J Glob Optim
DOI 10.1007/s10898-015-0287-2

A proximal alternating direction method of multipliers for a minimization problem with nonconvex constraints

Zheng Peng · Jianli Chen · Wenxing Zhu

Received: 28 December 2013 / Accepted: 3 March 2015

Abstract In this paper, a proximal alternating direction method of multipliers is proposed for solving a minimization problem with Lipschitz nonconvex constraints. Such problems are raised in many engineering fields, such as the analytical global placement of very large scale integrated circuit design. The proposed method is essentially a new application of the classical proximal alternating direction method of multipliers. We prove that, under some suitable conditions, any subsequence of the sequence generated by the proposed method globally converges to a Karush–Kuhn–Tucker point of the problem. We also present a practical implementation of the method using a certain self-adaptive rule of the proximal parameters. The proposed method is used as a global placement method in a placer of very large scale integrated circuit design. Preliminary numerical results indicate that, compared with some state-of-the-art global placement methods, the proposed method is applicable and efficient.

Keywords Alternating direction method of multipliers · Proximal point method · Nonconvex minimization · VLSI · Global convergence

1 Introduction

In the design of very large scale integrated circuit (VLSI for short), placement is the process of determining the locations of circuit devices on a die surface. Placement stage can be divided into three steps: global placement, legalization and detailed placement. Global placement is the first and the most important step, which generates a rough placement solution maintaining a global view of the whole netlist. The quality of the final placement result depends heavily on

This work was supported by the Natural Science Foundation of China (61170308), Major Science Foundation of FuJian Province Department of Education (JA14037), and Talent Foundation of Fuzhou University (XRC-1043).

Z. Peng (✉) · J. Chen · W. Zhu
College of Mathematics and Computer Science, Fuzhou University,
Fuzhou 350108, People's Republic of China
e-mail: pzheng@fzu.edu.cn

Published online: 10 March 2015

this step. Many researchers have presented various models of global placement, see Chu [11]. All the proposed models try to minimize the wire-length subject to no cell overlaps. Among these models, the density control half-perimeter wire length (HPWL) model is common in analytical placement, see Chang et al. [12].

Suppose that there are n cells to be placed in a chip area, and the chip area is divided into T bins. In view of graph theory, a circuit can be formulated as a graph $G(V, E)$, where $V = \{v_1, v_2, \ldots, v_i, \ldots, v_n\}$ is the vertex set and v_i labels the cell i with center position (x_i, y_i); $E = \{e_j, j = 1, 2, \ldots, m\}$ is the edge set, where each edge is an unordered pair of vertices which labels a net of the integrated circuits.

The density control HPWL model is a constrained minimization problem of the form:

$$\begin{cases} \min \ W(V, E) = \sum\limits_{e \in E} \left\{ \max\limits_{v_i, v_j \in e} \left| x_i - x_j \right| + \max\limits_{v_i, v_j \in e} \left| y_i - y_j \right| \right\}, \\ \text{s.t. } D_b(x, y) = M_b, \quad \forall b = 1, 2, \ldots, T. \end{cases} \tag{1.1}$$

The objective function $W(V, E)$ is known as HPWL, and the constraint $D_b(x, y) = M_b$ is used to control the density of the cells placed in bin b.

For simplicity, we assume that one edge $e(\in E)$ connects two cells, v_i and v_j. In this case, the model (1.1) is referred to as a regular graph model, and the wire-length $W(V, E)$ is reduced to

$$W(V, E) = \sum_{v_i, v_j \in e, e \in E} \left\{ \left| x_i - x_j \right| + \left| y_i - y_j \right| \right\}, \tag{1.2}$$

which can be rewritten to the following compact form

$$W(V, E) = \|Ax\|_1 + \|Ay\|_1, \tag{1.3}$$

where A is determined by the given netlist describing the connected information of all cells to be placed. The matrix A has the form

$$A = \begin{pmatrix} 1 & \cdots & -1 & \cdots \\ \cdots & 1 & \cdots & -1 \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}_{m \times n},$$

where m is the number of edges and n is the number of vertices. For example, if a graph consists of 5 vertices $\{v_1, \ldots, v_5\}$ and 7 edges $\{v_1v_2, v_1v_3, v_1v_4, v_2v_4, v_2v_5, v_3v_4, v_4v_5\}$, then

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}_{7 \times 5}.$$

The density function $D_b(x, y)$ has the form

$$D_b(x, y) = \sum_{v \in V} P_{b,v}(x) Q_{b,v}(y) \tag{1.4}$$

where $P_{b,v}(x)$ and $Q_{b,v}(y)$ are piecewise-linear functions.

Based on this observation, the density control HPWL (regular graph) model can be rewritten as the following form:

$$\begin{cases} \min\limits_{x,y} \ \|Ax\|_1 + \|Ay\|_1 \\ \text{s.t.} \quad D_b(x, y) - M_b = 0, \quad b = 1, 2, \ldots, T. \end{cases} \tag{1.5}$$

In the minimization problem (1.5), the objective function is the sum of two convex functions, the constraints consist of T equations and each function in the equations is nonconvex and Lipschitzian.

In this paper, we consider a generalization of the problem (1.5), i.e.,

$$\begin{cases} \min \ \theta_1(x) + \theta_2(y) \\ \text{s.t.} \quad \Phi(x, y) - b = 0, \ x \in R^{n_1}, y \in R^{n_2}, \end{cases} \tag{1.6}$$

where $\theta_1 : R^{n_1} \to R$ and $\theta_2 : R^{n_2} \to R$ are closed convex, and $\Phi : R^{n_1} \times R^{n_2} \to R^m$ is a vector-valued function whose components are bounded and Lipschitzian.

The goal of this paper is to propose a proximal alternating direction method of multipliers (proximal ADMM for short) for the problem (1.6), which is a minimization problem with Lipschitz nonconvex constraints. This is a new application of ADMM combining the proximal point algorithm (PPA). The "new application" means in the sense that, the proposed proximal ADMM is used to solve a nonconvex minimization problem, while the PPA and ADMM are commonly used in convex optimization. The PPA was originally proposed by Martinet [21,22] and extended by Rockafellar [24], Bertsekas and Tseng [6], and Kaplan and Tichatschke [20] , etc. The proximal ADMMs are state-of-the-art algorithms for structured convex optimization problems, see Eckstein and Bertsekas [13], Eckstein [14], and Eckstein and Fukushima [15], etc. He et al. [17] and Yuan [29] presented some effective proximal ADMMs for variational inequalities. More recently, Chen, He and Yuan [9] proposed an ADMM for the matrix completion; Yang et al. [28] proposed a PPA for log-determinant optimization with group Lasso regularization; and so on.

In general, a convex minimization problem with separable structure is of the form

$$\begin{cases} \min \ \theta_1(x) + \theta_2(y), \\ \ \text{s.t.} \ \ Ax + By = b, \\ \qquad\ x \in X \ , y \in Y, \end{cases} \tag{1.7}$$

where $X \subset R^{n_1}$ and $Y \subset R^{n_2}$ are closed convex subsets; $\theta_1 : X \to R$ and $\theta_2 : Y \to R$ are convex functions; $A \in R^{m \times n_1}$, $B \in R^{m \times n_2}$ and $b \in R^m$. The augmented Lagrangian function of the problem (1.7) is

$$L_\rho(x, y, \lambda) = \theta_1(x) + \theta_2(y) + \lambda^T(Ax + By - b) + \frac{\rho}{2}\|Ax + By - b\|^2. \tag{1.8}$$

Given (x^k, y^k, λ^k), the proximal ADMM for (1.7) produces the next iterate $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ via the following scheme:

$$x^{k+1} = \text{Arg}\min_{x \in X} \left\{ L_\rho(x, y^k, \lambda^k) + \frac{r_k}{2}\|x - x^k\|^2 \right\}, \tag{1.9a}$$

$$y^{k+1} = \text{Arg}\min_{y \in Y} \left\{ L_\rho(x^{k+1}, y, \lambda^k) + \frac{s_k}{2}\|y - y^k\|^2 \right\}, \tag{1.9b}$$

$$\lambda^{k+1} = \lambda^k + \gamma\rho(Ax^{k+1} + By^{k+1} - b). \tag{1.9c}$$

The relaxation factor $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ was first given by Glowinski [16] in the classical alternating directions method, which was showed by Xu [27] that it is valid in the proximal ADMM.

The major difference between (1.7) and (1.6) is that, the constraint of the former one is linear and consequently the problem is a convex minimization problem, while the constraint of the later one is nonconvex. The nonconvexity leads some difficulties in designing an efficient algorithm.

The rest of this paper is organized as follows. The next section describes the proposed proximal ADMM and its properties. Sect. 3 gives some preliminaries for convergence of the proximal ADMM. Section 4 shows the global convergence to a KKT point of the proposed method under some suitable conditions. Section 5 suggests a self-adaptive style for choosing proximal parameters to guarantee the subproblems satisfying the convergence conditions, and presents some preliminary numerical results on the VLSI problem to indicate the validity and efficiency of the proposed proximal ADMM. Finally, Sect. 6 concludes the paper with some final remarks.

2 The proposed method

The augmented Lagrangian function associated to the problem (1.6) is:

$$L_\rho(x, y, \lambda) = \theta_1(x) + \theta_2(y) + \lambda^T(\Phi(x, y) - b) + \frac{\rho}{2}\|\Phi(x, y) - b\|^2. \tag{2.1}$$

Given a triple (x^k, y^k, λ^k), the proposed proximal ADMM for (1.6) produces the new iterate $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ via the following scheme:

$$x^{k+1} = \operatorname{Arg}\min_{x\in R^{n_1}} \left\{L_{\rho_k}(x, y^k, \lambda^k) + \frac{r_k}{2}\|x - x^k\|^2\right\}, \tag{2.2a}$$

$$y^{k+1} = \operatorname{Arg}\min_{y\in R^{n_2}} \left\{L_{\rho_k}(x^{k+1}, y, \lambda^k) + \frac{s_k}{2}\|y - y^k\|^2\right\}, \tag{2.2b}$$

$$\lambda^{k+1} = \operatorname{Arg}\max_{\lambda\in R^{m}} \left\{L_{\rho_{k+1}}(x^{k+1}, y^{k+1}, \lambda) - \frac{1}{2\alpha_k\rho_{k+1}}\|\lambda - \lambda^k\|^2\right\}. \tag{2.2c}$$

In the implementation of the scheme (2.2), the penalty parameter is heuristically updated by

$$\rho_{k+1} = \max\{10 * \|\lambda_k\|_\infty, \rho_k\}, \tag{2.3}$$

which is enlightened by Bertesekas [7], Nocedal and Wright [23]. Thus $\rho_{k+1} \ge \rho_k$. The subproblem (2.2c) has the closed solution:

$$\lambda^{k+1} = \lambda^k + \alpha_k\rho_{k+1}(\Phi(x^{k+1}, y^{k+1}) - b). \tag{2.4}$$

By (2.2c), we have

$$L_{\rho_{k+1}}(x^{k+1}, y^{k+1}, \lambda^{k+1}) - \frac{1}{2\alpha_k\rho_{k+1}}\|\lambda^{k+1} - \lambda^k\|^2 \ge L_{\rho_{k+1}}(x^{k+1}, y^{k+1}, \lambda^k).$$

Since $\rho_{k+1} \ge \rho_k > 0$, which implies $L_{\rho_{k+1}}(x^{k+1}, y^{k+1}, \lambda^k) \ge L_{\rho_k}(x^{k+1}, y^{k+1}, \lambda^k)$. Then

$$L_{\rho_{k+1}}(x^{k+1}, y^{k+1}, \lambda^{k+1}) - \frac{1}{2\alpha_k\rho_{k+1}}\|\lambda^{k+1} - \lambda^k\|^2 \ge L_{\rho_k}(x^{k+1}, y^{k+1}, \lambda^k). \tag{2.5}$$

Hence, we select the step length $\alpha_k > 0$ by a line-search technique such that $\alpha_k \rho_{k+1} \geq \beta > 0$ and

$$L_{\rho_{k+1}}(x^{k+1}, y^{k+1}, \lambda^{k+1}) - L_{\rho_k}(x^{k+1}, y^{k+1}, \lambda^k) \leq \frac{\eta_k}{\alpha_k}, \tag{2.6}$$

where

$$\sum_{k=0}^{\infty} \eta_k < \infty. \tag{2.7}$$

The scheme (2.2) has a form similar to (1.9), except that we use a line-search technique to select a step length α_k in (2.2c) satisfying the conditions stated above, and update the penalty parameter ρ_k by a heuristic rule. The non-convexity of the augmented Lagrangian function (2.1) leads essential difference between scheme (2.2) and scheme (1.9). The choices of proximal parameters r_k and s_k play a crucial role in solving the subproblems (2.2a) and (2.2b), respectively. Inspired by He et al. [18], we propose a self-adaptive style for choosing these parameters to guarantee that the subproblems (2.2a) and (2.2b) satisfy the convergence conditions of the scheme (2.2) under some suitable assumptions.

Throughout this paper, all mentioned subgradients are in the sense of Clark's generalized subgradient. The following notations are used for convenience in the future discussion: Let $f(x, y, \lambda)$ be a subgradient of $L_\rho(x, y, \lambda)$ w.r.t. x for fixed y and λ, and $g(x, y, \lambda)$ be a subgradient of $L_\rho(x, y, \lambda)$ w.r.t. y for fixed x and λ, respectively. That is:

$$f(x, y, \lambda) \in \partial_x L_\rho(x, y, \lambda), \qquad g(x, y, \lambda) \in \partial_y L_\rho(x, y, \lambda).$$

Let $J(x, y)$ be a generalized Jacobi matrix of $\Phi(x, y)$ w.r.t. x for fixed y, and $K(x, y)$ be a generalized Jacobi matrix of $\Phi(x, y)$ w.r.t. y for fixed x, respectively. That is:

$$J(x, y) \in \Phi'_x(x, y), \qquad K(x, y) \in \Phi'_y(x, y).$$

Let $\zeta(x) \in \partial\theta_1(x)$ be a subgradient of $\theta_1(x)$, and $\varsigma(y) \in \partial\theta_2(y)$ be a subgradient of $\theta_2(y)$. Unless otherwise stated, $\|u\|$ denotes 2-norm if u is a vector, and F-norm if u is a matrix.

We make the following assumptions to specialize the problem to be solved.

A. The vector-valued function $\Phi(x, y)$ is uniformly bounded and Lipschitzian. That is, there exists $B > 0$, such that

$$\|\Phi(x, y)\| \leq B, \quad \forall (x, y) \in R^{n_1} \times R^{n_2}, \tag{2.8}$$

and there exists $L_\Phi > 0$ such that

$$\|(\Phi(x + h, y + l) - \Phi(x, y + l)) - (\Phi(x + h, y) - \Phi(x, y))\| \leq L_\Phi(\|h\| + \|l\|) \tag{2.9}$$

for all $(x, y), (x + h, y), (x, y + l), (x + h, y + l) \in R^{n_1} \times R^{n_2}$, see Antipin [1,2].

B. The generalized Jacobi matrix of function $\Phi(x, y)$ is uniformly bounded and Lipschitzian. That is, there exits $M > 0$ such that

$$\|J(x, y)\| \leq M, \qquad \|K(x, y)\| \leq M, \quad \forall (x, y) \in R^{n_1} \times R^{n_2}, \tag{2.10}$$

and there exists $L_D > 0$ such that

$$\|(J(x+h, y+l) - J(x, y+l)) - (J(x+h, y) - J(x, y))\| \leq L_D(\|h\| + \|l\|), \tag{2.11a}$$

$$\|(K(x+h, y+l) - K(x, y+l)) - (K(x+h, y) - K(x, y))\| \leq L_D(\|h\| + \|l\|), \tag{2.11b}$$

for all $(x, y), (x+h, y), (x, y+l), (x+h, y+l) \in R^{n_1} \times R^{n_2}$.

3 Preliminaries

The Lagrangian function of the problem (1.6) (without the penalty term $\frac{\rho}{2}\|\Phi(x, y) - b\|^2$) is

$$L_0(x, y, \lambda) = \theta_1(x) + \theta_2(y) + \lambda^T(\Phi(x, y) - b). \tag{3.1}$$

The augmented Lagrangian function (2.1) can be viewed as the Lagrangian associated to the problem

$$\begin{cases} \min\ \theta_1(x) + \theta_2(y) + \frac{\rho}{2}\|\Phi(x, y) - b\|^2 \\ \text{s.t.}\quad \Phi(x, y) - b = 0. \end{cases} \tag{3.2}$$

For regularity, the following assumption is common:

C. There exists at least a triple $(x^*, y^*, \lambda^*) \in R^{n_1} \times R^{n_2} \times R^m$ satisfying the Karush-Kuhn-Tucker conditions of the constrained minimization problem (1.6). That is, there exist vectors $\zeta(x^*) \in \partial\theta_1(x^*)$ and $\varsigma(y^*) \in \partial\theta_2(y^*)$, and $\lambda^* \in R^m$ such that

$$\zeta(x^*) + J(x^*, y^*)^T\lambda^* = 0, \tag{3.3a}$$

$$\varsigma(y^*) + K(x^*, y^*)^T\lambda^* = 0, \tag{3.3b}$$

$$\Phi(x^*, y^*) - b = 0. \tag{3.3c}$$

Assume that the LICQ condition holds at the point (x^*, y^*). Moreover, we assume that the sequence $\{\lambda^k\}$ generated by (2.2) is bounded, i.e., there exists a constant $\Lambda > 0$ such that $\|\lambda^k\| \leq \Lambda$ for all k, see Bertesekas [7].

By the assumption, there is a threshold value of the penalty parameter $\bar{\rho} = \|\lambda^*\|_\infty > 0$ such that, for all $\rho > \bar{\rho}$ the pair (x^*, y^*) is a stationary point of the augmented Lagrangian function $L_\rho(x, y, \lambda)$ with $\lambda = \lambda^*$. The update rule of the penalty parameter ρ_k, i.e., (2.3), is also inspired by this assertion, see Nocedal and Wright [23].

Note that

$$\partial_x\left(\frac{\rho}{2}\|\Phi(x, y) - b\|^2\right) \ni \rho J(x, y)^T[\Phi(x, y) - b],$$

$$\partial_y\left(\frac{\rho}{2}\|\Phi(x, y) - b\|^2\right) \ni \rho K(x, y)^T[\Phi(x, y) - b].$$

Due to the scheme (2.2), we have the following monotonicity theorem.

Theorem 3.1 *Suppose the assumptions A and B hold. Then, for the given $\rho_k > 0$ and $\tau > 0$, there exists $r_k > 0$ such that*

$$\left\langle x - x', \left(f(x, y^k, \lambda^k) - f(x', y^k, \lambda^k)\right) + r_k(x - x')\right\rangle \geq \tau\|x - x'\|^2, \quad \forall x, x' \in R^{n_1}. \tag{3.4}$$

Similarly, there exists $s_k > 0$ such that

$$\Big\langle y - y', \big(g(x^{k+1}, y, \lambda^k) - g(x^{k+1}, y', \lambda^k)\big) + s_k(y - y')\Big\rangle \geq \tau\|y - y'\|^2, \quad \forall y, y' \in R^{n_2}. \tag{3.5}$$

Proof We only prove (3.4). The assertion (3.5) can be proved by the same way.

By a direct computation, we get

$$\begin{aligned}
&\Big\langle x - x', f(x, y^k, \lambda^k) - f(x', y^k, \lambda^k) + r_k(x - x')\Big\rangle \\
&\quad = \langle x - x', \zeta(x) - \zeta(x')\rangle + \Big\langle x - x', \Big[J(x, y^k) - J(x', y^k)\Big]^T \lambda^k\Big\rangle \\
&\qquad + \rho_k\Big\langle x - x', J(x, y^k)^T(\Phi(x, y^k) - b) - J(x', y^k)^T(\Phi(x', y^k) - b)\Big\rangle + r_k\|x - x'\|^2,
\end{aligned} \tag{3.6}$$

where $\zeta(x) \in \partial\theta_1(x)$. Since θ_1 is convex, we have

$$\langle x - x', \zeta(x) - \zeta(x')\rangle \geq 0. \tag{3.7}$$

Since $J(x, y)$ is Lipschitzian and $\{\lambda^k\}$ is bounded, by Cauchy–Schwarz inequality we get

$$\begin{aligned}
&\Big\langle x - x', \Big[J(x, y^k) - J(x', y^k)\Big]^T \lambda^k\Big\rangle \\
&\quad \geq -\|x - x'\| \cdot \Big\|\Big[J(x, y^k) - J(x', y^k)\Big]^T \lambda^k\Big\| \\
&\quad \geq -\|x - x'\| \cdot \|J(x, y^k) - J(x', y^k)\| \cdot \|\lambda^k\| \\
&\quad \geq -L_D\Lambda\|x - x'\|^2.
\end{aligned} \tag{3.8}$$

Now we deal with the third term of the right-hand side of (3.6). By a manipulation,

$$\begin{aligned}
&\big\langle x - x', J(x, y^k)^T(\Phi(x, y^k) - b) - J(x', y^k)^T(\Phi(x', y^k) - b)\big\rangle \\
&\quad = \big\langle x - x', J(x, y^k)^T(\Phi(x, y^k) - b) - J(x, y^k)^T(\Phi(x', y^k) - b)\big\rangle \\
&\qquad + \big\langle x - x', J(x, y^k)^T(\Phi(x', y^k) - b) - J(x', y^k)^T(\Phi(x', y^k) - b)\big\rangle \\
&\quad = \big\langle x - x', J(x, y^k)^T\big[\Phi(x, y^k) - \Phi(x', y^k)\big]\big\rangle \\
&\qquad + \Big\langle x - x', \Big[J(x, y^k) - J(x', y^k)\Big]^T \big(\Phi(x', y^k) - b\big)\Big\rangle.
\end{aligned} \tag{3.9}$$

Noting that $J(x, y)$ is bounded and $\Phi(x, y)$ is Lipschitzian, by the Cauchy–Schwarz inequality we have

$$\begin{aligned}
&\Big\langle x - x', J(x, y^k)^T\Big[\Phi(x, y^k) - \Phi(x', y^k)\Big]\Big\rangle \\
&\quad \geq -\|x - x'\| \cdot \Big\|J(x, y^k)^T\Big[\Phi(x, y^k) - \Phi(x', y^k)\Big]\Big\| \\
&\quad \geq -\|x - x'\| \cdot \|J(x, y^k)\| \cdot \|\Phi(x, y^k) - \Phi(x', y^k)\| \\
&\quad \geq -\|x - x'\| \cdot M \cdot L_\Phi\|x - x'\| = -L_\Phi M\|x - x'\|^2.
\end{aligned} \tag{3.10}$$

Noting that $J(x, y)$ is Lipschitzian and $\Phi(x, y)$ is bounded, again using the Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
&\left\langle x-x', \left[J(x,y^k)-J(x',y^k)\right]^T \left(\Phi(x',y^k)-b\right)\right\rangle \\
&\quad \geq -\|x-x'\|\cdot\left\|\left[J(x,y^k)-J(x',y^k)\right]^T\left(\Phi(x',y^k)-b\right)\right\| \\
&\quad \geq -\|x-x'\|\cdot\|J(x,y^k)-J(x',y^k)\|\cdot\|\Phi(x',y^k)-b\| \\
&\quad \geq -L_D(B+\|b\|)\|x-x'\|^2. \qquad (3.11)
\end{aligned}
$$

Substituting (3.10)–(3.11) into (3.9), we obtain

$$
\begin{aligned}
&\left\langle x-x', J(x,y^k)^T(\Phi(x,y^k)-b)-J(x',y^k)^T(\Phi(x',y^k)-b)\right\rangle \\
&\quad \geq -(L_\Phi M+L_D(B+\|b\|))\|x-x'\|^2. \qquad (3.12)
\end{aligned}
$$

Combining (3.7)–(3.8) and (3.12), we have

$$
\begin{aligned}
&\left\langle x-x', f(x,y^k,\lambda^k)-f(x',y^k,\lambda^k)+r_k(x-x')\right\rangle \\
&\quad \geq \Big(r_k-\big(L_D\Lambda+\rho_k(L_D(B+\|b\|)+L_\Phi M)\big)\Big)\|x-x'\|^2. \qquad (3.13)
\end{aligned}
$$

Then by letting

$$
r_k \geq \tau + L_D\Lambda + \rho_k(L_D(B+\|b\|)+L_\Phi M),
$$

we have (3.4). Symmetrically, by letting

$$
s_k \geq \tau + L_D\Lambda + \rho_k(L_D(B+\|b\|)+L_\Phi M),
$$

we have (3.5). This completes the proof. □

Theorem 3.1 shows that, for a given $\rho_k > 0$, there exist $r_k > 0$ and $s_k > 0$ such that the objective functions of the subproblems (2.2a) and (2.2b) are strictly convex [19]. However, the lower bounds of r_k and s_k depend on some unknown parameters, such as L_D, L_ϕ, B and M. So, in practical computation we use an adaptive strategy to get some suitable choices of r_k and s_k, respectively. Please see Sect. 5 for our adaptive strategy.

By the strict convexity, x^{k+1} is the unique solution of the subproblem (2.2a), and y^{k+1} is the unique solution of the subproblem (2.2b). Thus we have

$$
L_{\rho_k}(x^{k+1},y^k,\lambda^k)+\frac{r_k}{2}\|x^{k+1}-x^k\|^2 \leq L_{\rho_k}(x,y^k,\lambda^k)+\frac{r_k}{2}\|x-x^k\|^2, \forall x\in R^{n_1}, \qquad (3.14)
$$

and

$$
L_{\rho_k}(x^{k+1},y^{k+1},\lambda^k)+\frac{s_k}{2}\|y^{k+1}-y^k\|^2 \leq L_{\rho_k}(x^{k+1},y,\lambda^k)+\frac{s_k}{2}\|y-y^k\|^2, \forall y\in R^{n_2}. \qquad (3.15)
$$

The equalities of (3.14) and (3.15) hold if and only if $x = x^{k+1}$ and $y = y^{k+1}$.

By letting $x = x^k$ in (3.14), we get

$$
L_{\rho_k}(x^{k+1},y^k,\lambda^k)+\frac{r_k}{2}\|x^{k+1}-x^k\|^2 \leq L_{\rho_k}(x^k,y^k,\lambda^k), \qquad (3.16)
$$

and by letting $y = y^k$ in (3.15), we obtain

$$
L_{\rho_k}(x^{k+1},y^{k+1},\lambda^k)+\frac{s_k}{2}\|y^{k+1}-y^k\|^2 \leq L_{\rho_k}(x^{k+1},y^k,\lambda^k). \qquad (3.17)
$$

Adding (3.16) and (3.17) yields

$$
L_{\rho_k}(x^{k+1},y^{k+1},\lambda^k)+\frac{r_k}{2}\|x^{k+1}-x^k\|^2+\frac{s_k}{2}\|y^{k+1}-y^k\|^2 \leq L_{\rho_k}(x^k,y^k,\lambda^k). \qquad (3.18)
$$

The inequality (3.18) implies that the scheme (2.2) is primal descent.

4 Convergence analysis

By the assumptions A and C, the level set $H = \{(x, y, \lambda) \in R^{n_1} \times R^{n_2} \times R^m : L_\rho(x, y, \lambda) \leq l_0\}$ is closed and bounded for all fixed $\rho < \infty$ and all $l_0 \in R$. By (3.18), there are $\rho > 0$ and $l_0 > 0$ large enough such that $\{(x^k, y^k, \lambda^k)\} \subset H$ for all k. By the Weierstrass theorem, the sequence $\{(x^k, y^k, \lambda^k)\}$ admits an accumulation point. Then, there exists a convergent subsequence $\{(x^k, y^k, \lambda^k)\}_{k\in\mathcal{K}}$. For convenience, assume that:

$$(x^\infty, y^\infty, \lambda^\infty) := \lim_{k\to\infty, k\in\mathcal{K}} (x^k, y^k, \lambda^k).$$

We prove the convergence to a KKT point of the proposed proximal ADMM, by verifying that the accumulation point $(x^\infty, y^\infty, \lambda^\infty)$ of the convergent subsequence $\{x^k, y^k, \lambda^k\}_{k\in\mathcal{K}}$ satisfies the KKT conditions of the problem (1.6). For simplification, the notation '$\lim\limits_{k\to\infty}$' is used to denote '$\lim\limits_{k\to\infty, k\in\mathcal{K}}$' in what follows.

Lemma 4.1 *Suppose the assumption C holds and the sequence $\{\lambda^k\}$ is generated by the scheme (2.2). Then we have*

$$\lim_{k\to\infty} \|\lambda^{k+1} - \lambda^k\|^2 = 0. \tag{4.1}$$

Proof By (2.5), we get

$$\frac{1}{2\rho_{k+1}} \|\lambda^{k+1} - \lambda^k\|^2 \leq \alpha_k \Big(L_{\rho_{k+1}}(x^{k+1}, y^{k+1}, \lambda^{k+1}) - L_{\rho_k}(x^{k+1}, y^{k+1}, \lambda^k)\Big). \tag{4.2}$$

Adding (4.2) from $k = 0$ to ∞, by (2.6) and (2.7), we have

$$\sum_{k=0}^{\infty} \frac{1}{2\rho_{k+1}} \|\lambda^{k+1} - \lambda^k\|^2 \leq \sum_{k=0}^{\infty} \eta_k < \infty.$$

On the other hand, by the assumption C and (2.3), we have $\rho_{k+1} \geq \rho_k > 0$, and there exists a real number $\varrho > 0$ sufficiently large such that $\rho_k < \varrho$ for all k. Thus

$$\frac{1}{2\varrho} \sum_{k=0}^{\infty} \|\lambda^{k+1} - \lambda^k\|^2 < \infty$$

which follows (4.1) and completes the proof. □

Theorem 4.1 *Suppose the sequence $\{(x^k, y^k)\}$ is generated by the scheme (2.2). Then the primal residual satisfies*

$$\lim_{k\to\infty} \|\Phi(x^k, y^k) - b\| = 0. \tag{4.3}$$

Theorem 4.1 can be immediately derived from Lemma 4.1, and it provides the primal feasibility of the iterate (x^k, y^k) at the limit. The condition (3.3c) is a direct consequence of this theorem.

Lemma 4.2 *Suppose the sequence $\{(x^k, y^k)\}$ is generated by the scheme (2.2), and the assumptions A and B hold. Assume that the proximal parameters r_k and s_k satisfy the conditions given by Theorem 3.1. Then we have*

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0, \tag{4.4a}$$

$$\lim_{k\to\infty} \|y^{k+1} - y^k\| = 0. \tag{4.4b}$$

Proof By the assumptions A, B and Theorem 3.1, the subproblems (2.2a) and (2.2b) are strictly convex which results in (3.18). That is

$$L_{\rho_k}(x^{k+1}, y^{k+1}, \lambda^k) + \frac{r_k}{2}\|x^{k+1} - x^k\|^2 + \frac{s_k}{2}\|y^{k+1} - y^k\|^2 \le L_{\rho_k}(x^k, y^k, \lambda^k). \quad (4.5)$$

By the identity

$$\Phi(x^{k+1}, y^{k+1}) - b = \frac{1}{\alpha_k \rho_{k+1}}(\lambda^{k+1} - \lambda^k),$$

we have

$$\begin{aligned} &L_{\rho_k}(x^{k+1}, y^{k+1}, \lambda^k) - L_{\rho_{k+1}}(x^{k+1}, y^{k+1}, \lambda^{k+1}) \\ &= (\lambda^k - \lambda^{k+1})^T \big(\Phi(x^{k+1}, y^{k+1}) - b\big) - \frac{\rho_{k+1} - \rho_k}{2}\|\Phi(x^{k+1}, y^{k+1}) - b\|^2 \\ &= -\left\{\frac{1}{\alpha_k \rho_{k+1}}\left(1 + \frac{\rho_{k+1} - \rho_k}{2\alpha_k \rho_{k+1}}\right)\right\}\|\lambda^{k+1} - \lambda^k\|^2. \end{aligned} \quad (4.6)$$

Substituting (4.6) to (4.5) yields

$$\begin{aligned} &L_{\rho_{k+1}}(x^{k+1}, y^{k+1}, \lambda^{k+1}) + \frac{r_k}{2}\|x^{k+1} - x^k\|^2 + \frac{s_k}{2}\|y^{k+1} - y^k\|^2 \\ &\le L_{\rho_k}(x^k, y^k, \lambda^k) + \left\{\frac{1}{\alpha_k \rho_{k+1}}\left(1 + \frac{\rho_{k+1} - \rho_k}{2\alpha_k \rho_{k+1}}\right)\right\}\|\lambda^{k+1} - \lambda^k\|^2. \end{aligned}$$

Since $\alpha_k \rho_{k+1} \ge \beta > 0$ and $\rho_k < \varrho$ for all k, we get

$$\begin{aligned} &L_{\rho_{k+1}}(x^{k+1}, y^{k+1}, \lambda^{k+1}) + \frac{r_k}{2}\|x^{k+1} - x^k\|^2 + \frac{s_k}{2}\|y^{k+1} - y^k\|^2 \\ &< L_{\rho_k}(x^k, y^k, \lambda^k) + \left(\frac{1}{\beta} + \frac{\varrho}{2\beta^2}\right)\|\lambda^{k+1} - \lambda^k\|^2. \end{aligned} \quad (4.7)$$

Adding (4.7) from $k = 0$ to ∞, we get

$$\begin{aligned} &\sum_{k=0}^{\infty}\left[\frac{r_k}{2}\|x^{k+1} - x^k\|^2 + \frac{s_k}{2}\|y^{k+1} - y^k\|^2\right] \\ &< L_{\rho_0}(x^0, y^0, \lambda^0) - \lim_{k\to\infty} L_{\rho_{k+1}}(x^{k+1}, y^{k+1}, \lambda^{k+1}) + \left(\frac{1}{\beta} + \frac{\varrho}{2\beta^2}\right)\sum_{k=0}^{\infty}\|\lambda^{k+1} - \lambda^k\|^2. \end{aligned} \quad (4.8)$$

By the boundedness of $L_\rho(x, y, \lambda)$ and Theorem 4.1, we have

$$\sum_{k=0}^{\infty}\left[\frac{r_k}{2}\|x^{k+1} - x^k\|^2 + \frac{s_k}{2}\|y^{k+1} - y^k\|^2\right] < \infty. \quad (4.9)$$

By Theorem 3.1, r_k and s_k are strictly positive (bounded away from zero), hence we have (4.4) and complete the proof. □

By the optimality condition of the subproblems (2.2a) and (2.2b), there exist $f(x^{k+1}, y^k, \lambda^k) \in \partial_x L_{\rho_k}(x^{k+1}, y^k, \lambda^k)$ and $g(x^{k+1}, y^{k+1}, \lambda^k) \in \partial_y L_{\rho_k}(x^{k+1}, y^{k+1}, \lambda^k)$, such that

$$f(x^{k+1}, y^k, \lambda^k) + r_k(x^{k+1} - x^k) = 0, \quad (4.10a)$$

$$g(x^{k+1}, y^{k+1}, \lambda^k) + s_k(y^{k+1} - y^k) = 0. \quad (4.10b)$$

Theorem 4.2 *Suppose the sequence $\{x^k, y^k, \lambda^k\}$ is generated by the proposed proximal ADMM, and the assumptions A and B hold. Then, there exist a vector $\varsigma(y^{k+1}) \in \partial\theta_2(y^{k+1})$ and a matrix $K(x^{k+1}, y^{k+1}) \in \Phi'_y(x^{k+1}, y^{k+1})$, such that*

$$\lim_{k\to\infty} \left\| \varsigma(y^{k+1}) + K(x^{k+1}, y^{k+1})^T \lambda^{k+1} \right\| = 0. \tag{4.11}$$

Proof By (2.4), there exist a vector $\varsigma(y^{k+1}) \in \partial\theta_2(y^{k+1})$ and a matrix $K(x^{k+1}, y^{k+1}) \in \Phi'_y(x^{k+1}, y^{k+1})$, such that

$$\begin{aligned}
&g(x^{k+1}, y^{k+1}, \lambda^k)\\
&\quad = \varsigma(y^{k+1}) + K(x^{k+1}, y^{k+1})^T \lambda^k + \rho_k K(x^{k+1}, y^{k+1})^T \big(\Phi(x^{k+1}, y^{k+1}) - b\big)\\
&\quad = \varsigma(y^{k+1}) + K(x^{k+1}, y^{k+1})^T \Big\{\lambda^k + \alpha_k\rho_{k+1}\big(\Phi(x^{k+1}, y^{k+1}) - b\big)\Big\}\\
&\qquad + K(x^{k+1}, y^{k+1})^T \big(\rho_k - \alpha_k\rho_{k+1}\big)\big(\Phi(x^{k+1}, y^{k+1}) - b\big)\\
&\quad = \varsigma(y^{k+1})\\
&\qquad + K(x^{k+1}, y^{k+1})^T \lambda^{k+1} + \big(\rho_k - \alpha_k\rho_{k+1}\big) K(x^{k+1}, y^{k+1})^T \big(\Phi(x^{k+1}, y^{k+1}) - b\big).
\end{aligned}$$

By the triangle inequality and taking limit on the both sides of (4.10b), we get

$$\begin{aligned}
&\lim_{k\to\infty} \Big\{\|g(x^{k+1}, y^{k+1}, \lambda^k)\| - \|s_k(y^{k+1} - y^k)\|\Big\}\\
&\quad \le \lim_{k\to\infty} \left\| g(x^{k+1}, y^{k+1}, \lambda^k) + s_k(y^{k+1} - y^k) \right\| = 0.
\end{aligned} \tag{4.12}$$

Adding (4.12) and (4.4b) yields

$$\lim_{k\to\infty} \left\| g(x^{k+1}, y^{k+1}, \lambda^k) \right\| = 0. \tag{4.13}$$

By the Cauchy–Schwarz inequality and the assumption B, we get

$$\begin{aligned}
&\|\big(\rho_k - \alpha_k\rho_{k+1}\big) K(x^{k+1}, y^{k+1})^T \big(\Phi(x^{k+1}, y^{k+1}) - b\big)\|\\
&\quad \le |\rho_k - \alpha_k\rho_{k+1}| \times \|K(x^{k+1}, y^{k+1})\| \times \|\Phi(x^{k+1}, y^{k+1}) - b\|\\
&\quad \le |\rho_k - \alpha_k\rho_{k+1}| \cdot M \times \|\Phi(x^{k+1}, y^{k+1}) - b\|.
\end{aligned}$$

Using the triangle inequality again, we get

$$\begin{aligned}
&\|g(x^{k+1}, y^{k+1}, \lambda^k)\|\\
&\quad = \|\varsigma(y^{k+1}) + K(x^{k+1}, y^{k+1})^T \lambda^{k+1}\\
&\qquad + \big(\rho_k - \alpha_k\rho_{k+1}\big) K(x^{k+1}, y^{k+1})^T \big(\Phi(x^{k+1}, y^{k+1}) - b\big)\|\\
&\quad \ge \|\varsigma(y^{k+1}) + K(x^{k+1}, y^{k+1})^T \lambda^{k+1}\|\\
&\qquad - \|\big(\rho_k - \alpha_k\rho_{k+1}\big) K(x^{k+1}, y^{k+1})^T \big(\Phi(x^{k+1}, y^{k+1}) - b\big)\|\\
&\quad \ge \|\varsigma(y^{k+1}) + K(x^{k+1}, y^{k+1})^T \lambda^{k+1}\| - |\rho_k - \alpha_k\rho_{k+1}| \cdot M \times \|\Phi(x^{k+1}, y^{k+1}) - b\|.
\end{aligned}$$

Taking limits on the both sides of the above inequality and noting $|\rho_k - \alpha_k\rho_{k+1}| < \varrho + \beta$ for all k, by Corollary 4.1 and (4.13), we have

$$\lim_{k\to\infty} \left\| \varsigma(y^{k+1}) + K(x^{k+1}, y^{k+1})^T \lambda^{k+1} \right\| = 0. \tag{4.14}$$

□

Theorem 4.2 implies that the accumulation point $(x^\infty, y^\infty, \lambda^\infty)$ satisfies the condition (3.3b).

Theorem 4.3 *Suppose that the sequence $\{(x^k, y^k, \lambda^k)\}$ is generated by the scheme (2.2), and the assumptions A and B hold. Then, there exist a subgradient $\zeta(x^k) \in \partial\theta_1(x^k)$ and a matrix $J(x^k, y^k) \in \Phi_x'(x^k, y^k)$, such that*

$$\lim_{k\to\infty} \|\zeta(x^k) + J(x^k, y^k)^T\lambda^k\| = 0. \tag{4.15}$$

Proof By the optimality condition of subproblem (2.2a), we have

$$f(x^{k+1}, y^k, \lambda^k) + r_k(x^{k+1} - x^k) = 0.$$

By (2.2c), there exist a vector $\zeta(x^{k+1}) \in \partial\theta_1(x^{k+1})$ and matrices $J(x^{k+1}, y^k) \in \Phi_x'(x^{k+1}, y^k)$ and $J(x^{k+1}, y^{k+1}) \in \Phi_x'(x^{k+1}, y^{k+1})$, such that

$$\begin{aligned}
&f(x^{k+1}, y^k, \lambda^k)\\
&\quad= \zeta(x^{k+1}) + J(x^{k+1}, y^k)^T\lambda^k + \rho_k J(x^{k+1}, y^k)^T\left(\Phi(x^{k+1}, y^k) - b\right)\\
&\quad= \zeta(x^{k+1}) + J(x^{k+1}, y^{k+1})^T\lambda^{k+1} + \xi^k,
\end{aligned}$$

where

$$\xi^k = \left\{\begin{array}{l}
\left[J(x^{k+1}, y^k) - J(x^{k+1}, y^{k+1})\right]^T\lambda^k\\
+\rho_k\left[J(x^{k+1}, y^k) - J(x^{k+1}, y^{k+1})\right]^T\left(\Phi(x^{k+1}, y^k) - b\right)\\
+J(x^{k+1}, y^{k+1})^T\left\{\begin{array}{l}\rho_k\left(\Phi(x^{k+1}, y^k) - (\Phi(x^{k+1}, y^{k+1})\right)\\ +(\rho_k - \alpha_k\rho_{k+1})(\Phi(x^{k+1}, y^{k+1}) - b))\end{array}\right\}
\end{array}\right\}.$$

By the assumptions and Lemma 4.2, we deduce that

$$\lim_{k\to\infty} \|\xi^k\| = 0. \tag{4.16}$$

On the other hand, taking limits on the both sides of (4.10a) we get

$$\begin{aligned}
&\lim_{k\to\infty}\left\{\|f(x^{k+1}, y^k, \lambda^k)\| - \|r_k(x^{k+1} - x^k)\|\right\}\\
&\quad\le \lim_{k\to\infty}\|f(x^{k+1}, y^k, \lambda^k) + r_k(x^{k+1} - x^k)\| = 0.
\end{aligned} \tag{4.17}$$

Multiplying r_k on the both sides of (4.4a), adding the resulting equation and (4.17), we obtain

$$\lim_{k\to\infty}\|f(x^{k+1}, y^k, \lambda^k)\| = 0. \tag{4.18}$$

Hence, by the triangle inequality we have

$$\begin{aligned}
&\lim_{k\to\infty}\left\{\|\zeta(x^{k+1}) + J(x^{k+1}, y^{k+1})^T\lambda^{k+1}\| - \|\xi^k\|\right\}\\
&\quad\le \lim_{k\to\infty}\|\zeta(x^{k+1}) + J(x^{k+1}, y^{k+1})^T\lambda^{k+1} + \xi^k\|\\
&\quad= \lim_{k\to\infty}\|f(x^{k+1}, y^k, \lambda^k)\| = 0.
\end{aligned} \tag{4.19}$$

The assertion of Theorem 4.3 follows from (4.16) and (4.19) directly. □

Theorem 4.3 indicates that the accumulation point $(x^\infty, y^\infty, \lambda^\infty)$ satisfies (3.3a).

Up to now, we have checked that, the accumulation point $(x^\infty, y^\infty, \lambda^\infty)$ of the convergent subsequence $\{(x^k, y^k, \lambda^k)\}_{k\in\mathcal{K}}$ satisfies the KKT conditions (3.3). Thus, it is a KKT point of the problem (1.6).

5 Practical implementation and numerical results

The inequalities (3.16) and (3.17) play a critical role in the convergence of the scheme (2.2). However, the proximal parameters r_k and s_k depend on some unknown parameters. Thus, a self-adaptive strategy is presented in this section for the suitable choice of these parameters.

5.1 Practical implementation

Inspired by [18], in the practical implementation of the proposed method (2.2), we choose the proximal parameters r_k of (2.2a), and respectively, s_k of (2.2b), via a self-adaptive rule. Based on this idea, we propose a practical implementation in what follows.

Let $\nu = \frac{\sqrt{5}+1}{2}$, for a given triple (x^k, y^k, λ^k) and $\rho_k > 0$, the new iterate $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is generated via the following scheme:

s1. Let x^{k+1} be a solution of (2.2a).
Let $\delta_x^k = L_{\rho_k}(x^k, y^k, \lambda^k) - L_{\rho_k}(x^{k+1}, y^k, \lambda^k)$.
while $(\delta_x^k < \frac{r_k}{4}\|x^{k+1} - x^k\|^2)$
set $r_k = \nu * r_k$; let x^{k+1} be a solution of (2.2a),
and $\delta_x^k = L_{\rho_k}(x^k, y^k, \lambda^k) - L_{\rho_k}(x^{k+1}, y^k, \lambda^k)$.
end(while)
if $(\delta_x^k > 2r_k\|x^{k+1} - x^k\|^2)$ set $r_k = r_k/\nu$; **end** (if)

s2. Let y^{k+1} be a solution of (2.2b).
Let $\delta_y^k = L_{\rho_k}(x^{k+1}, y^k, \lambda^k) - L_{\rho_k}(x^{k+1}, y^{k+1}, \lambda^k)$.
while $(\delta_y^k < \frac{s_k}{4}\|y^{k+1} - y^k\|^2)$
set $s_k = \nu * s_k$; let y^{k+1} be a solution of (2.2b),
and $\delta_y^k = L_{\rho_k}(x^{k+1}, y^k, \lambda^k) - L_{\rho_k}(x^{k+1}, y^{k+1}, \lambda^k)$.
end(while)
if $(\delta_y^k > 2s_k\|y^{k+1} - y^k\|^2)$ set $s_k = s_k/\nu$; **end** (if)

s3. Update the penalty parameter by (2.3). Let $\lambda^{k+1} = \lambda^k + \alpha_k\rho_{k+1}(\Phi(x^{k+1}, y^{k+1}) - b)$, where the step size α_k is chosen to satisfy $\alpha_k\rho_{k+1} \geq \beta > 0$ and the conditions (2.6)–(2.7).

The self-adaptive rule stated above is different from that in [18]. The former one depends on the decrement of the augmented Lagrange function L_{ρ_k}, while the later one essentially depends on the change of its partial-gradient. We do not use the existing rule because of the computational expensiveness of the partial-gradient of the problem (1.6). The finite termination property of the " **while** " loops in step 1 and step 2 is guaranteed by Theorem 3.1.

The use of " **if** " in s1 (resp. s2) of the self-adaptive rule is to prepare a reduced r_k (resp. s_k) for the next iteration if the step size is too small, which is the same as that in s1.1 (resp. s1.2) of the self-adaptive approximate PPA based prediction-correction method [18].

The practical termination criterion of the scheme (2.2) is set as

$$\max\left\{\|x^{k+1} - x^k\|_\infty, \|y^{k+1} - y^k\|_\infty, \|\lambda^{k+1} - \lambda^k\|_\infty\right\} < \varepsilon \tag{5.1}$$

with a given small $\varepsilon > 0$.

An efficient method for solving the subproblems (2.2a) and (2.2b) is actually problem-based. In the numerical experiments, the minimization problem (1.5) raised in VLSI is solved by the proximal ADMM. In the constraints of this problem, the density $D_b(x, y)$ is the sum of products of two piecewise linear functions associated to x and y respectively, see (1.4). To simplify the computation of the subgradient, a Gauss-Seidel-like coordinate descent algorithm is performed for solving the subproblems (2.2a) and (2.2b). The Gauss-Seidel-like coordinate descent method is stated as follows:

Gauss-Seidel coordinate descent method [for $\min_{x \in R^n} \phi(x)$]
For $i = 1, 2, \ldots, n$, let

$$x_i^{k+1} = \text{Arg}\min_{x_i} \phi(x_1^{k+1}, \ldots, x_{i-1}^{k+1}, x_i, x_{i+1}^k, \ldots, x_n^k). \tag{5.2}$$

5.2 Numerical results

To demonstrate the efficiency, we used the proximal ADMM as a global placement method for VLSI, and tested the resulting placer on the IBM standard cell benchmarks [30] shown in Table 1. To be fair and convenient for comparisons, the legalization and detailed placement methods of NTUplace3 [10] were directly used in conjunction with our placer. The unique and distinguish difference between our placement tool and NTUplace3 is that, for solving the problem (1.5) in the global placement stage, the proximal ADMM is performed in our placement tool, while a penalty function method is performed in NTUplace3. The NTUplace3 obtained high quality results in the International Symposium on Physical Design (ISPD) 2005, and achieved the best average HPWLs on ISPD 2006 contest benchmarks. So these comparisons are sufficient to show the efficiency of the proposed method.

We also compared our placer with the other state-of-the-art placement methods. They are: mPL6 [8], fastplace3.0 [26], fengshui5.0 [3] and capo10.5-umpack [25]. The mPL6, one of the strongest existing placers, uses the generalized force-directed placement to guide global placement at each level of the clustering hierarchy. Fastplace3.0 is a scalable and fast multilevel placer, which uses the quadratic placement technique. Fengshui5.0, an hMETIS partitioner based placer, is a state-of the-art min-cut placer. The capo10.5-umpack, the latest version, gains interests in the industrial and academic worlds.

All the placement methods were run on the same computer with Intel Core(TM) i5-2400 3.10GHz CPU and 4GB memory. One difference is that, capo10.5 was tested under the Windows operating system, and the other placers were tested under the Linux operating system. For fair comparisons, each of the compared placers was run in its default parameters, no manual tuning of parameters was allowed.

Table 2 lists the experimental results of mPL6, fastplace3.0, fengshui5.0, capo10.5, NTU-place3 and our placement method on the IBM standard cell benchmarks. The best result is remarked in bold for each circuit generated by the list placement methods.

Compared with NTUplace3, our placement method achieves shorter HPWL in all the 18 benchmarks. For examples, Figs. 1 and 2 show the placement results obtained by our placement method and NTUplace3 on the instances ibm01 and ibm08, respectively. It can be seen from these figures that our placement method can place the cells more compact, and results in less HPWL. Indeed, our placement method reduces the HPWL, compared with NTUplace3, about 3.21 and 6.97 % on the instances ibm01 and ibm08, respectively.

Among these placement methods on the 18 benchmarks, our placement method achieves 15 best HPWLs, mPL6 achieves 3 best HPWLs. The experimental results show that our placement method achieves high solution quality. Thus we can conclude from these experiments that, the proposed proximal ADMM is effective on solving the problem under consideration in this paper.

Table 1 The characteristics of IBM standard cell benchmarks

Circuit	No. of cells	No. of pads	No. of nets	No. of pins	No. of rows
ibm01	12,506	246	14,111	50,566	96
ibm02	19,342	259	19,584	81,199	109
ibm03	22,853	283	27,401	93,573	121
ibm04	27,220	287	31,970	105,859	136
ibm05	28,146	1201	28,446	126,308	139
ibm06	32,332	166	34,826	128,182	126
ibm07	45,639	287	48,117	175,639	166
ibm08	51,023	286	50,513	204,890	170
ibm09	53,110	285	60,902	222,088	183
ibm10	68,685	744	75,196	297,567	234
ibm11	70,152	406	81,454	280,786	208
ibm12	70,439	637	77,240	317,760	242
ibm13	83,709	490	99,666	357,075	224
ibm14	147,088	517	152,772	546,816	305
ibm15	161,187	383	186,608	715,823	303
ibm16	182,980	504	190,048	778,823	347
ibm17	184,752	743	189,581	860,036	379
ibm18	210,341	272	201,920	819,697	361

Table 2 The HPWL comparisons on IBM standard cell benchmarks among the placement methods

Circuit	mPL6	fastplace3.0	fengshui5.0	capo10.5	NTUplace3	Our method
ibm01	1624316	1727729	1760429	1750630	1686950	**1623806**
ibm02	3588165	3598083	3660498	3716180	3526805	**3456869**
ibm03	4729925	4646751	4656522	5488720	4733598	**4590319**
ibm04	5801143	5732461	5880070	5820890	5775934	**5666603**
ibm05	**9319907**	9884011	9818974	9787560	9944742	9356888
ibm06	**4848977**	4936872	5102631	5220080	5008886	4905972
ibm07	8065562	8227714	9014011	9000210	8358811	**8056921**
ibm08	8879299	9048117	ABORT	9551110	9189443	**8539797**
ibm09	9208787	9667507	9715123	11583200	9421964	**9052080**
ibm10	17451071	17268496	18026404	18357500	17381894	**17042556**
ibm11	13837322	14635573	14453292	14373500	13946979	**13776863**
ibm12	21734406	22838560	23517296	23971800	22228242	**21918813**
ibm13	16705163	16882824	17704332	18264000	16772447	**16333783**
ibm14	31723086	31896196	32622444	32961900	31571929	**31021837**
ibm15	37895733	38524492	40728792	41130600	38579492	**37051784**
ibm16	42596120	44243820	46035824	46588800	43237931	**41769648**
ibm17	**59208028**	59920616	65473152	64615000	63003317	61251040
ibm18	40669287	40676484	42303792	42284100	43298326	**39069479**

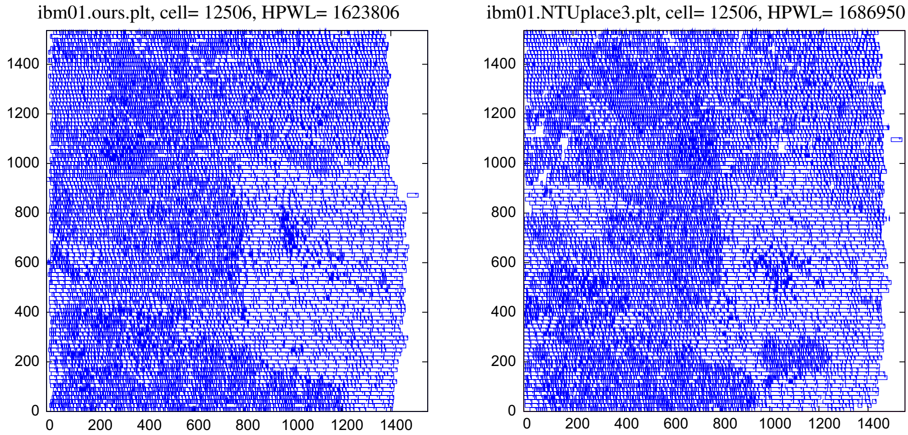

Fig. 1 Comparison of the placement results on ibm01. *Left* the placement result generated by our placement method, HPWL = 1623806. *Right* the placement result generated by NTUplace3, HPWL = 1686950

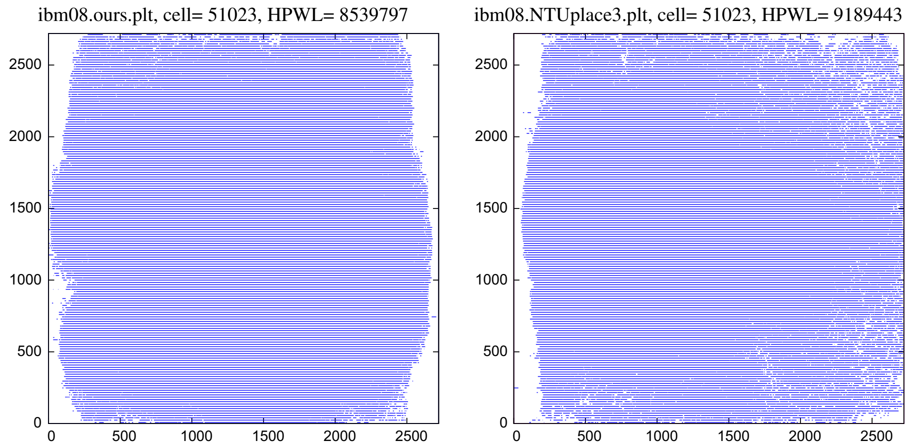

Fig. 2 Comparison of the placement results on ibm08. *Left* the placement result generated by our placement method, HPWL = 8639797. *Right* the placement result generated by NTUplace3, HPWL = 9189443

6 Concluding remarks

The proximal ADMM is very efficient in solving convex optimization problems and monotone variational inequalities, especially with separable structure. However, it is seldom used in nonconvex optimization. In this paper, we considered a special nonconvex minimization problem, in which the objective function is the sum of two convex functions, and the constraints involve Lipschitz nonconvex functions. We proposed a proximal ADMM for solving this problem. Under some assumptions and suitable choice of the proximal parameters, we proved the convergence to a KKT point of the proposed method.

Using the proposed proximal ADMM as a global placement tool, and combining the existing legalization and detailed placement methods of NTUplace3 placer, we designed a new placement method of VLSI. The experimental results on the IBM standard cell benchmarks show that, the new placement method achieves high solution quality compared with some state-of-the-art placement methods. These results indicate that, the proposed method is efficient in solving the encountered nonconvex constrained minimization problem of this paper.

Acknowledgments The authors are very grateful to the editor and the anonymous reviewers for their constructive comments which play a crucial role on the final version of this paper.

References

1. Antipin, A.S.: Gradient approach of computing fixed points of equilibrium problems. J. Glob. Optim. **24**(3), 285–309 (2002)
2. Antipin, A.S.: Extra-proximal methods for solving two-person nonzero-sum games. Math. Program. **120**, 147–177 (2009)
3. Agnihotri, A.R., Ono, S., Li, C., Yildiz, M.C., Khatkhate, A., Koh, C.K., Madden, P.H.: Recursive bisection placement: feng shui 5.0 implementation detail. In: Proceeding of the International Symposium on Physical Design, pp. 230–232 (2005)
4. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
5. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2010)
6. Bertesekas, D.P., Tseng, P.: Partial proximal minimization algorithms for convex programming. SIAM J. Optim. **4**(3), 551–572 (1994)
7. Bertesekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Academic Press. Inc., New York (1996)
8. Chan, T., Cong, J., Shinnerl, J., Sze, K., Xie, M.: mPL6: Enhanced multilevel mixed-sized placement. In: Proceeding of the International Symposium on Physical Design, pp. 212–214 (2006)
9. Chen, C.H., He, B.S., Yuan, X.M.: Matrix completion via alternating direction methods. IMA J. Numer. Anal. **32**, 227–245 (2012)
10. Chen, T.C., Jiang, Z.W., Hsu, T.C., Chen, H.C., Chang, Y.W.: NTUplace3: an analytical placer for large-scale mixed-size design with preplaced blocks and density constraints. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **27**(7), 1228–1240 (2008)
11. Chu, C.: Placement. In: Wang L.T., Chang Y.W., Cheng K.T. (eds.) Electronic Design Automation: Synthesis, Verification, and Testing, chap. 11, pp. 635–684. Elsevier, Morgan Kaufmann (2008)
12. Chang, Y.W., Jiang, Z.W., Chen, T.C.: Essential issues in analytical placement algorithms. Inf. Media Technol. **4**(4), 815–836 (2009)
13. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**(1–3), 293–318 (1992)
14. Eckstein, J.: Some saddle-function splitting methods for convex programming. Optim. Methods Softw. **4**, 75–83 (1994)
15. Eckstein, J., Fukushima, M.: Some reformulations and applications of the alternating direction method of multipliers. In: Hager, W.W., Hearn, D.W., Pardalos, Panos M. (eds.) Large Scale Optimization: State of the Art, pp. 115–134. Kluwer Academic Publishers, Dordrecht (1994)
16. Glowinski, R.: Numerical Methods for Nonlinear Variational Problems. Springer, New York (1984)
17. He, B.S., Liao, L.Z., Han, D.R., Yang, H.: A new inexact alternating directions method for monotone variational inequalities. Math. Program. **92**(1), 103–118 (2002)
18. He, B.S., Liao, L.Z., Qian, M.J.: Alternating projection based prediction-correction methods for structured variational inequalities. J. Comput. Math. **24**, 693–710 (2006)
19. Karamardian, S., Schaible, S.: Seven kinds of monotone maps. J. Optim. Theory Appl. **66**(1), 37–46 (1990)
20. Kaplan, A., Tichatschke, R.: Proximal point methods and nonconvex optimization. J. Glob. Optim. **13**, 389–406 (1998)
21. Martinet, B.: Regularisation d'indquations variationelles par approximations successives. Revue Francaise d'Automatique et Informatique Recherche Opérationelle **4**, 154–159 (1970)
22. Martinet, B.: Determination approch*é*e d'un point fixe d'une application pseudo-contractante. Compte Rendu Academie des Sciences de Paris **274**, 163–165 (1972)
23. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer Science & Business Media, LLC (2006)

24. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
25. Roy, J.A., Papa, D.A., Adya, S.N., Chan, H.H., Ng, A.N., Lu, J.F., Markov, I.L.: Capo: robust and scalable open-source min-cut floorplacer. In: Proceeding of International Symposium on Physical Design, pp. 224–227 (2005)
26. Viswanathan, N., Pan, M., Chu, C.: Fastplace3.0: a fast multilevel quadratic placement algorithm with placement congestion control. In: Proceeding of Asia and South Pacific Design Automation Conference, pp. 135–140 (2007)
27. Xu, M.H.: Proximal alternating directions method for structured variational inequalities. J. Optim. Theory Appl. **134**(1), 107–117 (2009)
28. Yang, J.F., Sun, D.F., Toh, K.C.: A proximal point algorithm for log-determinant optimization with group Lasso regularization. SIAM J. Optim. **23**(2), 857–893 (2013)
29. Yuan, X.M.: An improved proximal alternating direction method for monotone variational inequalities with separable structure. Comput. Optim. Appl. **49**(1), 17–29 (2011)
30. IBM standard cell benchmarks, available online at: http://vlsicad.eecs.umich.edu/BK/Slots/cache/www.public.iastate.edu/nataraj/ISPD04-Bench.html