Optimization Methods and Software
Vol. 00, No. 00, Month 200x, 1–11

On the Local Quadratic Convergence of the Primal-Dual Augmented Lagrangian Method

ROMAN A. POLYAK
Department of SEOR and Mathematical Sciences Department
George Mason University
4400 University Dr, Fairfax VA 22030
rpolyak@gmu.edu

Dedicated to Professor M. Powell on the occasion of his 70th birthday.

(Received 00 Month 200x; In final form 00 Month 200x)

We consider a Primal-Dual Augmented Lagrangian (PDAL) method for optimization problems with equality constraints. Each step of the PDAL requires solving the Primal-Dual linear system of equations. We show that under the standard second-order optimality condition the PDAL method generates a sequence, which locally converges to the primal-dual solution with quadratic rate.

Keywords: Augmented Lagrangian; Primal-dual system; Merit function; Quadratic rate of convergence

2000 Mathematics Subject Classifications: 90C25; 90C26

1 Introduction

The Augmented Lagrangian (AL) and correspondent multipliers method for equality constraints optimization were introduced in the late sixties by M.R. Hestenes [8] and M.J.D. Powell [13]. Over the last almost forty years the AL theory and multipliers method have been extensively studied and currently made up a considerable part of optimization theory (see [1]– [3] and references therein).

The multipliers method at each step alternates the unconstrained minimization of the AL in primal space with Lagrange multipliers update while the penalty parameter can be fixed or updated from step to step. A fixed penalty parameter allows to avoid the ill-conditioning phenomenon typical for penalty methods (see [6]).

Under the standard second-order optimality condition the multipliers method converges with linear rate for any fixed, but large enough penalty parameter. The ratio is inversely proportional to the penalty parameter. Therefore, the superlinear rate of convergence can be achieved by increasing the penalty parameter from step to step. Unfortunately, it leads to ill-conditioning of the AL Hessian, which reduces the size of the area where Newton's method is well-defined (see [9]). It increases the number of damped Newton steps per update and reduces the overall efficiency of the Newton AL method.

In this paper we consider the PDAL, which eliminates the basic drawbacks of the Newton AL method. The PDAL reduces drastically the computational effort per step and at the same time improves substantially the rate of convergence as compared with the Newton AL method.

A step of the Augmented Lagrangian multipliers method is equivalent to solving the Primal-Dual AL nonlinear system of equations, which consists of the optimality criteria for the primal AL minimizer and formulas for the Lagrange multipliers update.

Application of Newton's method to the Primal–Dual AL nonlinear system leads to solving the Primal–Dual linear system of equations for finding the Primal–Dual direction. This linear system happened to be identical to the Primal-Dual system introduced in (see [1] p 240) as a variation of Newton's method for Lagrange system of equations.

It has been pointed out in [1] that if the penalty parameter infinitely grows then a local super linear rate of convergence can be expected, however the way the penalty parameter should be update was not specified in [1].

We specify the penalty parameter update and prove that PDAL method converges with quadratic rate.

Each step of the PDAL method consists of finding the Primal–Dual direction, updating the Primal–Dual approximation following by the penalty parameter update.

The key element of the PDAL method is the merit function, It turned out that by taking the penalty parameter as an inverse of the merit function, we make the primal-dual direction very close to the correspondent direction of the Newton method for the Lagrange system of equations. It allows us proving local quadratic rate of convergence of the PDAL method under the standard second-order optimality conditions.

For other super–linear and quadratically convergent Primal–Dual methods for both equality and inequality constraints (see [4], [5], [10], [12] and references therein).

The paper is organized as follows. In the next section, we state the problem and introduce the basic assumptions. In section 3 we briefly recall the basic AL results. In section 4 we describe the PDAL method and prove its local quadratic rate of convergence. In section 5 we made few comments concerning future research

2 Statement of the problem and basic assumptions

We consider $q+1$ twice continuously differentiable functions $f,\ c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1,\ldots,q$ and the feasible set

$$\Omega = \{x : c_i(x) = 0, \quad i = 1,\ldots,q\}.$$

The problem consists of finding

$$\text{(P)} \qquad f(x^*) = \min\{f(x)|x \in \Omega\}\,.$$

The Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}$ for the problem (P) is given by formula

$$L(x,\lambda) = f(x) - \sum_{i=1}^{q} \lambda_i c_i(x)\,.$$

We consider the vector–function $c^T(x) = (c_1(x),\ldots,c_q(x))$, the Jacobian

$$J(c(x)) = \nabla c(x) = (\nabla c_1(x),\ldots,\nabla c_q(x))^T$$

and assume that

$$\text{rank } \nabla c(x^*) = q < n\,, \tag{1}$$

i.e. the gradients $\nabla c_i(x^*)$, $i = 1,\ldots,q$ are linearly independent. Then there exists a vector of Lagrange multipliers $\lambda^* \in \mathbb{R}^q$, such that the necessary conditions for x^* to be the minimizer are satisfied, i.e.

$$\nabla_x L(x^*,\lambda^*) = \nabla f(x^*) - \sum_{i=1}^{q} \lambda_i^* \nabla c_i(x^*) = 0 \tag{2}$$

$$c_i(x^*) = 0, \quad i = 1,\ldots,q\,. \tag{3}$$

Let us consider the Hessian of the Lagrangian $L(x,\lambda)$.

$$\nabla^2_{xx} L(x,\lambda) = \nabla^2 f(x) - \sum_{i=1}^{q} \lambda_i \nabla^2 c_i(x) \,.$$

The condition (1) together with sufficient condition for the minimizer x^* to be isolated

$$(\nabla^2 L(x^*,\lambda^*)y,y) \geq m(y,y),\ \forall y : \nabla c(x^*)y = 0, m > 0 \tag{4}$$

comprise the standard second-order optimality conditions for the problem (P).

We consider the vector norm $||x|| = \max_{1\leq i\leq n} |x_i|$ and for a given matrix $A \in \mathbb{R}^{n,n}$ the corresponding matrix norm

$$||A|| = \max_{i\leq i\leq n} \left(\sum_{j=1}^{n} |a_{ij}| \right) \,.$$

For a small enough $\epsilon_0 > 0$, we consider the neighbourhood $\Omega_{\epsilon_0}(x^*) = \{x : \|x - x^*\| \leq \epsilon_0\}$ of the primal solution x^*.

We assume that the Lipschitz condition for the Hessians $\nabla^2 f$ and $\nabla^2 c_i$, $i = 1,\dots,q$ is satisfied, i.e. there are $L_0 > 0, L_i > 0, i = 1,\dots,q$ such that

$$||\nabla^2 f(x_1) - \nabla^2 f(x_2)|| \leq L_0 ||x_1 - x_2|| \tag{5}$$

$$||\nabla^2 c_i(x_1) - \nabla^2 c_i(x_2)|| \leq L_i ||x_1 - x_2|| \tag{6}$$

for all $(x_1,x_2) \in \Omega_{\epsilon_0}(x^*) \times \Omega_{\epsilon_0}(x^*)$.

We conclude the section by recalling the Debreu lemma [1], which will be used later.

Lemma 2.1 *Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a symmetric matrix, $B : \mathbb{R}^n \to \mathbb{R}^r$, $m > 0$ and $(Ay,y) \geq m(y,y)$, $\forall y \in \mathbb{R}^n : By = 0$, then there exists $\kappa_0 > 0$ large enough that the inequality*

$$\left(\left(A + \kappa B^T B\right) y, y \right) \geq \mu(y,y) \,,\ \ \forall y \in \mathbb{R}^n$$

holds for $0 < \mu < m$ and any $\kappa \geq \kappa_0$.

3 Augmented Lagrangian and multipliers method

The Augmented Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}_{++} \to \mathbb{R}$ is defined by formula (see [8,13])

$$\mathcal{L}(x,\lambda,k) = f(x) - \sum_{i=1}^{q} \lambda_i c_i(x) + \frac{k}{2} \sum_{i=1}^{q} c_i^2(x) \,. \tag{7}$$

For the primal-dual solution (x^*,λ^*) and any $k > 0$ from (7) follows

$$\text{a)}\ \ \mathcal{L}(x^*,\lambda^*,k) = f(x^*),\ \text{b)}\ \nabla_x \mathcal{L}(x^*,\lambda^*,k) = \nabla_x L(x^*,\lambda^*) = 0 \tag{8}$$

and

$$\nabla^2_{xx} \mathcal{L}(x^*,\lambda^*,k) = \nabla_{xx} L(x^*,\lambda^*) + k \nabla c(x^*)^T \nabla c(x^*) \,. \tag{9}$$

From (9), standard second-order optimality condition (1), (4) and Debreu Lemma 2.1 follows that for $\forall\, k \geq k_0$ the AL Hessian $\nabla^2_{xx}\mathcal{L}(x^*, \lambda^*, k)$ is positive definite, i.e. there is $\mu > 0$:

$$\left(\nabla^2 \mathcal{L}_{xx}(x^*, \lambda^*, k)v, v\right) \geq \mu(v, v)\,, \quad \forall\, v \in \mathbb{R}^n\,. \tag{10}$$

Under (5)–(6) there exists a small enough $0 < \delta < \epsilon_0$ that the type (10) inequality holds for any $y \in \Omega_\delta(y^*) = \{y = (x, \lambda)\,:\, ||y - y^*|| \leq \delta\}$.

Without restricting the generality we can assume that

$$\inf_{x \in \mathbb{R}^n} f(x) \geq 0\,. \tag{11}$$

Otherwise it is sufficient to replace $f(x)$ by

$$f(x) := \ln\left(1 + e^{f(x)}\right)\,.$$

It follows from (8) and (10) that x^* is a local minimizer of $\mathcal{L}(x, \lambda^*, k)$ in the neighborhood $\Omega_{\epsilon_0}(x^*)$, i.e.

$$\mathcal{L}(x^*, \lambda^*, k) \leq \mathcal{L}(x, \lambda^*, k)\,, \quad \forall\, x \in \Omega_{\epsilon_0}(x^*)\,. \tag{12}$$

Let us rewrite $\mathcal{L}(x, \lambda^*, k)$ as follows:

$$\mathcal{L}(x, \lambda^*, k) = f(x) + \frac{k}{2}\sum_{i=1}^{q}\left(c_i(x) - \frac{\lambda_i^*}{k}\right)^2 - \frac{1}{2k}\sum_{i=1}^{q}\lambda_i^{*2}\,.$$

Outside of $\Omega_{\epsilon_0}(x^*)$, the second term can be made as large as one wants by increasing $k > 0$. Therefore, keeping in mind (11), we can find a large enough $k_0 > 0$ that the inequality (12) can be extended on the entire $\mathbb{R}^n$, i.e.

$$x^* = \arg\min\{\mathcal{L}(x, \lambda^*, k) | x \in \mathbb{R}^n\} \tag{13}$$

for any $k \geq k_0$.

Due to (5)–(6) the AL $\mathcal{L}(x, \lambda^*, k)$ is strongly convex in the neighbourhood $\Omega_{\epsilon_0}(x^*)$.

Moreover, it remains true for any pair (λ, k) from the extended dual set

$$D_\alpha(\lambda^*, k_0) = \{(\lambda, k) : |\lambda_i - \lambda_i^*| \leq \alpha k, \quad k \geq k_0\}$$

where $\alpha > 0$ is small enough and $k_0 > 0$ is large enough. In other words, for any $(\lambda, k) \in D_\alpha(\lambda^*, k_0)$, there exists a unique minimizer

$$\hat{x} \equiv \hat{x}(\lambda, k) = \arg\min\{\mathcal{L}(x, \lambda, k) | x \in \mathbb{R}^n\} \tag{14}$$

and the AL $\mathcal{L}(x, \lambda, k)$ is strongly convex in x in the neighborhood of $\hat{x}$.

Each step of the AL method maps $(\lambda, k) \in D_\alpha(\lambda^*, k_0)$ into a pair $\hat{y} = (\hat{x}, \hat{\lambda})$ given by the following formulas

$$\hat{x} \equiv \hat{x}(\lambda, k) = \arg\min\{\mathcal{L}(x, \lambda, k) | x \in \mathbb{R}^n\} \tag{15}$$

$$\hat{\lambda} = \hat{\lambda}(\lambda, k) = \lambda - kc(\hat{x})\,. \tag{16}$$

It follows from (15)–(16) that for any $(\lambda, k) \in D_\alpha(\lambda^*, k_0)$ we have

$$\nabla_x \mathcal{L}(\hat{x}, \lambda, k) = \nabla f(\hat{x}) - \sum (\lambda_i - k c_i(\hat{x})) \nabla c_i(\hat{x}) = \nabla_x L\left(\hat{x}, \hat{\lambda}\right) = 0 \tag{17}$$

The AL method (15)–(16) is equivalent to the dual quadratic prox method (see [14]).

$$\hat{\lambda} = \arg\max \left\{ d(u) - \frac{1}{2k} ||u - \lambda||^2 | u \in \mathbb{R}^q \right\}. \tag{18}$$

We would like to emphasize however that neither the primal sequence generated by (15) nor the dual sequence generated by (18) provide sufficient information for the analysis of the AL multipliers method.

Only the Primal-Dual system (16)–(17) provides such information for proving the basic AL results, which are stated in the following Theorem [1].

Theorem 3.1 *If $f, c_i \in \mathcal{C}^2$, $i = 1, \ldots, q$ and the standard second-order optimality conditions (1), (4) are satisfied, then for any fixed pair $(\lambda, k) \in D_\alpha(\lambda^*, k_0)$*

(i) there exist $\hat{x}$ and $\hat{\lambda}$ given by formulas (15)–(16);
(ii) the following bound holds

$$\max \left\{ ||\hat{x} - x^*||, ||\hat{\lambda} - \lambda^*|| \right\} \leq \frac{\sigma}{k} ||\lambda - \lambda^*|| \tag{19}$$

and $\sigma > 0$ is independent of $k \geq k_0$;
(iii) the AL $\mathcal{L}(x, \lambda, k)$ is strongly convex in x in the neighborhood of $\hat{x} = \hat{x}(\lambda, k)$.

The PD system (16)–(17) plays also a critical role in the PDAL method, which we consider in the following section.

Meanwhile using the arguments similar to those which we used for proving

$$x^* = \arg\min\{\mathcal{L}(x, \lambda^*, k) \mid x \in \mathbb{R}^n\},$$

0ne can show that for $k_0 > 0$ large enough and any $k \geq k_0$ the local minimizer $\hat{x} \equiv \hat{x}(\lambda, k)$ is, in fact, a unique global minimizer of $\mathcal{L}(x, \lambda, k)$ in x, i.e.

$$\hat{x} \equiv \hat{x}(\lambda, k) = \arg\min\{\mathcal{L}(x, \lambda, k) \mid x \in \mathbb{R}^n\}.$$

We conclude this section by pointing out that a vector $\bar{y} = (\bar{x}, \bar{\lambda})$ which satisfies the Lagrange system of equations

$$\nabla_x L(x, \lambda) = 0, \quad c(x) = 0 \tag{20}$$

is not necessarily a primal-dual solution. In particular, $\bar{x}$ can be a local or global maximum of f on Ω. However, the following remark holds true.

Remark 1 If the standard optimality conditions are satisfied and $k_0 > 0$ is large enough, then any pair $\hat{y} = (\hat{x}, \hat{\lambda})$:

$$\hat{y} \in \hat{Y} = \{\hat{y} = (\hat{x}, \hat{\lambda}) = (\hat{x}(\lambda, k), \hat{\lambda}(\lambda, k) : (\lambda, k) \in D_\alpha(\lambda^*, k_0)\},$$

which satisfies the Lagrange system (20), is the primal-dual solution, i.e. if $\hat{x}(\lambda, k) = \hat{x}$ and $\hat{\lambda}(\lambda, k) = \hat{\lambda}$ satisfy (20) then $\hat{\lambda} = \lambda^*$ and

$$\hat{x}(\lambda^*, k) = x^*, \hat{\lambda}(\lambda^*, k) = \lambda^*, \quad \forall\, k > k_0.$$

4 Primal-Dual Augmented Lagrangian method

In this section we describe a PDAL method and show its local convergence with quadratic rate. The PDAL method requires at each step solving one primal-dual linear system of the equation following the penalty parameter update. The key element of the PDAL method is the penalty parameter, which is taken as an inverse to the merit function.From this point on we assume $\Omega_\delta(y^*) := \Omega_\delta(y^*) \cap \hat{Y}$.

The merit function $\nu : \Omega_\delta(y^*) \to \mathbb{R}_+$, we define by the following formula

$$\nu(y) = \max\{||\nabla_x L(x,\lambda)||, ||c(x)||\}\,. \tag{21}$$

It follows from (21) that $\nu(y) \geq 0$. Assuming that the standard second order optimality conditions (1), (4) are satisfied and keeping in mind Remark *1*, we obtain

$$\nu(y) = 0 \Leftrightarrow y = y^*. \tag{22}$$

On the one hand $\nu(y)$ is measuring the distance from $y \in \Omega_\delta(y^*)$ to the primal-dual solution y^*.

On the other hand the merit function $\nu(y)$ will be used for controlling the penalty parameter $k > 0$.

Let us consider the Primal–Dual AL system

$$\nabla_x L(\hat{x},\hat{\lambda}) = \nabla f(\hat{x}) - \sum_{i=1}^{q} \hat{\lambda}_i \nabla c_i(\hat{x}) = 0, \tag{23}$$

$$\hat{\lambda} = \lambda - kc(\hat{x})\,. \tag{24}$$

Solving the PD system (23)–(24) for $(\hat{x},\hat{\lambda})$ under fixed $\lambda \in \mathbb{R}^q$ and $k \geq k_0 > 0$ is equivalent to one step of the AL method (15)–(16).

Application of the Newton method for solving the PD system (23)–(24) for $(\hat{x},\hat{\lambda})$ using $y = (x,\lambda) \in \Omega_\delta(y^*)$ as a starting point leads to finding Primal–Dual direction.

By linearizing (23)–(24) at $y = (x,\lambda)$ we obtain

$$\nabla f(x) + \nabla^2 f(x)\nabla x - \sum (\lambda_i + \triangle\lambda_i)\left(\nabla c_i(x) + \nabla^2 c_i(x)\triangle x\right) = 0, \tag{25}$$

$$\lambda + \triangle\lambda = \lambda - k\,(c(x) + \nabla c(x)\triangle x)\,. \tag{26}$$

Ignoring terms of the second and higher orders we can rewrite the system (25)–(26) as follows

$$M_k(x,\lambda)\triangle y = \begin{pmatrix} \nabla^2_{xx}L(x,\lambda) & -\nabla c(x)^T \\ \nabla c(x) & k^{-1}I^q \end{pmatrix} \begin{pmatrix} \triangle x \\ \triangle\lambda \end{pmatrix} = \begin{pmatrix} -\nabla_x L(x,\lambda) \\ -c(x) \end{pmatrix} = a(y) \tag{27}$$

which is exactly the system (35)–(36)([1] p 240).

Let $0 < \delta < \epsilon_0$ be small enough and a pair $(y,k) : y \in \Omega_\delta(y^*)$, $y \neq y^*$ and $k = (\upsilon(y))^{-1} \geq k_0$ be a starting point.

The PDAL method consists of the following operations.

1. Find the primal-dual direction $\triangle y = (\triangle x, \triangle\lambda)$ from the system (27) if $\triangle y = O$ then stop.
2. Find the new primal-dual vector $\hat{y} = (\hat{x},\, \hat{\lambda})$:

$$\hat{x} = x + \triangle x, \hat{\lambda} = \lambda + \triangle\lambda \tag{28}$$

3. Update the scaling parameter

$$\hat{k} := (\nu(\hat{y}))^{-1}. \tag{29}$$

4. Set

$$y := \hat{y},\ k := \hat{k} \text{ and go to (1)}. \tag{30}$$

The PDAL method generates a primal–dual sequence $\{y^s = (x^s, \lambda^s)\}_{s=1}^{\infty}$.

The following technical lemmas(see [7]) will be used for proving the local convergence of the primal–dual sequence $\{y^s\}_{s=1}^{\infty}$ to the primal–dual solution $y^* = (x^*, \lambda^*)$ with quadratic rate.

Lemma 4.1 *If the standard second-order optimality conditions (1), (4) and Lipschitz conditions (5)–(6) are satisfied, then there exists small enough* $0 < \delta < \epsilon_0$ *and* $0 < l < L < \infty$ *such that for any* $y \in \Omega_\delta(y^*)$ *the following bounds hold*

$$l\|y - y^*\| \le \nu(y) \le L\|y - y^*\|. \tag{31}$$

Proof. The right inequality follows from $\nu(y^*) = 0$, boundedness of $\Omega_\delta(y^*)$, and Lipschitz condition for the gradients $\nabla f, \nabla c_i$, $i = 1, \ldots, q$.

On the other hand

$$\|\nabla_x L(x, \lambda)\| \le \nu(y),\ \|c(x)\| \le \nu(y). \tag{32}$$

By linearizing the Lagrange system (20) at $x = x^*$, $\lambda = \lambda^*$ and keeping in mind (2)–(3) we obtain the following system

$$\begin{pmatrix} \nabla^2_{xx} L(x^*, \lambda^*) & -\nabla c(x^*)^T \\ \nabla c(x^*) & O \end{pmatrix} \begin{pmatrix} x - x^* \\ \lambda - \lambda^* \end{pmatrix} = M_\infty(x^*, \lambda^*)(y - y^*)$$
$$= \begin{pmatrix} \nabla_x L(x, \lambda) + o(\|x - x^*\|)e_n \\ c(x) + o(\|x - x^*\|)e_q \end{pmatrix} \tag{33}$$

where $e_r = (1, \ldots, 1) \in \mathbb{R}^r$. □

It follows from (1) and (4) that the matrix $M_\infty(x^*, \lambda^*)$ is nonsingular and there is $M_0 > 0$ that

$$\|M_\infty^{-1}(x^*, \lambda^*)\| \le M_0. \tag{34}$$

From (33) we obtain

$$\begin{pmatrix} x - x^* \\ \lambda - \lambda^* \end{pmatrix} = M_\infty^{-1}(x^*, \lambda^*) \begin{pmatrix} \nabla_x L(x, \lambda) + o(||x - x^*||)e_n \\ c(x) + o(\|x - x^*\|)e_q \end{pmatrix}$$

Keeping in mind (34) we have

$$||y - y^*|| \le M_0 \nu(y) + o(\|x - x^*\|).$$

Therefore $\nu(y) \ge l||y - y^*||$, where $l = (2M_0)^{-1}$.

It follows from (31) that the merit function $\nu(y)$ in the neighbourhood $\Omega_\delta(y^*)$ behaves similar to the norm of a gradient of a strongly convex and smooth enough function in the neighbourhood of the minimum.

Lemma 4.2 *Let* $A \in \mathbb{R}^{n,n}$ *be a nonsingular matrix and* $||A^{-1}|| \le N$*; then there exists small enough* $\beta > 0$ *that any matrix* $B \in \mathbb{R}^{n,n} : ||A - B|| \le \beta$ *is nonsingular and the following bound holds*

$$a)\ ||B^{-1}|| \le 2N,\ b)\ ||A^{-1} - B^{-1}|| \le 2N^2\beta. \tag{35}$$

The proof can be found in [7].

Lemma 4.3 *If the standard second order optimality condition (1),(4) are satisfied, then the matrices*

$$M_\infty(x,\lambda) = \begin{pmatrix} \nabla^2 L(x,\lambda) & -\nabla c(x)^T \\ \nabla c(x) & O \end{pmatrix} \quad \text{and} \quad M_k(x,\lambda) = \begin{pmatrix} \nabla^2 L(x,\lambda) & -\nabla c(x)^T \\ \nabla c(x) & k^{-1} I \end{pmatrix}$$

are nonsingular and the following bound holds

$$\max\{||M_\infty^{-1}(y)||, ||M_k^{-1}(y)||\} \le 2N \tag{36}$$

for any pair $(y,k) : y \in \Omega_\delta(y^*)$ and $k \ge k_0$.

Proof. The nonsingularity of

$$M_\infty(x^*,\lambda^*) = \begin{bmatrix} \nabla^2_{xx} L(x^*,\lambda^*) & -\nabla c(x^*)^T \\ \nabla c(x^*) & 0 \end{bmatrix}$$

follows from the standard second-order optimality condition (1), (4). Therefore, there exists $N > 0$ that $\|M_\infty^{-1}(y^*)\| \le N$.

Using Lemma 4.2 and Lipschitz condition (5)–(6) we can find a small enough $0 < \delta < \epsilon_0$ such that the following bound holds

$$\max\left\{\left\|M_\infty^{-1}(x,\lambda)\right\|\ ,\ \left\|M_k^{-1}(x,\lambda)\right\|\right\} \le 2N\ ,\ \forall\, y \in \Omega_\delta(y^*) \quad \texttt{and} \quad \forall k \ge k_0\,.$$

Now we are ready to prove convergence of the PDAL sequence $\{y^s\}_{s=1}^\infty$ to the primal–dual solutions $y^* = (x^*,\lambda^*)$ and establish the quadratic convergence rate.

Theorem 4.4 *If the standard second-order optimality condition (1), (4) and the Lipschitz conditions (5), (6) are satisfied, then*

(i) there exist a small enough $0 < \delta < \epsilon_0$ such that for any starting pair $(y,k) : y \in \Omega_\delta(y^*)$ and $k = (\upsilon(y))^{-1} \ge k_0$ the PDAL step (27)–(30) generates an approximation $\hat{y} = (\hat{x},\hat{\lambda})$ that the following bound holds

$$||\hat{y} - y^*|| \le \rho ||y - y^*||^2\,. \tag{37}$$

and $\rho > 0$ is independent on $y \in \Omega_\delta(y^*)$ and $k = (\upsilon(y))^{-1}$.

(ii) for a given $0 < q_0 < 1$ there exists a small enough $0 < \delta_0 < \min\{\delta, \rho^{-1}\min\{q_0, l^{-1}L\}\}$ such that for any $y \in \Omega_{\delta_0}(y^*)$ and $k = (\upsilon(y))^{-1} > k_0$ we have $\hat{y} \in \Omega_{\delta_0}(y^*)$, $\hat{k} = (\upsilon(\hat{y}))^{-1} > k$ and the following bound hold

$$||\hat{y} - y^*|| \le q_0 ||y - y^*||\,. \tag{38}$$

(iii) for a given $0 < q < 1$ there is small enough $0 < \gamma < \delta_0$ such that for any starting point $y \in \Omega_\gamma(y^*)$ and $k = \upsilon(y))^{-1}$

the primal-dual sequence $\{y^s\}_{s=1}^\infty$ converges to y^* and the following bond holds

$$||y^s - y^*|| \le \rho^{-1} q^{2^s}, \quad s \ge 1 \tag{39}$$

Proof.
(i) The primal-dual system (27) has a unique solution, for any $y \in \Omega_\delta(y^*)$ and $k = \upsilon(y))^{-1} \ge k_0$ as soon as $\delta > 0$ is small enough and k_0 is large enough. We will show later that there exist $0 < \delta_0 < \delta$ that for any starting point $y \in \Omega_{\delta_0}(y^*)$, the entire primal-dual sequence $\{y^s\}_{s=1}^\infty$ belongs to $\Omega_{\delta_0}(y^*)$, therefore the PDAL method is well defined. The key ingredient of the convergence proof is that PD direction $\Delta y = (\Delta x, \Delta x)$ is

close to the Newton direction for Lagrange system (20) at any starting point $y \in \Omega_\delta(y^*)$ and $k = (\upsilon(y))^{-1}$ as soon as $0 < \delta < \epsilon_0$ is small enough.

We consider the Lagrange system (20) at the same starting point $y = (x, \lambda) \in \Omega_\delta(y^*)$. Let $\overline{\triangle y} = (\overline{\triangle x}, \overline{\triangle\lambda})$ be the Newton direction obtained by solving the following system of linear equations

$$M_\infty(y)\overline{\triangle y} = a(y)\,.$$

The new approximation $\bar{y} = (\bar{x}, \bar{\lambda})$ for the Lagrange system (20) is

$$\begin{aligned}&\bar{x} = x + \overline{\triangle x}, \bar{\lambda} = \lambda + \overline{\triangle\lambda}, \quad \text{or}\\ &\bar{y} = y + \overline{\triangle y}\,.\end{aligned}$$

Under the standard second-order optimality condition (1), (4) and Lipschitz conditions (5)–(6) there exists $\rho_1 > 0$ independent of $y \in \Omega_\delta(y^*)$ such that the following bound holds (see [10, Theorem 9, Ch 8])

$$||\bar{y} - y^*|| \le \rho_1 ||y - y^*||^2\,. \tag{40}$$

Let us prove a similar bound for the PDAL approximation $\hat{y}$. We have

$$\begin{aligned}||\hat{y} - y^*|| &= ||y + \triangle y - y^*|| = ||y + \overline{\triangle y} + \triangle y - \overline{\triangle y} - y^*||\\ &\le ||\bar{y} - y^*|| + ||\triangle y - \overline{\triangle y}||\,.\end{aligned}$$

For $||\triangle y - \overline{\triangle y}||$ we obtain

$$\begin{aligned}||\triangle y - \overline{\triangle y}|| &= ||(M_k^{-1}(y) - M_\infty^{-1}(y))a(y)||\\ &\le ||(M_k^{-1}(y) - M_\infty^{-1}(y))||||a(y)||\,.\end{aligned} \tag{41}$$

Using (35b) and keeping in mind

$$||(M_k(y) - M_\infty(y))|| = k^{-1}$$

from (41) we obtain

$$||\triangle y - \overline{\triangle y}|| \le 2k^{-1}N^2||a(y)||\,. \tag{42}$$

In view of $\nabla_x L(x^*, \lambda^*) = 0$, $c(x^*) = 0$, Lipschitz condition (5), (6) and $0 < \delta < \epsilon_0$ there exists $L_0 > 0$ such that

$$||a(y)|| \le L_0||y - y^*||, \quad \forall y \in \Omega_\delta(y^*)\,. \tag{43}$$

Using the right inequality (31) and (28), (42), (43) we obtain

$$\begin{aligned}||\triangle y - \overline{\triangle y}|| &\le 2N^2 L_0 \nu(y)||y - y^*||\\ &\le 2N^2 L L_0||y - y^*||^2\\ &= \rho_2||y - y^*||^2\,.\end{aligned} \tag{44}$$

From (40) and (44) we have

$$||\hat{y} - y^*|| \le ||\bar{y} - y^*|| + ||\triangle y - \triangle\bar{y}|| \le \rho||y - y^*||^2, \tag{45}$$

where $\rho = 2\max\{\rho_1, \rho_2\}$ is independent on $y \in \Omega_\delta(y^*)$ and $k \geq k_0$.
(ii) For a given $0 < q_0 < 1$ we can find $\delta_0 < min\{\delta, q_0\rho^{-1}\}$ such that from (45)we obtain

$$||\hat{y} - y^*|| \leq \rho||y - y^*||^2 \leq \rho\delta_0||y - y^*|| \leq q_0||y - y^*||, \forall y \in \Omega_{\delta_0}(y^*)$$

Also from the right inequality (31) and (45) follows

$$\nu(\hat{y}) \leq L||\hat{y} - y^*|| \leq L\rho||y - y^*||^2 \leq L\rho\delta_0||y - y^*||, \forall y \in \Omega_{\delta_0}(y^*).$$

Using the left inequality (31) we obtain $\nu(\hat{y}) \leq Ll^{-1}\rho\delta_0\nu(y)$, therefore for $\delta_0 < \rho^{-1}lL^{-1}$ we obtain

$$\hat{k} = (\nu(\hat{y}))^{-1} > (\nu(y))^{-1} = k.$$

In other words, there is small enough $0 < \delta_0 < \min\{\delta, \rho^{-1}\min\{\gamma, \ell L^{-1}\}\}$ and large enough k_0 that for any $y \in \Omega_{\delta_0}(y^*)$ and $k = (\upsilon(y))^{-1} \geq k_0$ we have $\hat{y} \in \Omega_{\delta_0}(y^*), \hat{k} = (\upsilon(\hat{y}))^{-1} > k$ and bound (38) holds.

Therefore for any $y \in \Omega_{\delta_0}(y^*)$ and $k = (\nu(y))^{-1} \geq k_0$ as a starting pair, the primal-dual sequence $\{y^s\}_{s=1}^{\infty} \subset \Omega_{\delta_0}(y^*)$ and converges to y^*.
(iii) It follows from (37) that for $y \in \Omega_{\delta_0}(y^*)$ and $k = (\upsilon(y))^{-1}$ as a starting point we have

$$||y^{s+1} - y^*|| \leq \rho||y^s - y^*||^2 \leq \rho^{-1}\left(\rho||y^s - y^*||\right)^2, \quad s \geq 1. \tag{46}$$

For a given $0 < q < 1$ we can find $0 < \gamma < \min\{\delta_0, \rho^{-1}q\}$ such that for any $y \in \Omega_\gamma(y^*)$ we have $\rho||y - y^*|| < q$. By iterating (46) we obtain

$$||y^{s+1} - y^*|| \leq \rho^{-1}q^{2^s}.$$

□

5 Concluding remarks

The basic operation at each step of the PDAL method consists of solving one linear system of equations (27), whereas the basic operation of the AL method step requires solving an unconstrained optimization problem (15). On the other hand, the PDAL sequence locally converges to the primal-dual solution with quadratic rate (see Theorem 4.4), while the AL sequence converges with Q-linear rate (seeTheorem 3.1).

The local PDAL method can be extended into a globally convergent primal-dual AL method for convex optimization in the framework of the nonlinear rescaling theory with dynamic scaling parameters update (see [11, 12]). We are going to cover these results in the upcoming paper.

Acknowledgement. The author would like to thank the Associate Editor and the referees for their useful suggestions.
The research was supported by NSF Grant CCF–0836338.

References

[1] Bertsekas, D., 1982. Constrained Optimization and Lagrange Multipliers Methods (New York: Academic Press).
[2] Bertsekas, D., 1999. Nonlinear Programming, Second Edition Constrained Optimization and Lagrange Multipliers Methods (New York: Academic Press).
[3] DiPillo, G., G. Grippo L., 1982. An Augumented Lagrangian for inequality constraints in nonlinear programming problems *J. Optim. Theory Appl.*, **36**, 495–519.
[4] DiPillo, G., Lucidi, S. and Palagi, L., 2000. A superlinearly convergent primal–dual algorithm model for unconstrained optimization problems with bounded variable *Optimization Methods and Software*, **14**, 49–73.
[5] Facchinei, F., and Lucidi, S., 1995. Quadratically and superlinearly convergent algorithms for the solution of inequality constrained minimization problems *J. Optim. Theory Appl.*, **85**, 265–289.
[6] Fiacco, A. V. and McCormick, G. P., 1990. Nonlinear Programming: Sequential Unconstrained Minimization Techniques, Classes in Applied Optimization (Philadelphia, PA: SIAM).

[7] Griva, I. and Polyak, R., 2006. Primal-dual nonlinear rescaling method with dynamic scaling parameter update. *Mathematical Programming*, **106** (2), 237–259.
[8] Hestenes, M. R., 1969. Multipliers and gradient method. *J. Optim. Theory Appl.*, **4**, 303–320.
[9] Nesterov, Y., 2004. Introductory Lectures in Convex Optimization (Boston/Dordrecht/London: Kluwer Academic Publishers).
[10] Polyak, B. T., 1987. Introduction to Optimization (New York: Software Inc).
[11] Polyak, R., 2006. Nonlinear rescaling as interior quadratic prox method for convex optimization. *Computational Optimization and Applications*, **35** (3), 347–373.
[12] Polyak, R., 2008. Primal-dual exterior point method for convex optimization. *Optimization Methods and Software*, **23** (1), 141–160.
[13] Powell, M. J. D., 1969. A method for nonlinear constraints in minimization problems. In: R. Fletcher (Ed) *Optimization* (London Academic Press), pp. 283–298.
[14] Rockafellar, R. T., 1976. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, **1**, 97–116.