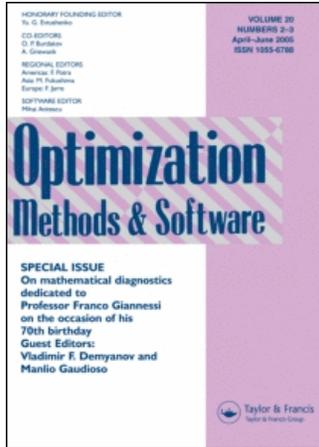


This article was downloaded by:[George Mason University]  
On: 21 December 2007  
Access Details: [subscription number 768416529]  
Publisher: Taylor & Francis  
Informa Ltd Registered in England and Wales Registered Number: 1072954  
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Optimization Methods and Software

Publication details, including instructions for authors and subscription information:  
<http://www.informaworld.com/smpp/title~content=t713645924>

### Primal-dual exterior point method for convex optimization

Roman A. Polyak<sup>a</sup>

<sup>a</sup> Department of SEOR and Mathematical Sciences Department, George Mason University, Fairfax, Virginia, USA

First Published on: 24 April 2007

To cite this Article: Polyak, Roman A. (2007) 'Primal-dual exterior point method for convex optimization', Optimization Methods and Software, 23:1, 141 - 160

To link to this article: DOI: 10.1080/10556780701363065

URL: <http://dx.doi.org/10.1080/10556780701363065>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## Primal–dual exterior point method for convex optimization

Roman A. Polyak\*

*Department of SEOR and Mathematical Sciences Department,  
George Mason University, Fairfax, Virginia, USA*

*(Received 5 June 2006; final version received 19 March 2007)*

We introduce and study the primal–dual exterior point (PDEP) method for convex optimization problems. The PDEP is based on the non-linear rescaling (NR) multipliers method with dynamic scaling parameters update. The NR method at each step alternates finding the unconstrained minimizer of the Lagrangian for the equivalent problem with both Lagrange multipliers and scaling parameters vectors update. The NR step is replaced by solving the primal–dual (PD) system of equations. The application of the Newton method to the PD system leads to the PDEP method.

We show that under the standard second-order optimality condition, the PDEP method generates a PD sequence, which globally converges to the PD solution with asymptotic quadratic rate.

**Keywords:** nonlinear rescaling; duality; interior quadratic prox; primal–dual exterior point method; quadratic convergence rate

### 1. Introduction

The non-linear rescaling (NR) method with dynamic scaling parameters update was first introduced for exponential transformation in ref. [20].

During the last decade, the NR method and its dual equivalent, Interior Prox, with second-order  $\varphi$ -divergence distance have been extensively studied, and a number of convergence results under various assumptions on the input data were obtained (see [2,4,15,16] and references therein).

The NR method alternates finding the primal minimizer of the Lagrangian for the equivalent problem with both Lagrange multipliers and scaling parameters update while the penalty parameter is fixed. Under the standard second-order optimality condition, the NR method converges with Q-linear rate if the penalty parameter is fixed but large enough.

To improve the rate of convergence, one has to increase the penalty parameter from step to step. This allows achieving Q-superlinear rate, but it leads to the ill-conditioned Hessian of the minimized function and significantly increases from step to step the computational effort for finding the approximation for the primal minimizer.

---

\*Email: [rpolyak@gmu.edu](mailto:rpolyak@gmu.edu)

The purpose of this paper is to introduce and analyse the primal–dual exterior point (PDEP) method which allows, to a large extent, elimination of the basic drawbacks of the original NR technique.

Each step of the NR method is equivalent to solving the primal–dual (PD) non-linear system of equations. Recently, the corresponding PD systems were used for developing globally convergent primal–dual NR methods with up to 1.5-Q superlinear rate (see [8,17] and references therein).

In this paper, we use specific properties of the PD system related to the NR method with dynamic scaling parameters update for developing the PDEP method.

The PDEP method requires solving, at each step, a linear PD system of equations and generates a PD sequence, which under the standard second-order optimality conditions converges to the PD solution from any starting point with asymptotic quadratic rate. This is our main contribution.

The PDEP method does not require finding the primal minimizer at each step, allows the unbounded increase of the penalty parameter without compromising both the accuracy and computational effort per step and it is free from any stringent conditions for accepting the Newton step, which are typical for constrained optimization problems.

There are three important features that make the PDEP method free from such restrictions.

First, the Lagrangian for the equivalent problem is defined on the entire primal space.

Second, after a few Lagrange multipliers updates, the terms of the Lagrangian for the equivalent problem corresponding to the passive constraints become negligibly small due to their super-quadratic convergence to zero. Therefore, on the one hand, these terms become irrelevant for finding the Newton direction. On the other hand, there is no need to enforce their non-negativity.

Third, the NR method is an exterior point method in the primal space. Therefore, there is no need to enforce the non-negativity of the slack variables for the active constraints, as it takes place in interior point methods (see [19]).

Due to the super-quadratic convergence to zero of the Lagrange multipliers for the passive constraints after very few Lagrange multipliers updates, the PD direction becomes practically identical to the Newton direction for the Lagrange system of equations corresponding to the active constraints.

From this point onwards, both primal and dual approximation are, generally speaking, infeasible and we are dealing with the PDEP method. At the same time, the PDEP method turns out to be very close to the Newton method for solving the Lagrange system of equations for the active constraints. It makes possible to prove the asymptotic quadratic rate of the PDEP method.

The paper is organized as follows. In the next section, we state the problem and introduce the basic assumptions on the input data. In Section 3, we recall the basic facts of the general NR method and describe some convergence results, which will be used later. In Section 4, we introduce the PDEP method and prove its local quadratic convergence under standard second-order optimality conditions. In Section 5, we consider the globally convergent PDEP method and show that the PD sequence converges to the PD solution with asymptotic quadratic rate. We conclude the paper with some remarks concerning future research.

## 2. Statement of the problem and basic assumptions

Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^1$  be convex and all  $c_i : \mathbb{R}^n \rightarrow \mathbb{R}^1, i = 1, \dots, q$  are concave and smooth functions. We consider the following convex optimization problem:

$$x^* \in X^* = \operatorname{argmin}\{f(x) \mid x \in \Omega\}, \quad (\text{P})$$

where  $\Omega = \{x : c_i(x) \geq 0, i = 1, \dots, q\}$ . We assume that:

**A:** The optimal set  $X^*$  is non-empty and bounded.

**B:** Slater's condition holds, i.e., there exists

$$\hat{x} : c_i(\hat{x}) > 0, \quad i = 1, \dots, q.$$

Let us consider the Lagrangian  $L(x, \lambda) = f(x) - \sum_{i=1}^q \lambda_i c_i(x)$ , the dual function

$$d(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda)$$

and the dual problem

$$\lambda \in L^* = \operatorname{argmax}\{d(\lambda) \mid \lambda \in \mathbb{R}_+^q\}. \quad (\text{D})$$

Due to assumption **B**, the optimal dual solution set  $L^*$  is bounded, and for any  $\lambda^* = (\lambda_1^*, \dots, \lambda_q^*) \in L^*$ , we have

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

and the complementary slackness conditions

$$\lambda_i^* c_i(x^*) = 0, \quad i = 1, \dots, q \quad (2)$$

are satisfied, i.e. the Karush–Kuhn–Tucker (KKT) conditions hold true.

Let  $I^* = \{i : c_i(x^*) = 0\} = \{1, \dots, r\}$ ,  $r < n$ , be the active constraints set. We consider the vector functions  $c^T(x) = (c_1(x), \dots, c_q(x))$ ,  $c_{(r)}^T(x) = (c_1(x), \dots, c_r(x))$  and their Jacobians  $\nabla c(x) = J(c(x))$  and  $\nabla c_{(r)}(x) = J(c_{(r)}(x))$ .

The sufficient regularity condition

$$\operatorname{rank} \nabla c_{(r)}(x^*) = r, \quad \lambda_i^* > 0, \quad i \in I^* \quad (3)$$

together with the sufficient condition for the minimum  $x^*$  to be isolated

$$(\nabla_{xx}^2 L(x^*, \lambda^*)y, y) \geq \mu(y, y), \quad \mu > 0 \quad \forall y \neq 0 : \nabla c_{(r)}(x^*)y = 0 \quad (4)$$

comprise the standard second-order optimality conditions, which guarantee the uniqueness of the primal and dual solution.

Let  $N \gg f(x^*)$  be large enough, then by adding one constraint  $c_0(x) = N - f(x) \geq 0$  to the given set of constraints  $c_i(x) \geq 0$ ,  $i = 1, \dots, q$ , we obtain a problem equivalent to the initial one. Obviously, the extra constraint does not affect the solution set  $X^*$  and the correspondent Lagrange multiplier  $\lambda_0^* = 0$ .

On the other hand, the equivalent problem has a bounded feasible set due to the boundedness of  $X^*$  and Corollary 20 in ref. [6]. Therefore, in the following, we assume that  $\Omega$  is bounded.

### 3. Non-linear rescaling method

We consider a class  $\Psi$  of twice continuous differentiable functions  $\psi : (-\infty, \infty) \rightarrow \mathbb{R}$  with the following properties (see [15,16]):

- 1<sup>0</sup>.  $\psi(0) = 0$ ;
- 2<sup>0</sup>. (a)  $\psi'(t) > 0$ ;
- (b)  $\psi'(0) = 1$ ;

- (c)  $\psi'(t) \leq at^{-1}$ ;
- (d)  $|\psi''(t)| \leq bt^{-2} \forall t \in [1, \infty), a > 0, b > 0$ ;
- 3<sup>0</sup>.  $-m^{-1} \leq \psi''(t) < 0 \forall t \in (-\infty, \infty)$ ;
- 4<sup>0</sup>.  $\psi''(t) \leq -M^{-1} \forall t \in (-\infty, 0]$  and  $0 < m < M < \infty$ ;
- 5<sup>0</sup>.  $-\psi''(t) \geq 0.5t^{-1}\psi'(t) \forall t \in [1, \infty)$ .

Due to the properties 1<sup>0</sup> and 2<sup>0</sup> for any given vector  $\mathbf{k} = (k_1, \dots, k_q) \in \mathbb{R}_{++}^q$ , we have

$$c_i(x) \geq 0 \iff k_i^{-1}\psi(k_i c_i(x)) \geq 0, \quad i = 1, \dots, q$$

The Lagrangian  $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}_+^q \times \mathbb{R}_{++}^q \rightarrow \mathbb{R}$  for the equivalent problem

$$\mathcal{L}(x, \lambda, \mathbf{k}) = f(x) - \sum_{i=1}^q k_i^{-1} \lambda_i \psi(k_i c_i(x)) \tag{5}$$

is our main instrument.

First, we describe the NR method.

Let  $x^0 \in \mathbb{R}^n, \lambda^0 \in \mathbb{R}_{++}^q, k > 0$  and  $\mathbf{k}^0 = (k_i^0 = k(\lambda_i^0)^{-1}, i = 1, \dots, q)$ .

The NR multipliers method maps the triple  $(x^s, \lambda^s, \mathbf{k}^s)$  into the triple  $(x^{s+1}, \lambda^{s+1}, \mathbf{k}^{s+1})$  defined by the following formulas:

$$x^{s+1} = \operatorname{argmin}\{\mathcal{L}(x, \lambda^s, \mathbf{k}^s) \mid x \in \mathbb{R}^n\}, \tag{6}$$

$$\lambda_i^{s+1} = \lambda_i^s \psi'(k_i^s c_i(x^{s+1})), \quad i = 1, \dots, q, \tag{7}$$

$$k_i^{s+1} = k(\lambda_i^{s+1})^{-1}, \quad i = 1, \dots, q. \tag{8}$$

The minimizer  $x^{s+1}$  in (6) exists for any  $\lambda^s \in \mathbb{R}_{++}^q$  and any  $\mathbf{k}^s \in \mathbb{R}_{++}^q$  due to the boundness of  $X^*$ , convexity of  $f$ , concavity of  $c_i$  and properties 3<sup>0</sup> and 4<sup>0</sup> of  $\psi \in \Psi$ . It can be proven using considerations similar to those in ref. [1], i.e. the NR method is well defined.

It is well known (see [2,4,20]) that the NR method (6)–(8) is equivalent to the following Interior Prox method

$$\lambda^{s+1} = \operatorname{argmax}\{d(\lambda) - k^{-1}D(\lambda, \lambda^s) \mid \lambda \in \mathbb{R}^q\}, \tag{9}$$

where  $D : \mathbb{R}_+^q \times \mathbb{R}_{++}^q \rightarrow \mathbb{R}_+$ , given by the formula  $D(u, v) = \sum_{i=1}^q v_i^2 \varphi(u_i/v_i)$ , is the second-order  $\varphi$ -divergence distance with the kernel  $\varphi = -\psi^* : \mathbb{R}_{++} \rightarrow \mathbb{R}_+$ , where  $\psi^*$  is the Fenchel transform of  $\psi$ .

On the other hand, NR method (6)–(8) is equivalent to the Interior Quadratic Prox for the dual problem in the rescaled from step to step dual space (see [15,16]), i.e.

$$\lambda^{s+1} = \operatorname{argmax} \left\{ d(\lambda) - \frac{1}{2}k^{-1} \|\lambda - \lambda^s\|_{R_s}^2 \mid \lambda \in \mathbb{R}^q \right\}, \tag{10}$$

where  $\|\lambda\|_{R_s}^2 = \lambda^T R_s \lambda, R_s = (-\Psi''_{[s]}(\cdot))^{-1}, \Psi''_{[s]}(\cdot) = \operatorname{diag}(\psi''_{[s,i]}(\cdot))_{i=1}^q, \psi''_{[s,i]}(\cdot) = \psi''(\theta_i^s k_i^s c_i(x^{s+1}))$  and  $0 < \theta_i^s < 1$ .

Let  $\Phi = \{\varphi : \varphi = -\psi^*, \psi \in \Psi\}$  be the class of kernels, which corresponds to the class  $\Psi$  of transformations  $\psi$ . The properties of the kernels  $\varphi$  induced by the properties 1<sup>0</sup>–5<sup>0</sup> of the original transformation  $\psi \in \Psi$  were established in the following theorem.

**THEOREM 3.1 [5]** *The kernels  $\varphi \in \Phi$  are convex twice continuously differentiable and possess the following properties:*

1.  $\varphi(s) \geq 0 \forall s \in (0, \infty)$  and  $\min_{s \geq 0} \varphi(s) = \varphi(1) = 0$ ;

2. (a)  $\lim_{s \rightarrow 0^+} \varphi'(s) = -\infty$ , (b)  $\varphi'(s)$  is monotone increasing and (c)  $\varphi'(1) = 0$ ;
3. (a)  $\varphi''(s) \geq m > 0 \forall s \in (0, \infty)$  and (b)  $\varphi''(s) \leq M < \infty \forall s \in [1, \infty)$ .

Unfortunately, several well-known transformations including exponential  $\psi_1(t) = 1 - e^{-t}$  [1], logarithmic  $\psi_2(t) = \ln(t + 1)$  and hyperbolic  $\psi_3(t) = t(t + 1)^{-1}$  Modified Barrier Function (MBF) [13] as well as log-sigmoid  $\psi_4(t) = 2(\ln 2 + t - \ln(1 + e^t))$  and modified Chen-Harker-Kanzow-Smale (CHKS) transformation  $\psi_5(t) = t - (t^2 + 4\eta)^{1/2} + 2\sqrt{\eta}$ ,  $\eta > 0$  (see [15]) do not satisfy 1<sup>0</sup>–5<sup>0</sup>. Transformations  $\psi_1 - \psi_3$  do not satisfy the property 3<sup>0</sup> ( $m = 0$ ), while for  $\psi_4$  and  $\psi_5$  the property 4<sup>0</sup> is violated ( $M = \infty$ ). This can be fixed (see [16]) by using the quadratic extrapolation idea, which was first applied in ref. [3] to modify the logarithmic MBF transformation  $\psi_2$ .

Other kernels  $\varphi \in \Phi$  were considered in ref. [2] (see also [18]). In particular for the regularized logarithmic MBF kernel  $\varphi(t) = 0.5\nu(t - 1)^2 + \mu(t - \ln t - 1)$  with  $\nu > 0$ ,  $\mu > 0$ , the authors in ref. [2] proved that the dual sequence generated by the Interior Prox method (9) converges to the dual solution with  $O((ks)^{-1})$  rate under very mild assumptions on the input data.

We would like to point out that the properties 3<sup>0</sup> and 4<sup>0</sup> of  $\psi \in \Psi$  and the correspondent properties 3(a) and 3(b) of the kernel  $\varphi \in \Phi$  are critical for our convergence proof of the NR method (6)–(8) and its dual equivalents (9) and (10) (see [15,16]).

In particular, properties 3(a) and 3(b) of the kernel  $\varphi \in \Phi$  lead to the following bounds

$$d(\lambda^{s+1}) - d(\lambda^s) \geq mk^{-1} \|\lambda^{s+1} - \lambda^s\|^2 \tag{11}$$

$$d(\lambda^{s+1}) - d(\lambda^s) \geq kmM^{-2} \sum_{i \in I^-(x^{s+1})} c_i^2(x^{s+1}), \tag{12}$$

where  $I^-(x) = \{i : c_i(x) < 0\}$ .

The convergence of the primal  $\{x^s\}_{s=0}^\infty$  and the dual  $\{\lambda^s\}_{s=0}^\infty$  sequences in value, i.e.

$$f(x^*) = \lim_{s \rightarrow \infty} f(x^s) = \lim_{s \rightarrow \infty} d(\lambda^s) = d(\lambda^*),$$

is a direct consequence of the assumptions **A** and **B**, the equivalence of the NR method (6)–(8) to the Interior Quadratic Prox (10) and bounds (11) and (12) (see [15,16]).

We would like to emphasize that if  $m = 0$  and/or  $M = \infty$ , then the bounds (11) and (12) are trivial and useless, and the convergence of the NR method (6)–(8) even for a particular exponential transformation become problematic (see [20, p. 3]).

Also, it follows from (12) that for any  $\tau < 0$  and any  $i = 1, \dots, q$ , the inequality  $c_i(x^{s+1}) \leq \tau$  is possible only for a finite number of steps. Therefore, from some point onwards, only original transformations  $\psi_1 - \psi_5$  are used in the NR method. In fact, for  $k > 0$  large enough, the quadratic branch will be used just once. Therefore, the asymptotic analysis and the numerical performance of both the NR method (6)–(8) and its dual equivalents (9)–(10) depend only on the properties of the original transformations  $\psi_1 - \psi_5$  and the corresponding original dual kernels  $\varphi_1 - \varphi_5$ . The transformations  $\psi_1 - \psi_5$  for  $t \geq \tau$  are infinite time differentiable and so is the Lagrangian  $\mathcal{L}(x, \lambda, \mathbf{k})$  if the input data has the corresponding property. This allows us to use the Newton method for solving the PD system, which is equivalent to (6)–(7). We will concentrate on it in Section 4.

Each second-order  $\varphi$ -divergence distance function  $D_i(u, v) = \sum_{i=1}^q v_i^2 \varphi_i(u_i/v_i)$  leads to a corresponding Interior Prox method (9) for finding a maximum of a concave function on  $\mathbb{R}_+^q$ .

Sometimes the origin of the function  $d(\lambda)$  is irrelevant for the convergence analysis of the Prox method (9) (see [2]). However, when the dual function  $d(\lambda)$  is a product of the Lagrangian duality, such analysis can produce only limited results, because neither the primal nor the dual sequence

controls the NR method. The NR method is controlled rather by the PD system, solving which is equivalent to the NR step.

The PD system is defined by the PD map. The properties of the map are critical for establishing the rate of convergence of NR methods under the standard second-order optimality conditions (3)–(4) (see [13,14]).

In the remaining part of this section, we just mention some of these results, which we will use later. The results are taking place for any transformation  $\psi \in \Psi$  and can be proven using the correspondent PD map and arguments similar to those in refs. [13,14].

From (8), we have  $\lim_{s \rightarrow \infty} k_i^s = k(\lambda_i^*)^{-1}$ ,  $i = 1, \dots, r$ , i.e. the scaling parameters corresponding to the active constraints grow linearly with  $k > 0$ . Therefore, the technique which has been used in refs. [13,14] can be applied for the asymptotic analysis of the method (6)–(8).

For a given small enough  $\delta > 0$ , and large enough  $k \geq k_0 > 0$ , we define the following set:

$$D(\lambda^*, \mathbf{k}, \delta) = \{(\lambda, \mathbf{k}) \in \mathbb{R}_+^q \times \mathbb{R}_{++}^q : \lambda_i \geq \delta, |\lambda_i - \lambda_i^*| \leq \delta k, i = 1, \dots, r, \\ 0 < \lambda_i \leq k\delta, k \geq k_0, i = r + 1, \dots, q; \mathbf{k} = (k_i = k\lambda_i^{-1}, i = 1, \dots, q)\}.$$

The following theorem is similar to Theorem 6.2 in ref. [14] and can be proven using the same technique.

**THEOREM 3.2** *If  $f, c_i \in C^2$ , and the standard second-order optimality conditions (3)–(4) hold, then there exists sufficiently small  $\delta > 0$  and large enough  $k_0 > 0$  such that for any  $(\lambda, \mathbf{k}) \in D(\cdot)$ , we have the following:*

(1) *There exists  $\hat{x} = \hat{x}(\lambda, \mathbf{k}) = \operatorname{argmin}\{\mathcal{L}(x, \lambda, \mathbf{k}) \mid x \in \mathbb{R}^n\}$  such that*

$$\nabla_x \mathcal{L}(\hat{x}, \lambda, \mathbf{k}) = 0$$

*and*

$$\hat{\lambda}_i = \lambda_i \psi'(k_i c_i(\hat{x})), \quad \hat{k}_i = k \hat{\lambda}_i^{-1}, \quad i = 1, \dots, q.$$

(2) *For the pair  $(\hat{x}, \hat{\lambda})$ , the bound*

$$\max\{\|\hat{x} - x^*\|, \|\hat{\lambda} - \lambda^*\|\} \leq ck^{-1} \|\lambda - \lambda^*\|$$

*holds and  $c > 0$  is independent on  $k \geq k_0$ .*

(3) *The Lagrangian transformation  $\mathcal{L}(x, \lambda, \mathbf{k})$  is strongly convex in the neighbourhood of  $\hat{x}$ .*

The results of Theorem 3.2 do not require convexity of  $f$  and all  $-c_i, i = 1, \dots, q$ . Therefore, the NR method can be used for solving non-convex optimization problems as long as we can find an approximation for the minimizer  $\hat{x}(\lambda, k)$  for a large enough  $k \geq k_0$ . Then, after the first Lagrange multipliers and scaling vector update, the NR method (6)–(8) at each step requires finding an approximation for the minimizer of a strongly convex function. To find an approximation for the first unconstrained minimizer, one can use the interesting cubic regularization of the Newton method recently developed in ref. [10].

Finding  $x^{s+1}$  requires solving an unconstrained minimization problem (6), which is generally speaking, an infinite procedure. The following stopping criteria (see [14]) allows to replace  $x^{s+1}$  by an approximation  $\bar{x}^{s+1}$ , which can be found in a finite number of Newton steps by minimizing  $\mathcal{L}(x, \bar{\lambda}^s, \bar{\mathbf{k}}^s)$  in  $x \in \mathbb{R}^n$ . Then, the primal approximation  $\bar{x}^{s+1}$  is used instead of  $x^{s+1}$  in (7) for the Lagrange multipliers update, and the dual approximation  $\bar{\lambda}^{s+1}$  is used instead of  $\lambda^{s+1}$  in (8) for the scaling vector update.

The bounds similar to those established in (2) of Theorem 3.2 remain true.

For a given small enough  $\sigma > 0$ , let us consider the sequence  $\{\bar{x}^s, \bar{\lambda}^s, \bar{\mathbf{k}}^s\}$ , which is generated by the following formulas:

$$\bar{x}^{s+1} : \|\nabla_x \mathcal{L}(\bar{x}^{s+1}, \bar{\lambda}^s, \bar{\mathbf{k}}^s)\| \leq \sigma k^{-1} \|\Psi'(\bar{\mathbf{k}}^s c(\bar{x}^{s+1}))\bar{\lambda}^s - \bar{\lambda}^s\|, \quad (13)$$

$$\bar{\lambda}^{s+1} = \Psi'(\bar{\mathbf{k}}^s c(\bar{x}^{s+1}))\bar{\lambda}^s, \quad (14)$$

where

$$\Psi'(\bar{\mathbf{k}}^s c(\bar{x}^{s+1})) = \text{diag}(\psi'(\bar{\mathbf{k}}_i^s c_i(\bar{x}^{s+1})))_{i=1}^q$$

and

$$\bar{\mathbf{k}}^{s+1} = \bar{k}_i^{s+1} = k(\bar{\lambda}_i^{s+1})^{-1}, \quad i = 1, \dots, q.$$

The following theorem can be proven using the same technique that we used for proving Theorem 7.1 in ref. [14].

**THEOREM 3.3** *If the standard second-order optimality conditions (3)–(4) hold and the Hessians  $\nabla^2 f(x)$  and  $\nabla^2 c_i(x)$ ,  $i = 1, \dots, m$  satisfy the Lipschitz conditions*

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L_0 \|x - y\|, \quad \|\nabla^2 c_i(x) - \nabla^2 c_i(y)\| \leq L_i \|x - y\|, \quad (15)$$

*then there is  $k_0 > 0$  large enough such that for the PD sequence  $\{\bar{x}^s, \bar{\lambda}^s\}$  generated by the formulas (13) and (14), the following bounds hold true and  $c > 0$  is independent of  $k \geq k_0$  for  $s \geq 0$  :*

$$\|\bar{x}^{s+1} - x^*\| \leq c(1 + \sigma)k^{-1} \|\bar{\lambda}^s - \lambda^*\|, \quad \|\bar{\lambda}^{s+1} - \lambda^*\| \leq c(1 + \sigma)k^{-1} \|\bar{\lambda}^s - \lambda^*\|. \quad (16)$$

To find an approximation  $\bar{x}^{s+1}$ , one can use the Newton method with step length for minimization  $\mathcal{L}(x, \bar{\lambda}^s, \bar{\mathbf{k}}^s)$  in  $x$ . It requires, generally speaking, several Newton steps to find  $\bar{x}^{s+1}$ . Then we update the vector of Lagrange multipliers  $\bar{\lambda}^s$  using  $\bar{x}^{s+1}$  instead of  $x^{s+1}$  in (7) and update the scaling vector  $\bar{\mathbf{k}}^s$  using  $\bar{\lambda}^{s+1}$  instead of  $\lambda^{s+1}$  in (8).

For the logarithmic MBF transformation  $\psi(t) = \ln(t + 1)$  and  $\lambda = e = (1, \dots, 1) \in \mathbb{R}^q$ ,  $\mathbf{k} = ke$ , the Lagrangian  $\mathcal{L}(x, e, ke) = f(x) - \sum_{i=1}^q k^{-1} \ln(kc_i(x) + 1)$  is a self-concordant function if for example  $f$  and  $-c_i$ ,  $i = 1, \dots, q$  are linear- or convex-quadratic functions. In such a case, the interior point methods (see [9]) can be used to find the first approximation  $\bar{x}^0$  for the primal minimizer of the Lagrangian  $\mathcal{L}(x, e, ke)$ .

Having  $\bar{x}^0$ , we can apply formulas (7)–(8) for finding  $\{\bar{\lambda}^0, \bar{\mathbf{k}}^0\} \in D(\lambda^*, \mathbf{k}, \delta)$ . It follows from Theorems 3.2 and 3.3 that the entire sequence  $\{\bar{\lambda}^s, \bar{\mathbf{k}}^s\} \subset D(\lambda^*, \mathbf{k}, \delta)$  if this is true for the pair  $\{\bar{\lambda}^0, \bar{\mathbf{k}}^0\}$  (see [13, p. 195]). Moreover, the Lagrangians  $\mathcal{L}(x, \bar{\lambda}^s, \bar{\mathbf{k}}^s)$  for the equivalent problem are strongly convex in  $x$  in the neighbourhood of  $\bar{x}^s$ . Therefore, the level sets  $\mathcal{L}_s = \{x : \mathcal{L}(x, \bar{\lambda}^s, \bar{\mathbf{k}}^s) \leq \mathcal{L}(\bar{x}^{s-1}, \bar{\lambda}^s, \bar{\mathbf{k}}^s)\}$  are bounded  $\forall s \geq 0$ .

It allows retaining the classical polynomial complexity bounds (see [9]) at the initial stage and speeds up the process substantially in the final stage using the PDEP method, which we consider in the following section.

Instead of finding  $\bar{x}^{s+1}$  and then updating the Lagrange multipliers, we consider a PD system. Finding an approximate solution for the PD system is equivalent to finding  $\bar{x}^{s+1}$  and  $\bar{\lambda}^{s+1}$ . The application of the Newton method for solving the PD system leads to the PDEP method.

#### 4. Local PDEP method

In this section, we describe the PDEP method and prove its local quadratic rate of convergence. One step of the NR method (6)–(8) maps the given triple  $(x, \lambda, \mathbf{k}) \in \mathbb{R}^n \times \mathbb{R}_{++}^q \times \mathbb{R}_{++}^q$  into a triple

$(\hat{x}, \hat{\lambda}, \hat{\mathbf{k}}) \in \mathbb{R}^n \times \mathbb{R}_{++}^q \times \mathbb{R}_{++}^q$  by formulas

$$\begin{aligned} \hat{x} : \nabla_x \mathcal{L}(\hat{x}, \lambda, \mathbf{k}) &= \nabla f(\hat{x}) - \sum \psi'(k_i c_i(\hat{x})) \lambda_i \nabla c_i(\hat{x}) \\ &= \nabla f(\hat{x}) - \sum \hat{\lambda}_i \nabla c_i(\hat{x}) = 0, \end{aligned} \tag{17}$$

$$\hat{\lambda} : \hat{\lambda}_i = \lambda_i \psi'(k_i c_i(\hat{x})), \quad i = 1, \dots, q, \tag{18}$$

$$\hat{\mathbf{k}} : \hat{k}_i = k \hat{\lambda}_i^{-1}, \quad i = 1, \dots, q. \tag{19}$$

By removing the scaling vector update formula (19) from the system (17)–(19), we obtain the primal–dual NR system

$$\nabla_x L(\hat{x}, \hat{\lambda}) = \nabla f(\hat{x}) - \sum_{i=1}^q \hat{\lambda}_i \nabla c_i(\hat{x}) = 0, \tag{20}$$

$$\hat{\lambda} = \Psi'(\mathbf{k}c(\hat{x}))\lambda, \tag{21}$$

where  $\Psi'(\mathbf{k}c(\hat{x})) = \text{diag}(\psi'(k_i c_i(\hat{x})))_{i=1}^q$ .

By solving the PD system, a given vector of Lagrange multipliers is mapped into a new PD pair, while the penalty parameter and the scaling parameters vector remain fixed. The contractibility properties of the corresponding map are critical for both the convergence and the rate of convergence. To understand the conditions under which the corresponding map is contractive and to find the contractibility bounds, one has to analyze the PD map (see [13,14]). It should be emphasized that neither the primal NR sequence  $\{x^s\}$  nor the dual sequence  $\{\lambda^s\}$  provides sufficient information for this analysis. Only the PD system (20)–(21) has all necessary components for such analysis. This reflects the important observation that for any multipliers method, neither the primal nor the dual sequences controls the computational process. The numerical process is governed rather by the PD system. The importance of the PD system associated with non-linear rescaling methods has been recognized for quite some time (see [12,13]). It was used recently (see [8,17]) for developing PD methods for convex optimization with up to 1.5-Q-superlinear rate.

In this section, we use the specific properties of the PD system (20)–(21) for developing the PDEP method, which locally converges to the PD solution with quadratic rate.

From the standard second-order optimality condition (3)–(4) follows the uniqueness of  $x^*$  and  $\lambda^*$  and the existence of  $\tau^* > 0$  that (a)  $\min\{c_i(x^*) \mid r + 1 \leq i \leq q\} \geq \tau^*$  and (b)  $\min\{\lambda_i^* \mid 1 \leq i \leq r\} \geq \tau^*$ .

Therefore, due to (16), there is  $k_0 > 0$  large enough such that for any  $k \geq k_0$  and  $s \geq 1$

$$(a) \min\{c_i(\bar{x}^s) \mid r + 1 \leq i \leq q\} \geq 0.5\tau^* \quad \text{and} \quad (b) \min\{\bar{\lambda}_i^s \mid 1 \leq i \leq r\} \geq 0.5\tau^*. \tag{22}$$

Using formula (14) and the property  $2^0(c)$ , we have

$$\bar{\lambda}_i^{s+1} = \psi'(k_i^s c_i(\bar{x}^{s+1})) \bar{\lambda}_i^s \leq 2a(k\tau^*)^{-1} (\bar{\lambda}_i^s)^2, \quad s \geq 1.$$

Hence, for any fixed  $k > \max\{k_0, 2a(\tau^*)^{-1}\}$ , we have

$$\bar{\lambda}_i^{s+1} \leq (\bar{\lambda}_i^s)^2, \quad s \geq 1, \quad r + 1 \leq i \leq q.$$

So for a given accuracy  $0 < \varepsilon \ll 1$ , in at most  $s = O(\ln \ln \varepsilon^{-1})$  Lagrange multipliers updates, the Lagrange multipliers for the passive constraints will be of the order  $o(\varepsilon^2)$ . From this point onwards, all terms of the Lagrangian  $\mathcal{L}(x, \lambda, \mathbf{k})$  related to the passive constraints will

be automatically ignored in the calculations. Therefore, the PD system (20)–(21) will actually be reduced to the following system for  $\hat{x}$  and  $\hat{\lambda} = (\hat{\lambda}_1, \dots, \hat{\lambda}_r)$ :

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

$$\hat{\lambda}_i = \psi'(k_i c_i(\hat{x})) \lambda_i, \quad i = 1, \dots, r. \tag{24}$$

To simplify the notation, we use  $L(x, \lambda)$  for the truncated Lagrangian, i.e.,  $L(x, \lambda) = f(x) - \sum_{i=1}^r \lambda_i c_i(x)$ , and  $c(x)$  for the active constraints vector-function, i.e.,  $c^T(x) = (c_1(x), \dots, c_r(x))$ .

We use the vector norm  $\|x\| = \max_{1 \leq i \leq n} |x_i|$  and the matrix  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$  norm  $\|A\| = \max_{1 \leq i \leq n} (\sum_{j=1}^n |a_{ij}|)$ . For a given  $\varepsilon_0 > 0$ , we define the  $\varepsilon_0$ -neighbourhood  $\Omega_{\varepsilon_0} = \{y = (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_{++}^q : \|y - y^*\| \leq \varepsilon_0\}$  of the PD solution  $y^* = (x^*, \lambda^*)$ .

We will measure the distance between the current approximation  $y = (x, \lambda)$  and the solution  $y^*$  using the following merit function:

$$v(y) = v(x, \lambda) = \max \left\{ \|\nabla_x L(x, \lambda)\|, - \min_{1 \leq i \leq q} c_i(x), \sum_{i=1}^q |\lambda_i \|c_i(x)\|, - \min_{1 \leq i \leq q} \lambda_i \right\},$$

assuming that the input data is properly scaled.

It follows from the KKT conditions (1) and (2) that

$$v(x, \lambda) = 0 \iff x = x^*, \lambda = \lambda^*.$$

Later, we will use the following lemma.

LEMMA 4.1 [8] *If the standard second-order optimality condition (3)–(4) and Lipschitz condition (15) are satisfied, then there exists  $0 < m_0 < M_0 < \infty$  and  $\varepsilon_0 > 0$  small enough such that*

$$m_0 \|y - y^*\| \leq v(y) \leq M_0 \|y - y^*\|, \quad \forall y \in \Omega_{\varepsilon_0}. \tag{25}$$

It follows from (25) that in the neighbourhood  $\Omega_{\varepsilon_0}$ , the merit function  $v(y)$  is similar to  $\|\nabla f(x)\|$  for a strongly convex and smooth enough function  $f(x)$ . The merit function  $v(y)$  will be used:

- (1) to update the penalty parameter  $k > 0$ ;
- (2) to control accuracy at each step as well as for the overall stopping criteria;
- (3) to identify ‘small’ and ‘large’ Lagrange multipliers at each PDEP step;
- (4) to decide whether the primal or PD direction has to be used at the current step.
- (5) to regularize the Hessian  $\nabla_{xx}^2 L(x, \lambda)$ .

At first, we consider the Newton method for solving the PD system (23)–(24) and show its local quadratic convergence. To find the Newton direction  $(\Delta x, \Delta \lambda)$ , we have to linearize the system (23)–(24) at  $y = (x, \lambda)$ .

We start with system (24). Due to  $3^0$ , the inverse  $\psi'^{-1}$  exists. Therefore, using the identity  $\psi'^{-1} = \psi'^*$  and keeping in mind  $\varphi = -\psi^*$  and (19), we can rewrite (24) as follows:

$$c_i(\hat{x}) = k^{-1}\lambda_i\psi'^{-1}\left(\frac{\hat{\lambda}_i}{\lambda_i}\right) = k^{-1}\lambda_i\psi'^*\left(\frac{\hat{\lambda}_i}{\lambda_i}\right) = -k^{-1}\lambda_i\varphi'\left(\frac{\hat{\lambda}_i}{\lambda_i}\right).$$

Assuming  $\hat{x} = x + \Delta x$  and  $\hat{\lambda} = \lambda + \Delta\lambda$ , keeping in mind  $\varphi'(1) = 0$  and ignoring the terms of the second- and higher-order, we obtain

$$\begin{aligned} c_i(\hat{x}) &= c_i(x) + \nabla c_i(x)\Delta x = -k^{-1}\lambda_i\varphi'\left(\frac{\lambda_i + \Delta\lambda_i}{\lambda_i}\right) \\ &= -k^{-1}\lambda_i\varphi'\left(1 + \frac{\Delta\lambda_i}{\lambda_i}\right) = -k^{-1}\varphi''(1)\Delta\lambda_i, \quad i = 1, \dots, r, \end{aligned}$$

or

$$c_i(x) + \nabla c_i(x)\Delta x + k^{-1}\varphi''(1)\Delta\lambda_i = 0, \quad i = 1, \dots, r.$$

Now, we linearize the system (23) at  $y = (x, \lambda)$ . We have

$$\nabla f(x) + \nabla^2 f(x)\Delta x - \sum_{i=1}^r (\lambda_i + \Delta\lambda_i)(\nabla c_i(x) + \nabla^2 c_i(x)\Delta x) = 0.$$

Again, ignoring the terms of the second- and higher-orders, we obtain the following linearized PD system:

$$\nabla_{xx}^2 L(x, \lambda)\Delta x - \nabla c^T(x)\Delta\lambda = -\nabla_x L(x, \lambda), \quad (26)$$

$$\nabla c(x)\Delta x + k^{-1}\varphi''(1)I^r\Delta\lambda = -c(x), \quad (27)$$

where  $I^r$  is the identity matrix in  $R^r$  and  $\nabla c(x) = J(c(x))$  the Jacobian of the vector-function  $c(x)$ . Let us introduce matrix

$$N_k(x, \lambda) \equiv N_k(y) \equiv N_k(\cdot) = \begin{bmatrix} \nabla_{xx}^2 L(\cdot) & -\nabla c^T(\cdot) \\ \nabla c(\cdot) & k^{-1}\varphi''(1)I^r \end{bmatrix}.$$

Then, we can rewrite the system (26)–(27) as follows:

$$N_k(\cdot) \begin{bmatrix} \Delta x \\ \Delta\lambda \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) \\ \nabla_\lambda L(\cdot) \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) \\ -c(\cdot) \end{bmatrix}.$$

The local PDEP method consists of the following operations:

(1) Find the PD Newton direction  $\Delta y = (\Delta x, \Delta\lambda)$  from the system

$$N_k(\cdot) \begin{bmatrix} \Delta x \\ \Delta\lambda \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) \\ -c(\cdot) \end{bmatrix}. \quad (28)$$

(2) Find the new PD vector  $\hat{y} = (\hat{x}, \hat{\lambda})$  by formulas

$$\hat{x} := x + \Delta x, \quad \hat{\lambda} := \lambda + \Delta\lambda. \quad (29)$$

(3) Update the scaling parameter

$$\hat{k} = (v(\hat{y}))^{-1}. \tag{30}$$

Along with the matrix  $N_k(\cdot)$ , we consider the matrix

$$N_\infty(y) = N_\infty(\cdot) = \begin{bmatrix} \nabla^2 L(\cdot) & -\nabla c^T(\cdot) \\ \nabla c(\cdot) & 0 \end{bmatrix}.$$

Later, we will use the following technical lemmas.

**LEMMA 4.2** *Let  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be an invertible matrix and  $\|A^{-1}\| \leq c_0$ . Then for small enough  $\delta > 0$  and any  $B : \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that  $\|A - B\| \leq \delta$ , the matrix  $B$  is invertible and the following bounds hold:*

$$(a) \quad \|B^{-1}\| \leq 2c_0 \quad \text{and} \quad (b) \quad \|A^{-1} - B^{-1}\| \leq 2c_0^2\delta. \tag{31}$$

*Proof* The (31(a)) was proven in ref. [8] (see Lemma 2). The (31(b)) follows from the identity  $A^{-1} - B^{-1} = B^{-1}(B - A)A^{-1}$ , the inequality (31(a)) and the upper bounds for  $\|A^{-1}\|$  and  $\|A - B\|$ . ■

**LEMMA 4.3** *If the standard second-order optimality conditions (3)–(4) and the Lipschitz conditions (15) hold, then there exists small enough  $\varepsilon_0 > 0$  and large enough  $k_0 > 0$  such that both matrices  $N_\infty(y)$  and  $N_k(y)$  are non-singular and there is  $c_0 > 0$  independent of  $y \in \Omega_{\varepsilon_0}$  and  $k \geq k_0$ , so that*

$$\max\{\|N_\infty^{-1}(y)\|, \|N_k^{-1}(y)\|\} \leq 2c_0 \quad \forall y \in \Omega_{\varepsilon_0}, \quad \forall k \geq k_0. \tag{32}$$

*Proof* It is well known (see [11, Lemma 2, chapter 8]) that under the standard second-order optimality conditions (3)–(4), the matrix

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

is non-singular; hence, there exists  $c_0 > 0$  such that  $\|N_\infty^{-1}(y^*)\| \leq c_0$ . Due to the Lipschitz condition (15), there exists  $L > 0$  such that  $\|N_k(y) - N_\infty(y^*)\| \leq L\|y - y^*\| + k^{-1}\varphi''(1)$  and  $\|N_\infty(y) - N_\infty(y^*)\| \leq L\|y - y^*\|$ . Therefore, for a given small enough  $\delta > 0$ , there exist small  $\varepsilon_0 > 0$  and large  $k_0 > 0$  so that

$$\max\{\|N_k(y) - N_\infty(y^*)\|, \|N_\infty(y) - N_\infty(y^*)\|\} \leq \delta, \quad \forall y \in \Omega_{\varepsilon_0}, \quad \forall k \geq k_0.$$

Applying Lemma 4.2 first with  $A = N_\infty(y^*)$  and  $B = N_k(y)$  and then with  $A = N_\infty(y^*)$  and  $B = N_\infty(y)$ , we obtain (32). ■

The following theorem establishes the local quadratic convergence of the PDEP method.

**THEOREM 4.1** *If the standard second-order optimality conditions (3)–(4) and the Lipschitz condition (15) are satisfied, then there exists  $\varepsilon_0 > 0$  small enough such that for any PD pair  $y = (x, \lambda) \in \Omega_{\varepsilon_0}$ , the PDEP method (28)–(30) generates the PD sequence that converges to the PD solution with quadratic rate, i.e., the following bound holds:*

$$\|\hat{y} - y^*\| \leq c\|y - y^*\|^2, \quad \forall y \in \Omega_{\varepsilon_0},$$

and  $c > 0$  is independent on  $y \in \Omega_{\varepsilon_0}$ .

*Proof* We find the PD Newton direction  $\Delta y = (\Delta x, \Delta \lambda)$  from the system

$$N_k(y)\Delta y = b(y), \quad (33)$$

where

$$b(y) = \begin{bmatrix} -\nabla_x L(x, \lambda) \\ -c(x) \end{bmatrix}.$$

Along with the PD system (26)–(27), we consider the Lagrange system of equations, which corresponds to the active constraints at the same point  $y = (x, \lambda)$ :

$$\nabla_x L(x, \lambda) = \nabla f(x) - \nabla c(x)^T \lambda = 0, \quad (34)$$

$$c(x) = 0. \quad (35)$$

We apply the Newton method to Lagrange system (34)–(35) from the same starting point  $y = (x, \lambda)$ . We obtain the Newton direction  $\Delta \bar{y} = (\Delta \bar{x}, \Delta \bar{\lambda})$  for the system (34)–(35) from the following system of linear equations:

$$N_\infty(y)\Delta \bar{y} = b(y).$$

The new approximation for the system (34)–(35) is then obtained by formulas

$$\bar{x} = x + \Delta \bar{x}, \quad \bar{\lambda} = \lambda + \Delta \bar{\lambda} \quad \text{or} \quad \bar{y} = y + \Delta \bar{y}.$$

Under standard second-order optimality conditions (3) and (4) and the Lipschitz conditions (15), there is  $c_1 > 0$  independent on  $y \in \Omega_{\varepsilon_0}$  so that the following bound holds (see [11, Theorem 9, Chapter 8]):

$$\|\bar{y} - y^*\| \leq c_1 \|y - y^*\|^2. \quad (36)$$

Let us prove a similar bound for  $\|\hat{y} - y^*\|$ . We have

$$\begin{aligned} \|\hat{y} - y^*\| &= \|y + \Delta y - y^*\| = \|y + \Delta \bar{y} + \Delta y - \Delta \bar{y} - y^*\| \\ &\leq \|\bar{y} - y^*\| + \|\Delta y - \Delta \bar{y}\|. \end{aligned}$$

For  $\|\Delta y - \Delta \bar{y}\|$ , we obtain

$$\|\Delta y - \Delta \bar{y}\| = \|(N_k^{-1}(y) - N_\infty^{-1}(y))b(y)\| \leq \|N_k^{-1}(y) - N_\infty^{-1}(y)\| \|b(y)\|.$$

From Lemma 4.3, we have  $\max\{\|N_k^{-1}(y)\|, \|N_\infty^{-1}(y)\|\} \leq 2c_0$ . Besides,  $\|N_k(y) - N_\infty(y)\| = k^{-1}\varphi''(1)$ ; therefore, using Lemma 4.2 with  $A = N_k(y)$  and  $B = N_\infty(y)$ , we obtain

$$\|\Delta y - \Delta \bar{y}\| \leq 2k^{-1}\varphi''(1)c_0^2 \|b(y)\|. \quad (37)$$

In view of  $\nabla_x L(x^*, \lambda^*) = 0$ ,  $c(x^*) = 0$  and the Lipschitz condition (15), we have

$$\|b(y)\| \leq L \|y - y^*\|, \quad \forall y \in \Omega_{\varepsilon_0}.$$

Using (25), (30) and (37), we obtain

$$\begin{aligned} \|\Delta y - \Delta \bar{y}\| &\leq 2\varphi''(1)c_0^2 v(y)L \|y - y^*\| \\ &\leq 2\varphi''(1)c_0^2 M_0 L \|y - y^*\|^2. \end{aligned}$$

Therefore, for  $c_2 = 2\varphi''(1)c_0^2M_0L$ , which is independent of  $y \in \Omega_{\varepsilon_0}$ , we have

$$\|\Delta y - \Delta \bar{y}\| \leq c_2 \|y - y^*\|^2. \tag{38}$$

Using (36) and (38), for  $c = 2 \max\{c_1, c_2\}$ , we obtain

$$\|\hat{y} - y^*\| \leq \|\bar{y} - y^*\| + \|\Delta y - \Delta \bar{y}\| \leq c \|y - y^*\|^2, \quad \forall y \in \Omega_{\varepsilon_0}$$

and  $c > 0$  is independent of  $y \in \Omega_{\varepsilon_0}$ . We completed the proof. ■

### 5. Global PDEP method

In this section, we develop a globally convergent PDEP method and prove its asymptotic quadratic convergence rate under standard second-order optimality conditions.

The globally convergent PDEP method, roughly speaking, works as the Newton NR method in the initial phase and as the local PDEP method (28)–(30) in the final phase of the computational process.

We would like to emphasize that PDEP is not just a mechanical combination of two different methods, but is a unified procedure. Each step of the PDEP method consists of finding the PD direction  $\Delta y = (\Delta x, \Delta \lambda)$  by solving the linearized PD system (20)–(21). Then, we use either the PD Newton direction  $\Delta y$  to find a new PD approximation  $\hat{y}$  or the primal Newton direction  $\Delta x$  to minimize  $\mathcal{L}(x, \lambda, \mathbf{k})$  in  $x$ .

The choice at each step depends on the reduction of the merit function  $v(y)$  per step. If  $y \in \Omega_{\varepsilon_0}$ , then according to Theorem 4.1, each PDEP step produces a PD pair  $\hat{y} = (\hat{x}, \hat{\lambda})$  such that

$$\|\hat{y} - y^*\| \leq c \|y - y^*\|^2.$$

From the left inequality (25), we have  $\|y - y^*\| \leq m_0^{-1}v(y)$ ; therefore,  $\|\hat{y} - y^*\| \leq c \|y - y^*\|^2 \leq cm_0^{-2}(v(y))^2$ . Using the right inequality (25), we have

$$v(\hat{y}) \leq M_0 \|\hat{y} - y^*\| \leq c_0 M_0 \|y - y^*\|^2 \leq c M_0 m_0^{-2} v(y)^2.$$

Also,  $v(y) \leq M_0 \|y - y^*\| \leq M_0 \varepsilon_0, \forall y \in \Omega_{\varepsilon_0}$ . Thus for a small enough  $\varepsilon_0 > 0$ , we obtain

$$v(\hat{y}) \leq v(y)^{1.5}, \quad \forall y \in \Omega_{\varepsilon_0}$$

Therefore, if the PD step produces at least a 1.5-superlinear reduction of the merit function, then the PD step is accepted; otherwise, we use the primal direction  $\Delta x$  to minimize  $\mathcal{L}(x, \lambda, \mathbf{k})$  in  $x$ .

The important part of the method is the way the PD system (20)–(21) is linearized. Let us start with  $y = (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_{++}^q$  and compute  $v(y)$ .

By linearizing the system (20), we obtain

$$\nabla_{xx}^2 L(x, \lambda) \Delta x - \nabla c^T(x) \Delta \lambda = -\nabla_x L(x, \lambda). \tag{39}$$

The system (21) is split into two sub-systems. The first is associated with the set  $I_+(y) = \{i : \lambda_i > v(y)\}$  of ‘big’ Lagrange multipliers, whereas the second is associated with the set  $I_0(y) = \{i : \lambda_i \leq v(y)\}$  of ‘small’ Lagrange multipliers. Therefore,  $I_+(y) \cap I_0(y) = \emptyset$  and  $I_+(y) \cup I_0(y) =$

$\{1, \dots, q\}$ . We consider two sub-systems:

$$\hat{\lambda}_i = \psi'(k_i c_i(\hat{x})) \lambda_i, \quad i \in I_+(y), \quad (40)$$

and

$$\hat{\lambda}_i = \psi'(k_i c_i(\hat{x})) \lambda_i, \quad i \in I_0(y). \quad (41)$$

The system of Equations (40) can be rewritten as follows:

$$k_i c_i(\hat{x}) = \psi'^{-1}(\hat{\lambda}_i / \lambda_i) = -\varphi'(\hat{\lambda}_i / \lambda_i).$$

Let  $\hat{x} = x + \Delta x$  and  $\hat{\lambda} = \lambda + \Delta \lambda$ , then

$$c_i(x) + \nabla c_i(x) \Delta x = -k_i^{-1} \varphi'(1 + \Delta \lambda_i / \lambda_i), \quad i \in I_+(y).$$

Taking into account  $\varphi'(1) = 0$ , formula (8) and ignoring the terms of the second- and higher-order, we obtain

$$c_i(x) + \nabla c_i(x) \Delta x = -(k_i \lambda_i)^{-1} \varphi''(1) \Delta \lambda_i = -k^{-1} \varphi''(1) \Delta \lambda_i, \quad i \in I_+(y). \quad (42)$$

Let  $c_+(x)$  be the vector function associated with ‘big’ Lagrange multipliers, i.e.,  $c_+(x) = (c_i(x), i \in I_+(y))$ ,  $\nabla c_+(x) = J(c_+(x))$  the correspondent Jacobian and  $\Delta \lambda_+ = (\Delta \lambda_i, i \in I_+(y))$  the dual Newton direction associated with ‘big’ Lagrange multipliers. Then the system (42) can be rewritten as follows:

$$\nabla c_+(x) \Delta x + k^{-1} \varphi''(1) \Delta \lambda_+ = -c_+(x). \quad (43)$$

Now let us linearize the system (41). Keeping in mind (8) and ignoring the terms of the second- and higher-order, we obtain

$$\begin{aligned} \hat{\lambda}_i &= \lambda_i + \Delta \lambda_i = \psi'(k_i c_i(x) + \nabla c_i(x) \Delta x) \lambda_i \\ &= \psi'(k_i c_i(x)) \lambda_i + k \psi''(k_i c_i(x)) \lambda_i \Delta c_i(x) \Delta x \\ &= \bar{\lambda}_i + k \psi''(k_i c_i(x)) \lambda_i \nabla c_i(x) \Delta x, \quad i \in I_0(y). \end{aligned} \quad (44)$$

where  $\bar{\lambda}_i = \psi'(k_i c_i(x)) \lambda_i, i \in I_0(y)$ .

Let  $c_0(x)$  be the vector function associated with ‘small’ Lagrange multipliers,  $\nabla c_0(x) = J(c_0(x))$  the corresponding Jacobian,  $\lambda_0 = (\lambda_i, i \in I_0(y))$  the vector of ‘small’ Lagrange multipliers and  $\Delta \lambda_0 = (\Delta \lambda_i, i \in I_0(y))$  the corresponding dual Newton direction. Then (44) can be rewritten as follows:

$$-k \Psi''(\mathbf{k}_0 c_0(x)) \Delta c_0(x) \Lambda_0 \Delta x + \Delta \lambda_0 = \bar{\lambda}_0 - \lambda_0, \quad (45)$$

where  $\bar{\lambda}_0 = \Psi'(\mathbf{k}_0 c_0(x)) \lambda_0$ ,

$$\begin{aligned} \Psi'(\mathbf{k}_0 c_0(x)) &= \text{diag}(\psi'(k_i c_i(x)))_{i \in I_0(y)}, \quad \Lambda_0 = \text{diag}(\lambda_i)_{i \in I_0(y)} \\ \Psi''(\mathbf{k}_0 c_0(x)) &= \text{diag}(\psi''(k \lambda_i^{-1} c_i(x)))_{i \in I_0(y)}. \end{aligned}$$

Combining (39), (44), and (45), we obtain the following system for finding the PD direction  $\Delta y = (\Delta x, \Delta \lambda)$ , where  $\Delta \lambda = (\Delta \lambda_+, \Delta \lambda_0)$  and  $I_B$  and  $I_S$  are identity matrices in spaces of ‘big’

and ‘small’ Lagrange multipliers:

$$\begin{aligned}
 M(x, \lambda) \Delta y &= \begin{bmatrix} \nabla_{xx}^2 L(x, \lambda) & -\nabla c_+^T(x) & -\nabla c_0^T(x) \\ \nabla c_+(x) & k^{-1} \varphi''(1) I_B & 0 \\ -k \Psi''(\mathbf{k}_0 c_0(x)) \Lambda_0 \nabla c_0(x) & 0 & I_S \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda_+ \\ \Delta \lambda_0 \end{bmatrix} \\
 &= \begin{bmatrix} -\nabla_x L(x, \lambda) \\ -c_+(x) \\ \bar{\lambda}_0 - \lambda_0 \end{bmatrix}. \tag{46}
 \end{aligned}$$

To guarantee the existence of the PD direction  $\Delta y$  for any  $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_+^q$ , we replace the system (46) by the following regularized system, where  $I^n$  is the identity matrix in  $\mathbb{R}^n$ :

$$\begin{aligned}
 M_k(x, \lambda) \Delta y &= \begin{bmatrix} \nabla_{xx}^2 L(x, \lambda) + k^{-1} I^n & -\nabla c_+^T(x) & -\nabla c_0^T(x) \\ \nabla c_+(x) & k^{-1} \varphi''(1) I_B & 0 \\ -k \Psi'' \Lambda_0(\mathbf{k}_0 c_0(x)) \nabla c_0(x) & 0 & I_S \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda_+ \\ \Delta \lambda_0 \end{bmatrix} \\
 &= \begin{bmatrix} -\nabla_x L(x, \lambda) \\ -c_+(x) \\ \bar{\lambda}_0 - \lambda_0 \end{bmatrix}. \tag{47}
 \end{aligned}$$

Finding the PDEP direction  $\Delta y = (\Delta x, \Delta \lambda)$  from system (47), we call the PDEPD( $x, \lambda$ ) procedure.

Now, we are ready to describe the global PDEP method.

*Step 1.* Initialization: We choose an initial primal approximation  $x^0 \in \mathbb{R}^n$ , Lagrange multipliers vector  $\lambda^0 = (1, \dots, 1) \in \mathbb{R}^q$ , large enough penalty parameter  $k > 0$  and vector of scaling parameters  $\mathbf{k}^0 = k \lambda^0$ . Let  $\varepsilon > 0$  be the overall accuracy. We choose parameters  $0 < \eta < 0.5$ ,  $0.5 < \sigma \leq 1$  and  $0 < \theta < 0.25$ . Set  $x := x^0$ ,  $\lambda := \lambda^0$ ,  $v := v(x, \lambda)$ ,  $\lambda_c := \lambda^0$ ,  $\mathbf{k} := \mathbf{k}^0$  and  $\mathbf{k}_c := \mathbf{k}$ .

*Step 2.* If  $v \leq \varepsilon$  then stop. Output:  $x, \lambda$ .

*Step 3.* Find the PD direction:  $\Delta y = (\Delta x, \Delta \lambda) := \text{PDEPD}(x, \lambda_c)$ . Set  $\hat{x} := x + \Delta x$ ;  $\hat{\lambda} := \lambda + \Delta \lambda$ .

*Step 4.* If  $v < 1$  and  $v(\hat{x}, \hat{\lambda}) \leq v^{1.5}$ , set  $x := \hat{x}$ ,  $\lambda := \hat{\lambda}$ ,  $v := v(x, \lambda)$  and  $k := v^{-1}$ . Go to Step 2.

*Step 5.* Set  $t := 1$ , decrease  $0 < t \leq 1$  until  $\mathcal{L}(x + t \Delta x, \lambda_c, \mathbf{k}_c) - \mathcal{L}(x, \lambda_c, \mathbf{k}_c) \leq \eta t \langle \nabla_x \mathcal{L}(x, \lambda_c, \mathbf{k}_c), \Delta x \rangle$ .

*Step 6.* Set  $x := x + t \Delta x$  and  $\hat{\lambda} := \Psi'(\mathbf{k}_c c(x)) \lambda_c$ .

*Step 7.* If  $\|\nabla_x \mathcal{L}(x, \lambda_c, \mathbf{k}_c)\| \leq (\sigma/k) \|\hat{\lambda} - \lambda_c\|$ , go to step 9.

*Step 8.* Find the PD direction  $\Delta y = (\Delta x, \Delta \lambda) := \text{PDEPD}(x, \lambda_c)$ . Go to step 5.

*Step 9.* If  $v(x, \hat{\lambda}) \leq (1 - \theta)v$ , set  $\lambda := \hat{\lambda}$ ,  $\lambda_c := \lambda$ ,  $v := v(x, \lambda)$ ,  $k := \max\{v^{-1}, k\}$ ,  $\mathbf{k} := (k_i = k \lambda_i^{-1}, i = 1, \dots, q)$  and  $\mathbf{k}_c := \mathbf{k}$ . Go to step 2.

*Step 10.* Set  $k := k(1 + \theta)$ , go to step 8.

The following theorem proves the global convergence of the PDEP method and establishes the asymptotic quadratic rate.

**THEOREM 5.1** *If the standard second-order optimality conditions (3)–(4) and the Lipschitz condition (15) are satisfied, then the PDEP method generates a globally convergent PD sequence that converges to the PD solution with asymptotic quadratic rate.*

*Proof* First of all, the matrix  $M_k(y) \equiv M_k(x, \lambda)$  is non-singular for any  $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_{++}^q$ ,  $\lambda \in \mathbb{R}_+^q$  and any  $k > 0$ . Let us consider a vector  $w = (u, v_+, v_0)$ . Keeping in mind  $\psi''(t) < 0$ ,

convexity  $f(x)$ , concavity  $c_i(x)$  and the regularization term  $k^{-1}I^n$ , it is easy to see that  $M_k(y)w = 0 \rightarrow w = 0$ . Therefore,  $M_k^{-1}(y)$  exists and the primal–dual NR direction  $\Delta y$  can be found for any  $y = (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_+^q$  and any  $k > 0$ .

We consider now the PD sequence  $\{y^s = (x^s, \lambda^s)\}$  along with the sequence of scaling vectors  $\{\mathbf{k}^s\}$ , as well as the penalty parameter sequence  $\{k_s\}$  and the merit function sequence  $\{\nu_s = \nu(y^s)\}$  generated by the PDEP method.

It follows from steps 4, 9, and 10 that the sequence  $\{k_s\}$  is monotonically increasing. At the same time, it follows from steps 4 and 9 that  $\{\nu_s\}$  is monotonically decreasing.

Due to the monotonicity of  $\{\nu_s\}$  for a given small  $\varepsilon_0 > 0$ , steps 3 and 4 can occur not more than  $O(\ln \ln \varepsilon_0^{-1})$  times before we reach a PD approximation with accuracy  $\varepsilon_0 > 0$ . We will show later that from this point onwards, the PDEP method generates a sequence  $\{y^s\}_{s \geq s_0}$ , which converges to the PD solution with quadratic rate.

Let us concentrate now on steps 5–9. First of all, due to  $2^0(a)$  and formula (7) for the Lagrange multipliers update (see step 6), we have  $\lambda^s \in \mathbb{R}_{++}^q$ . Therefore, due to update formula (8), the scaling vectors  $\mathbf{k}^s \in \mathbb{R}_{++}^q$ . Keeping in mind the monotonicity of  $\{\nu_s\}$  and the definition of the merit function, we have

$$c_i(x^s) \geq -\nu_0 = \nu(y^0), \quad i = 1, \dots, q, \quad \forall s \geq 0.$$

Using the Corollary 20 from ref. [16], we conclude that the boundness of  $\Omega_0 = \{x : c_i(x) \geq -\nu_0, i = 1, \dots, q\}$  follows from the boundness of  $\Omega$ . It means that the primal sequence  $\{x^s\}$  is bounded. Moreover, it follows from steps 4, 9, and 10 that  $\lim_{s \rightarrow \infty} \nu_s = 0$ . Therefore, using the definition of the merit function again, we have

$$\lim_{s \rightarrow \infty} \|\nabla_x L(x^s, \lambda^s)\| = \lim_{s \rightarrow \infty} \|\nabla f(x^s) - \sum_{i=1}^q \lambda_i^s \nabla c_i(x^s)\| = 0.$$

We recall that  $\lambda^s \in \mathbb{R}_{++}^q$ . Let us show that  $\{\lambda^s\}$  is bounded.

Assuming the opposite, that the dual sequence  $\{\lambda^s\}$  is unbounded, we obtain  $\lambda_{i_s}^s = \max\{\lambda_i^s \mid 1 \leq i \leq q\} \rightarrow \infty$ . Therefore, keeping in mind the boundness of  $\{x^s\}$ , we have

$$\begin{aligned} & \lim_{s \rightarrow \infty} \left\| (\lambda_{i_s}^s)^{-1} \nabla f(x^s) - \sum_{i=1}^q (\lambda_{i_s}^s)^{-1} \lambda_i^s \nabla c_i(x^s) \right\| \\ &= \lim_{s \rightarrow \infty} \left\| (\lambda_{i_s}^s)^{-1} \nabla f(x^s) - \sum_{i=1}^q \bar{\lambda}_i^s \nabla c_i(x^s) \right\| \\ &= \lim_{s \rightarrow \infty} \left\| \sum_{i=1}^q \bar{\lambda}_i^s \nabla c_i(x^s) \right\| = 0. \end{aligned}$$

Now, using the boundness of  $\{\bar{\lambda}_i^s\}$ , we can find a converging subsequence of  $\{\bar{\lambda}_i^s\}$ . Without losing generality, we can assume that  $\lim_{s \rightarrow \infty} \lambda^s = \bar{\lambda}$  and  $\lim_{s \rightarrow \infty} x^s = \bar{x} \in \Omega$ . Let  $\bar{I} = \{i : c_i(\bar{x}) = 0\}$ , then  $c_i(\bar{x}) > 0, i \in \bar{I}$ . Due to  $2^0(c)$ , we have  $\bar{\lambda}_i = 0, i \in \bar{I}$ . Therefore,  $\sum_{i \in \bar{I}} \bar{\lambda}_i c_i(\bar{x}) = 0$ , which is impossible due to the Slater condition, which follows from (3). Thus, both the primal  $\{x^s\}$  and the dual sequence  $\{\lambda^s\}$  are bounded.

The existence of  $\tau > 0$  that  $\nu(y^s) \geq \tau, \forall y \notin \Omega_{\varepsilon_0}$ , follows from (25). Therefore, from (30), we have  $k^{-1} = \nu(y^s) \geq \tau$ .

After finding  $\Delta\lambda_+$  and  $\Delta\lambda_0$  from the second and third system in (47) and substituting their values into the first system, we obtain

$$P_k(y^s)\Delta x \equiv P_k(x^s, \lambda^s)\Delta x = -\nabla_x L(x^s, \bar{\lambda}^s) = -\nabla_x \mathcal{L}(x^s, \lambda^s, \mathbf{k}^s), \quad (48)$$

where

$$P_k(y^s) = \nabla_{xx}^2 L(x^s, \lambda^s) + k^{-1}I^n + k(\varphi''(1))^{-1}\nabla c_+^T(x^s)\nabla c_+(x^s) - k\nabla c_0^T(x^s)\Lambda_0^s\Psi''(\mathbf{k}_0c_0(x^s))\nabla c_0(x^s) \quad (49)$$

and  $\bar{\lambda}^s = (\bar{\lambda}_+^s, \bar{\lambda}_0^s)$ —where  $\bar{\lambda}_+^s = \lambda_+^s - k(\psi''(1))^{-1}c_+(x^s)$  and  $\bar{\lambda}_0^s = (\lambda_i^s = \psi'(k_i c_i(x^s)))\lambda_i^{s-1}$ ,  $i \in I_0(y^s)$ .

Due to the convexity  $f$  and concavity  $c_i$ ,  $i = 1, \dots, q$ , the Hessian  $\nabla_{xx}^2 L(x^s, \lambda^s)$  is non-negative definite. It follows from 3<sup>0</sup> that  $\Psi''(\mathbf{k}_0c_0(x^s))$  is a diagonal matrix with negative entries. It follows from Theorem 3.1 that  $\varphi''(1) > 0$ . Therefore, the third and the fourth terms in (49) are non-negative definite matrices. Keeping in mind the second term in (49), we conclude that the symmetric matrix  $P_k(y^s)$  is positive definite.

It follows from (25) that there is  $\tau > 0$  that  $k^{-1} = v(y^s) > \tau$ ,  $\forall y^s \notin \bar{\Omega}_{\varepsilon_0}$ , i.e. mineigenvalue  $P_k(y^s) = \mu_k(y^s) \geq \tau$ ,  $\forall y^s \notin \bar{\Omega}_{\varepsilon_0}$ . On the other hand, for any  $y^s \in \Omega_{\varepsilon_0}$  and  $k \geq k_0$ , we have  $\mu_k(y^s) \geq \rho > 0$  due to the Debreu lemma [5], the standard second-order optimality condition (3)–(4) and the Lipschitz condition (15).

Therefore,  $\mu_k(y) \geq \mu = \min\{\tau, \rho\} > 0$ ,  $\forall y \in \mathbb{R}^n \times \mathbb{R}_+^q$ ,  $\forall k \geq k_0$ , i.e.

$$\langle P_k(y^s)v, v \rangle \geq \mu \langle v, v \rangle, \quad \mu > 0, \quad \forall v \in \mathbb{R}^n, \quad \forall y^s \in \mathbb{R}^n \times \mathbb{R}_+^q.$$

From (48), we have

$$\langle \nabla_x \mathcal{L}(x^s, \lambda^s, \mathbf{k}^s), \Delta x \rangle = -\langle P_k(y^s)\Delta x, \Delta x \rangle \leq -\mu \|\Delta x\|_2^2. \quad (50)$$

Therefore, the primal Newton direction  $\Delta x$  defined by (47) is a descent direction for minimization  $\mathcal{L}(x, \lambda^s, \mathbf{k}^s)$  in  $x$ . It follows from Theorem 3.3 that step 10 can be used only finite times. Thus, there is  $\bar{k} > 0$  that the penalty parameter  $k_s \leq \bar{k}$ , which together with the boundness of the PD sequence  $\{x^s, \lambda^s\}$  means that there is  $\bar{M} > 0$  such that  $\|\nabla_{xx}^2 \mathcal{L}(x^s, \lambda^s, \mathbf{k}^s)\| \leq \bar{M}$ . Therefore, for  $0 < \eta \leq 0.5$ , we can find  $t \geq t_0 > 0$  such that

$$\mathcal{L}(x^s + t\Delta x, \lambda^s, \mathbf{k}^s) - \mathcal{L}(x^s, \lambda^s, \mathbf{k}^s) \leq \eta t \langle \nabla_x \mathcal{L}(x^s, \lambda^s, \mathbf{k}^s), \Delta x \rangle \leq -\mu t \eta \|\Delta x\|_2^2. \quad (51)$$

It follows from Theorem 3.2 that for any given pair  $(\lambda^s, \mathbf{k}^s) \in D(\lambda^*, \mathbf{k}, \delta)$ , the minimizer  $\hat{x}^{s+1} = \hat{x}(\lambda^s, \mathbf{k}^s)$  exists and  $\mathcal{L}(x, \lambda^s, \mathbf{k}^s)$  is strongly convex in the neighbourhood of  $\hat{x}^{s+1}$ . Therefore, the level set  $\mathcal{L}_s = \{x : \mathcal{L}(x, \lambda^s, \mathbf{k}^s) \leq \mathcal{L}(x^{s-1}, \lambda^s, \mathbf{k}^s)\}$  is bounded.

Hence, the primal sequence generated by PDEPD (see step 6) with  $t \geq t_0 > 0$  defined from (51) converges to  $\hat{x}^{s+1} = \hat{x}(\lambda^s, \mathbf{k}^s) : \nabla_x \mathcal{L}(\hat{x}^{s+1}, \lambda^s, \mathbf{k}^s) = \nabla_x L(\hat{x}^{s+1}, \hat{\lambda}^{s+1}) = 0$ .

Keeping in mind the standard second-order optimality condition (3)–(4) and the Lipschitz condition (15), it follows from Theorem 3.3 that for the PD approximation  $(\bar{x}^{s+1}, \bar{\lambda}^{s+1})$ , the bound (16) holds. It requires a finite number of Newton steps to find  $\bar{x}^{s+1}$ . After  $s_0 = O(\ln \varepsilon_0^{-1})$  Lagrange multipliers and scaling parameters updates, we find the PD approximation  $y^s \in \Omega_{\varepsilon_0}$ .

Let  $0 < \varepsilon \ll \varepsilon_0 < 1$  be the desired accuracy.

Keeping in mind the properties 2<sup>0</sup>(c) and 2<sup>0</sup>(d) for the transformation  $\psi \in \Psi$  as well as (22), (25) and (30) after  $s_1 = O(\ln \ln \varepsilon^{-1})$  updates, we obtain

$$\max\{\|k\Psi''k(\lambda_0^s)^{-1}c_0(x^s)\|, \|\bar{\lambda}_0^s - \lambda_0^s\|\} = o(\varepsilon^2), \quad i \in I_0. \quad (52)$$

For any  $y^s \in \Omega_{\varepsilon_0}$ , the term  $\|\nabla c_0(x^s)\|$  is bounded. The boundedness of  $\|\Delta x\|$  follows from the boundedness of  $\|\nabla_{xx}^2 L(x^s, \bar{\lambda}^s)\|$  and the fact that  $P_k(y^s)$  has a mineigenvalue bounded from below by  $\mu > 0$  uniformly in  $y \in \Omega_{\varepsilon_0}$ .

Let us consider the third part of the system (47), which is associated with the ‘small’ Lagrange multipliers

$$k\Psi''k(\lambda_0^s)^{-1}c_0(x^s)\nabla c_0(x^s)\Delta x^s + \Delta\lambda_0^s = \bar{\lambda}_0^s - \lambda_0^s.$$

It follows from (52) that  $\|\Delta\lambda_0^s\| = o(\varepsilon^2)$ . This means that after  $s = \max\{s_0, s_1\}$  updates, the part of the system (47) associated with ‘small’ Lagrange multipliers becomes irrelevant for the calculation of a Newton direction from (47).

In other words, the PDEP automatically reduces (47) to the following system:

$$\bar{M}_k(x^s, \lambda^s)\Delta\bar{y}^s = \bar{b}(x^s, \lambda^s), \quad (53)$$

where  $(\Delta\bar{y}^s)^T = (\Delta x^s, \Delta\lambda_+^s)$ ,  $\bar{b}(x^s, \lambda^s)^T = (-\nabla_x L(x^s, \lambda^s), -c_+(x^s))$  and

$$\bar{M}_k(x^s, \lambda^s) = \begin{bmatrix} \nabla_{xx}^2 L(x^s, \lambda^s) + k^{-1}I^n & -\nabla c_{(+)}^T(x^s) \\ \nabla c_{(+)}(x^s) & k^{-1}\varphi''(1)I_B \end{bmatrix}.$$

At this point, we have  $y^s \in \Omega_{\varepsilon_0}$ ; therefore, it follows from (25) that  $v(y^s) \leq M_0\varepsilon_0$ . Hence, for small enough  $\varepsilon_0 > 0$  from  $|\lambda_i^s - \lambda_i^*| \leq \varepsilon_0$ , we obtain  $\lambda_i^s \geq v(y^s)$ ,  $i \in I^*$ . On the other hand, we have  $v(y^s) > \lambda_i^s = O(\varepsilon^2)$ ,  $i \in I_0$ ; otherwise, we obtain  $v(y^s) \leq O(\varepsilon^2)$  and from (25) follows  $\|y^s - y^*\| = o(\varepsilon^2)$ . So, if after  $s = \max\{s_0, s_1\}$  Lagrange multipliers updates, we have not solved the problem with a given accuracy  $\varepsilon > 0$ , then  $I_+(y) = I^* = \{1, \dots, r\}$  and  $I_0(y) = I_0^* = \{r + 1, \dots, q\}$ .

From this point onwards, the global PDEP is, in fact, automatically reduced to (28)–(30) with matrix

$$\bar{M}_k(y^s) = \bar{M}_k(x^s, \lambda^s) = \begin{bmatrix} \nabla_{xx}^2 L(x^s, \lambda^s) + k^{-1}I^n & -\nabla c_{(r)}^T(x^s) \\ \nabla c_{(r)}(x^s) & k^{-1}\varphi''(1)I^r \end{bmatrix}$$

instead of  $N_k(\cdot)$ , where  $L(x^s, \lambda^s) = f(x^s) - \sum_{i=1}^r \lambda_i^s c_i(x_i^s)$  is the truncated Lagrangian.

Therefore, we have

$$\|\Delta y^s - \Delta\bar{y}^s\| = \|(\bar{M}_k^{-1}(y^s) - N_\infty^{-1}(y^s))b(y^s)\| \leq \|\bar{M}_k^{-1}(y^s) - N_\infty^{-1}(y^s)\| \|b(y^s)\|.$$

On the other hand,  $\|\bar{M}_k(y^s) - N_\infty(y^s)\| \leq k^{-1}(1 + \varphi''(1))$ . From Lemma 4.3, we have  $\max\{\|\bar{M}_k^{-1}(y^s)\|, \|N_\infty^{-1}(y^s)\|\} \leq 2c_0$ . Keeping in mind (15), (25) and (30), we obtain the following bound:

$$\begin{aligned} \|\Delta y^s - \Delta\bar{y}^s\| &\leq 2c_0^2 k^{-1}(1 + \varphi''(1))\|\bar{b}(y^s)\| = 2c_0^2 v(y^s)(1 + \varphi''(1))\|\bar{b}(y^s)\| \\ &\leq 2(1 + \varphi''(1))c_0^2 M_0 L \|y^s - y^*\|^2 = c_3 \|y^s - y^*\|^2, \end{aligned} \quad (54)$$

where  $c_3 > 0$  is independent of  $y \in \Omega_{\varepsilon_0}$ .

By using the bound (54) instead of (38) and repeating the arguments we used in Theorem 4.4, we conclude that the PD sequence generated by the PDEP method converges to the PD solution  $(x^*, \lambda^*)$  with asymptotic quadratic rate. Therefore, in the case when  $\varepsilon \ll \varepsilon_0$ , it takes another  $O(\ln \ln \varepsilon^{-1})$  PDEP steps to obtain approximation for  $(x^*, \lambda^*)$  with accuracy  $\varepsilon > 0$  starting with any  $y \in \Omega_{\varepsilon_0}$ . ■

## 6. Concluding remarks

The PDEP is fundamentally different from the Newton NR method. The distinct characteristic of the PDEP is its global convergence with asymptotic quadratic rate. The PDEP combines the best

features of the Newton NR method and the Newton method for the Lagrange system of equations corresponding to the active constraints. At the same time, the PDEP is free from their critical drawbacks. In the initial phase, PDEP is similar to the Newton NR method, i.e. the Newton method for minimization of the Lagrangian  $\mathcal{L}(x, \lambda, \mathbf{k})$  in  $x$  followed by Lagrange multipliers and scaling parameters update. Such a method (see Theorem 3.3) converges under a fixed penalty parameter. This allows us to reach the  $\varepsilon_0$ -neighbourhood of the PD solution in  $O(\ln \varepsilon_0^{-1})$  Lagrange multipliers updates without compromising the condition number of the Hessian,  $\nabla_{xx}^2 \mathcal{L}(x, \lambda, \mathbf{k})$ .

Once in the neighbourhood of the PD solution, the penalty parameter, which is inversely proportional to the merit function, grows extremely fast. Again, the unbounded increase of the scaling parameter at this point does not compromise the numerical stability, because instead of unconstrained minimization, the PDEP method solves the primal–dual NR system. Moreover, the PD direction  $\Delta y$  turns out to be very close to the Newton direction (see (54)) for the Lagrange system of equations corresponding to the active constraints. This guarantees the asymptotic quadratic convergence.

The situation in some sense recalls the Newton method with step length for unconstrained smooth convex optimization.

Several issues remain for future research.

First, the neighbourhood of the PD solution where the quadratic rate of convergence occurs needs to be characterized through parameters that measure the non-degeneracy of constrained optimization problems.

Second, the value of the scaling parameter  $k_0 > 0$  is *a priori* unknown and depends on the condition number of the problem (P), which can be estimated through parameters of a constrained optimization problem at the solution (see [13]). These parameters are obviously unknown. Therefore, it is important to find an efficient way to adjust the penalty parameter  $k > 0$  using the merit function value.

Third, we have to understand to what extent the PDEP method can be used in the non-convex case. In this regard, recent results from ref. [10] together with local convexity properties of the Lagrangian (5) that follow from the Debreu lemma [5] may play an important role.

Fourth, numerical experiments using various versions of primal–dual NR methods produce very encouraging results (see [7,8,17]). On the other hand, the PDEP method has certain specific features that require more numerical work to better understand its practical efficiency.

## Acknowledgements

The author is grateful to the anonymous referee for comments that helped to improve the paper. The research was supported by NSF Grant CCF-0324999.

## References

- [1] A. Auslender, R. Cominetti, and M. Haddou, *Asymptotic analysis of penalty and barrier methods in convex and linear programming*, Math. Oper. Res. 22 (1997), pp. 43–62.
- [2] A. Auslender, M. Tebboule, and S. Ben-Tiba, *Interior proximal and multipliers methods based on secondorder homogeneous kernels*, Math. Oper. Res. 24 (1999), pp. 645–668.
- [3] A. Ben-Tal, B. Yuzefovich, and M. Zibulevsky, *Penalty-barrier multipliers methods for minimax and constrained smooth convex optimization*, Research Report 9-92, Optimization Laboratory, Technion, Israel, 1992.
- [4] A. Ben-Tal and M. Zibulevsky, *Penalty-barrier methods for convex programming problems*, SIAM J. Control Optim. 7 (1997), pp. 347–366.
- [5] G. Debreu, *Definite and semidefinite quadratic forms*, Econometrica 20 (1952), pp. 295–300.
- [6] A. Fiacco and G. McCormick, *Nonlinear Programming Sequential Unconstrained Minimization Technique*, SIAM Classic in Applied Mathematics, SIAM, Philadelphia, PA, 1990.
- [7] I. Griva, *Numerical experiments with an interior-exterior point method for nonlinear programming*, Computational Optimization and Applications, 29 (2004), pp. 173–195.

- [8] I. Griva and R. Polyak, *Primal-dual nonlinear rescaling method with dynamic scaling parameter update*, Math. Program. 106 (2006), pp. 237–259.
- [9] Yu. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, 1994.
- [10] Yu. Nesterov and B. Polyak, *Cubic regularization of Newton method and its global performance*, Math. Program. 108 (2006), pp. 177–205.
- [11] B. Polyak, *Introduction to Optimization*, Software Inc., NY, 1987.
- [12] R. Polyak, *Smooth optimization methods for minmax problems*, SIAM J. Control Optim. 26 (1988), pp. 1274–1286.
- [13] R. Polyak, *Modified barrier functions*, Math. Program. 54 (1992), pp. 177–222.
- [14] R. Polyak, *Log-sigmoid multipliers method in constrained optimization*, Ann. Oper. Res. 101 (2001), pp. 427–460.
- [15] R. Polyak, *Nonlinear rescaling vs. smoothing technique in convex optimization*, Math. Program. 92 (2002), pp. 197–235.
- [16] R. Polyak, *Nonlinear rescaling as interior quadratic prox method for convex optimization*, Comput. Optim. Appl. 35 (2006), pp. 347–373.
- [17] R. Polyak and I. Griva, *Primal-dual nonlinear rescaling methods for convex optimization*, JOTA 122 (2004), pp. 111–156.
- [18] R. Polyak and M. Teboulle, *Nonlinear rescaling and proximal-like methods in convex programming*, Math. Program. 76 (1997), pp. 265–284.
- [19] D.F. Shanno and R.J. Vanderbei, *An interior point algorithm for nonconvex nonlinear programming*, COAP 13 (1999), pp. 231–252.
- [20] P. Tseng and D. Bertsekas, *On the convergence of the exponential multipliers method for convex programming*, Math. Program. 60 (1993), pp. 1–19.