345-0, Fax: 49-6221-345-4229
Email: subscriptions@springer.com

Cancellations must be received by September 30 to take effect at the end of the same year.

Changes of address: Allow for six weeks for all changes to become effective. All communications should include both old and new addresses (with postal codes) and should be accompanied by a mailing label from a recent issue. According to § 4 Sect. 3 of the German Postal Services Data Protection Regulations, if a subscriber's address changes, the German Federal Post Office can inform the publisher of the new address even if the subscriber has not submitted a formal application for mail to be forwarded. Subscribers not in agreement with this procedure may send a written complaint to Customer Service Journals, Karin Tiks, within 14 days of publication of this issue.

Microform editions are available from: ProQuest. Further information available at http://www.il.proquest.com/uni.

Electronic edition:
An electronic version is available at springerlink.com.

Production:
Science China Press
16 Donghuangchenggen North Street, Beijing 100717, China
Tel: 86-10-64034559 or 86-10-64034134
Fax: 86-10-64016350
Printed in the People's Republic of China
Jointly Published by
Science China Press and Springer

Go To Website

SCIENCE CHINA Mathematics

CONTENTS

Vol. 56 No. 10 October 2013

Progress of Projects Supported by NSFC

Articles

math.scichina.com

www.springer.com/scp

Indexed by:
SCI-CD
MR
Z Math
MathSciNet

ISSN 1674-7283

9 771674 728132 10