# LAGRANGIAN TRANSFORMATION IN CONVEX OPTIMIZATION

ROMAN A. POLYAK

Abstract. We introduce the Lagrangian Transformation(LT) and develop a general LT method for convex optimization problems.

A class $\Psi$ of strongly concave and smooth functions $\psi: \mathbb{R} \to \mathbb{R}$ with specific properties is used to transform the terms of the classical Lagrangian associated with constraints. The transformation is scaled by a positive vector of scaling parameters one for each constraint.

Each step of the LT method alternates unconstrained minimization of LT in primal space with both Lagrange multipliers and scaling parameters update.

The scaling parameters are updated inversely proportional to the square of the Lagrange multipliers. The updating the scaling parameters in such a way makes the LT multipliers method equivalent to the Interior Quadratic Prox for the dual problem in the rescaled dual space. We used this equivalence to prove convergence and to estimate the rate of convergence of the LT method under minimum assumptions on the input data.

Our main focus is on the primal-dual LT(PDLT) method. The PDLT generates a primal-dual sequence that is globally convergent to the primal-dual solution. We proved that under the standard second order optimality condition the PDLT sequence converges to the primal-dual solution with an asymptotic quadratic rate.

Keywords. Lagrangian transformation, duality, interior quadratic prox, primal-dual LT method.

## 1. Introduction

The Lagrangian Transformation is defined and smooth on the entire primal space.The LT combines the best properties of the classical Lagrangian and smooth penalty functions and at the same time is free from their basic drawbacks.

The LT multipliers method has several important properties.

First, both the Lagrange multipliers and the scaling parameters vector remain positive without any particular care, just due to the way they are updated from step to step. This allows to eliminate the combinatorial nature of constrained optimization problems with inequality constraints.

Second, the Lagrange multipliers that corresponding to the passive constraints converge to zero with at least a quadratic rate, thus allowing the detection of the active constraints in the early stage of the computational process.

Third, the scaling parameters that corresponding to the active constraints remain bounded when the primal-dual sequence approaches the primal-dual solution. This keeps the condition number of the LT Hessian stable, improves the rate of convergence as compared to the penalty methods and makes the computational process equally stable both far and near to solution.

The LT multipliers method is equivalent to the Interior Quadratic Prox for the dual problem in the rescaled dual space. It allows us to prove convergence as well as estimate the rate of convergence under minimum assumptions on the input data.

Our main focus is on the primal-dual LT(PDLT) method. We replace the unconstrained minimization of the LT in the primal space and the update of the Lagrange multipliers by solving the primal-dual(PD) system of equation. The application of Newton's method to the PD system leads to the PDLT method.

The PDLT method generates a globally convergent primal-dual sequence that under the standard second order optimality conditions converges to the primal-dual solution with a quadratic rate. This is our main contribution.

By solving the PD system a given vector of Lagrange multipliers is mapped into a new primal-dual pair, while the vector of scaling parameters remains fixed. The contractability properties of the corresponding map are critical both for the convergence and for 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 primal-dual maps (see [17]–[19]). It should be emphasized that neither the primal LT sequence nor the dual sequence generated by the Interior Quadratic Prox provides sufficient information for this analysis. Only the PD system, which is equivalent to one LT step, has all necessary components for such an analysis. This reflects the important observation that for any multipliers method neither the primal sequence nor the dual sequence control the computational process. The numerical process is governed rather by the PD system. The importance of the PD system associated with nonlinear rescaling methods has been recognized for quite some time (see [17]–[18]).

Recently the corresponding PD systems were used to develop globally convergent primal-dual nonlinear rescaling methods with an up to 1.5-Q superlinear rate (see [21]–[22]).

In this paper a general primal-dual LT method is developed. The method generates a globally convergent primal-dual sequence that under the standard second order optimality conditions converges to the primal-dual solution with a quadratic rate. In many aspects it reminds Newton's method for smooth unconstrained optimization.

This similarity becomes possible due to:

1) the special properties of $\psi \in \Psi$;

2) the structure of the LT method, in particular, the way in which the Lagrange multipliers and scaling parameters are updated at each step;

3) the fact that the Lagrange multipliers corresponding to the passive constraints converge to zero with at least a quadratic rate;

4) the way in which we use the merit functions for updating the penalty parameter;

5) the way in which we employ the merit function for the Lagrangian regularization.

During the initial phase the PDLT works as the Newton LT method, i.e. Newton's method for LT minimization followed by the Lagrange multipliers and the scaling parameters update. At some point, the Newton LT method automatically turns into Newton's method for the Lagrange system of equations corresponding to the active constraints.

It should be emphasized that the PDLT is free from any stringent conditions on accepting the Newton step, which are typical for constrained optimization problems.

There are three important features, that make Newton's method for the primal-dual LT system free from such restrictions.

First, the LT is defined on the entire primal space.

Second, after a few Lagrange multipliers update the terms of the LT corresponding to the passive constraints become negligibly small due to the quadratic convergence to zero of the Lagrange multipliers corresponding to the passive constraints. Therefore on the one hand these terms became irrelevant for finding the Newton direction. On the other hand, there is no need to enforce their nonnegativity.

Third, the LT multipliers method is, generally speaking, an exterior point method in the primal space. Therefore there is no need to enforce the nonnegativity of the the slack variables for the active constraints as it takes place in the Interior Point Methods (see [26]).

After a few Lagrange multipliers update the primal-dual LT direction becomes practically identical to the Newton direction for the Lagrange system of equations corresponding to the active constraints. This makes possible to prove the asymptotic convergence of the primal-dual sequence with quadratic rate.

The paper is organized as follows. In the next section we state the problem and the basic assumptions on the input data. In section 3 we describe the LT method and show that it is equivalent to an Interior Quadratic Prox. In section 4 we describe the convergence results for the LT method. In section 5 we introduce the Primal-Dual LT method and prove its local quadratic convergence under the standard second order optimality conditions. In section 6 we consider the globally convergent primal-dual LT method and show that the primal-dual sequence converges with an asymptotic quadratic rate. We conclude the paper with some remarks concerning possible future research.

## 2. Statement of the problem and basic assumptions

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

$$x^* \in X^* = \operatorname{Argmin}\{f(x) \mid x \in \Omega\}, \tag{P}$$

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

**A:** The optimal set $X^*$ is not empty and bounded.

**B:** The 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

$$\lambda \in L^* = \operatorname{Argmax}\{d(\lambda) \mid \lambda \in \mathbb{R}^q_+\}. \tag{D}$$

Due to the assumption **B**, the Karush-Kuhn-Tucker (KKT) conditions hold true, i. e. there exists a vector $\lambda^* = (\lambda_1^*, \dots, \lambda_q^*) \in \mathbb{R}^q_+$ such that

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i=1}^{q} \lambda_i^* \nabla c_i(x^*) = 0 \tag{2.1}$$

and the complementary slackness conditions

$$\lambda_i^* c_i(x^*) = 0, \quad i = 1, \dots, q \tag{2.2}$$

are satisfied. We assume that the active constraints set at $x^*$ is $I^* = \{i : c_i(x^*) = 0\} = \{1, \dots, r\}$. Let us consider the vector-functions $c^T(x) = (c_1(x), \dots, c_q(x))$, $c^T_{(r)}(x) = (c_1(x), \dots, c_r(x))$ and their Jacobians $\nabla c(x) = J(c(x))$ and $\nabla c_{(r)}(x) = J(c_{(r)}(x))$.

The sufficient regularity condition

$$\text{rank}\nabla c_{(r)}(x^*) = r, \quad \lambda_i^* > 0, \quad i \in I^* \tag{2.3}$$

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

$$(\nabla^2_{xx} L(x^*, \lambda^*)y, y) \geq \mu(y, y), \quad \mu > 0 \quad \forall y \neq 0 : \nabla c_{(r)}(x^*)y = 0 \tag{2.4}$$

comprise the standard second order optimality condition.

## 3. Lagrangian transformation

We consider a class $\Psi$ of twice continuous differentiable functions $\psi : (-\infty, \infty) \to \mathbb{R}$ with the following properties:

$1^0$. $\psi(0) = 0$;

$2^0$. a) $\psi'(t) > 0$; b) $\psi'(0) = 1$; c) $\psi'(t) \leq at^{-1}$; d) $|\psi''(t)| \leq bt^{-2}$ $\forall t \in [1, \infty)$, $a > 0$, $b > 0$;

$3^0$. $-m^{-1} \leq \psi''(t) < 0$ $\forall t \in (-\infty, \infty)$;

$4^0$. $\psi''(t) \leq -M^{-1}$ $\forall t \in (-\infty, 0]$ and $0 < m < M < \infty$;

$5^0$. $-\psi''(t) \geq 0.5t^{-1}\psi'(t)$ $\forall t \in [1, \infty)$.

The Lagrangian transformation $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^q_{++} \times \mathbb{R}^q_{++} \to \mathbb{R}$ we define by the following formula:

$$\mathcal{L}(x, \lambda, \mathbf{k}) = f(x) - \sum_{i=1}^{q} k_i^{-1} \psi(k_i \lambda_i c_i(x)), \tag{3.1}$$

and we assume $k_i \lambda_i^2 = k > 0$, $i = 1, \dots, q$. Due to concavity of $\psi(t)$, convexity of $f(x)$ and concavity of $c_i(x)$, $i = 1, \dots, q$ the LT $\mathcal{L}(x, \lambda, \mathbf{k})$ is a convex function in $x$ for any fixed $\lambda \in \mathbb{R}^q_{++}$ and $\mathbf{k} \in \mathbb{R}^q_{++}$. Also due to property $4^0$, assumption $\mathbf{A}$ and convexity of $f(x)$ and all $-c_i(x)$ for any given $\lambda \in \mathbb{R}^q_{++}$ and $\mathbf{k} = (k_1, \dots, k_q) \in \mathbb{R}^q_{++}$, the minimizer

$$\widehat{x} \equiv \widehat{x}(\lambda, \mathbf{k}) = \text{argmin}\{\mathcal{L}(x, \lambda, \mathbf{k}) \mid x \in \mathbb{R}^n\} \tag{3.2}$$

exists. It can be proved using arguments similar to those in [1]. Due to the complementarity condition (2.2) and properties $1^0$ and $2^0$b) for any KKT pair $(x^*, \lambda^*)$ and any $\mathbf{k} \in \mathbb{R}^q_{++}$ we have

1) $\mathcal{L}(x^*, \lambda^*, \mathbf{k}) = f(x^*)$;

2) $\nabla_x \mathcal{L}(x^*, \lambda^*, \mathbf{k}) = \nabla_x L(x^*, \lambda^*) = 0$, i.e.

$$x^* \in X^* = \text{Argmin}\{\mathcal{L}(x, \lambda^*, \mathbf{k}) \mid x \in \mathbb{R}^n\};$$

3) $\nabla^2_{xx}\mathcal{L}(x^*, \lambda^*, \mathbf{k}) = \nabla^2_{xx} L(x^*, \lambda^*) + \psi''(0)\nabla c_{(r)}(x^*) K_{(r)}, {\Lambda^*_{(r)}}^2 \nabla c_{(r)}(x^*)$, where $K_{(r)} = \text{diag}(k_i)_{i=1}^r$, $\Lambda_{(r)} = \text{diag}(\lambda_i)_{i=1}^r$. Therefore, for $K_{(r)} = k{\Lambda^*_{(r)}}^{-2}$ we have

$$\nabla^2_{xx}\mathcal{L}(x^*, \lambda^*, \mathbf{k}) = \nabla^2_{xx} L(x^*, \lambda^*) - k\psi''(0)\nabla c^T_{(r)}(x^*)\nabla c_{(r)}(x^*). \tag{3.3}$$

Before describing the LT multipliers method we would like to make a few comments about properties 1)–3). Let us consider a nonsmooth function $Q:\ \mathbb{R}^n \to \mathbb{R}_+$ given by the following formula

$$Q(x,x^*) = \max\{f(x) - f(x^*),\ -c_i(x),\ i = 1,\dots,q\}.$$

It is easy to see that the constrained optimization problem (P) is equivalent to the unconstrained nonsmooth optimization problem

$$Q(x^*,x^*) = \min\{Q(x,x^*) \mid x\in\mathbb{R}^n\} = 0.$$

For any fixed $\mathbf{k} \in \mathbb{R}^q_{++}$ due to 1) and 2) the function $P(x,\lambda^*,\mathbf{k}) = \mathcal{L}(x,\lambda^*,\mathbf{k}) - f(x^*)$ is an exact smooth approximation of $Q(x,x^*)$ at the solution $x^*$, i. e.

$$P(x^*,\lambda^*,\mathbf{k}) = \min\{P(x,\lambda^*,\mathbf{k}) \mid x \in \mathbb{R}^n\} = Q(x^*,x^*) = 0 \quad \forall \mathbf{k} \in \mathbb{R}^q_{++}.$$

The Log-barrier function $F(x,k) = f(x) - k^{-1}\sum_{i=1}^q \ln c_i(x)$ one can also view as an approximation for $Q(x,x^*)$ at $x = x^*$, however, for any fixed $k > 0$

$$\lim_{x\to x^*} (F(x,k) - Q(x^*,x^*)) = \infty.$$

Therefore there is only one way to guarantee convergence of the minimizers

$$x(k) = \operatorname{argmin}\{F(x,k) \mid x\in\mathbb{R}^n\}$$

to the solution $x^*$: unbounded increase of the penalty parameter $k > 0$. It leads to the well known numerical problems, which for NLP calculations have much stronger effect than for LP.

On the other hand, the LT $\mathcal{L}(x,\lambda,\mathbf{k})$ along with the penalty parameter $k > 0$ has two extra tools: the vector of the Lagrange multipliers $\lambda\in\mathbb{R}^q_{++}$ and the scaling parameter vector $\mathbf{k} \in \mathbb{R}^q_{++}$. We will show later that the LT method generates a primal-dual sequence converging to the primal-dual solution under any fixed penalty parameter $k > 0$. Moreover, using properly all three tools it becomes possible to develop a primal-dual LT method with a quadratic rate of convergence. This is the main purpose of the paper.

It follows from 3) (see (3.3)) that the LT Hessian $\nabla_{xx}\mathcal{L}(x^*,\lambda^*,\mathbf{k})$ is identical to the Hessian of the Quadratic Augmented Lagrangean (see [11], [23]–[25]) corresponding to the active constraints set. Moreover, due to Debreu lemma, (see [7]) under the standard second order optimality condition (2.3)–(2.4) the LT Hessian $\nabla^2_{xx}\mathcal{L}(x^*,\lambda^*,\mathbf{k})$ is positive definite for all $k \geq k_0$ if $k_0 > 0$ is large enough, whether $f(x)$ and all $-c_i(x)$ are convex or not. This is another important property of the LT $\mathcal{L}(x,\lambda,\mathbf{k})$ allowing to extend some of the obtained result for nonconvex optimization problems.

Now we are ready to introduce the LT method.

Let $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^q_{++}$, $k > 0$ and $\mathbf{k}^0 = (k_i^0 = k(\lambda_i^0)^{-2},\ i = 1,\dots,q)$. The LT multipliers method maps the triple $(x^s,\lambda^s,\mathbf{k}^s)$ into $(x^{s+1},\lambda^{s+1},\mathbf{k}^{s+1})$ defined by the following formulas:

$$x^{s+1} = \operatorname{argmin}\{\mathcal{L}(x,\lambda^s,\mathbf{k}^s) \mid x \in \mathbb{R}^n\}, \tag{3.4}$$

$$\lambda_i^{s+1} = \lambda_i^s\psi'(k_i^s\lambda_i^s c_i(x^{s+1})) = \lambda_i^s\psi'(k(\lambda_i^s)^{-1}c_i(x^{s+1})), \quad i = 1,\dots,q, \tag{3.5}$$

$$k_i^{s+1} = k(\lambda_i^{s+1})^{-2}, \quad i = 1,\dots,q. \tag{3.6}$$

**Theorem 3.1.** *If assumption* $\mathbf{A}$ *is satisfied then*

*1) the LT method (3.4)–(3.6) is well defined;*

*2) the LT method (3.4)–(3.6) is equivalent to an Interior Quadratic Prox for the dual problem;*

*3) the LT method (3.4)–(3.6) is equivalent to an Interior Prox method with second order $\varphi$-divergence distance function.*

*Proof.* 1) As we have mentioned already the existence of $x^{s+1}$ for any given $\lambda^s \in \mathbb{R}^q_{++}$ and $\mathbf{k}^s \in \mathbb{R}^q_{++}$ follows from assumption **A**, convexity of $f(x)$, concavity of $c_i(x)$ and properties $3^0$ and $4^0$ of the transformation $\psi\in\Psi$. Therefore the LT method (3.4)–(3.6) is well defined.

2) From (3.4) and (3.5) we obtain

$$\nabla_x \mathcal{L}(x^{s+1},\lambda^s,\mathbf{k}^s) = \nabla f(x^{s+1}) - \sum \lambda_i^s \psi'(k(\lambda_i^s)^{-1} c_i(x^{s+1}))\nabla c_i(x^{s+1}) = \nabla_x L(x^{s+1},\lambda^{s+1}) = 0. \tag{3.7}$$

From (3.5) and $2^0$a) follows $\lambda^{s+1} = (\lambda_i^{s+1},\dots,\lambda_q^{s+1}) \in \mathbb{R}^q_{++}$. Therefore from (3.6) we have $\mathbf{k}^{s+1} = (k_1^{s+1},\dots,k_q^{s+1}) \in \mathbb{R}^q_{++}$. It follows from (3.7) that

$$x^{s+1} = \operatorname{argmin}\{L(x,\lambda^{s+1}) \mid x \in \mathbb{R}^n\}, \tag{3.8}$$

so $d(\lambda^{s+1}) = L(x^{s+1},\lambda^{s+1})$ and $-c(x^{s+1}) \in \partial d(\lambda^{s+1})$, where $\partial d(\lambda^{s+1})$ is subdifferential of $d(\lambda)$ at $\lambda = \lambda^{s+1}$.

From (3.5), $2^0$b) and the mean value formula we obtain

$$\lambda_i^{s+1} - \lambda_i^s = \lambda_i^s(\psi'(k_i^s\lambda_i^s c_i(x^{s+1})) - \psi'(0)) = k_i^s(\lambda_i^s)^2\psi''(\theta_i^s k_i^s \lambda_i^s c_i(x^{s+1}))c_i(x^{s+1}) = k_i^s(\lambda_i^s)^2 \psi''_{[s,i]}(\cdot)c_i(x^{s+1}), \quad i = 1,\dots,q,$$

where $0 < \theta_i^s < 1$.

Using (3.6) we have

$$\lambda^{s+1} = \lambda^s + k\Psi''_{[s]}(\cdot)c(x^{s+1}), \tag{3.9}$$

where $\Psi''_{[s]}(\cdot) = \operatorname{diag}(\psi''_{[s,i]}(\cdot))_{i=1}^q$.

The equation (3.9) can be rewritten as follows:

$$-c(x^{s+1}) + k^{-1}(\Psi''_{[s]}(\cdot))^{-1}(\lambda^{s+1} - \lambda^s) = 0. \tag{3.10}$$

Keeping in mind $-c(x^{s+1}) \in \partial d(\lambda^{s+1})$ the equation (3.10) is the optimality criteria for the vector $\lambda^{s+1}$ to be a maximizer in the following problem:

$$d(\lambda^{s+1}) = \max\{d(\lambda) - \frac{1}{2}k^{-1}\sum(-\psi''_{[s,i]}(\cdot))^{-1}(\lambda_i - \lambda_i^s)^2 \mid \lambda \in \mathbb{R}^q\} = \max\{d(\lambda) - \frac{1}{2}k^{-1}||\lambda - \lambda^s||^2_{R_s} \mid \lambda \in \mathbb{R}^q\}, \tag{3.11}$$

where $R_s = (-\Psi''_{[s]}(\cdot))^{-1}$ and $\|\lambda\|^2_{R_s} = \lambda^T R_s \lambda$.

Therefore the LT method (3.4)–(3.6) is equivalent to the Quadratic Prox method (3.11) for the dual problem in the rescaled dual space. Keeping in mind that Quadratic Prox (3.11) generates positive dual sequence $\{\lambda^s\} \in \mathbb{R}^q_{++}$ we conclude that LT method (3.4)–(3.6) is equivalent to Interior Quadratic Prox (3.11).

3) The formula for the Lagrange multipliers update can be rewritten as follows:

$$k_i^s\lambda_i^s c_i(x^{s+1}) = {\psi'}^{-1}(\lambda_i^{s+1}/\lambda_i^s), \quad i = 1,\dots,q. \tag{3.12}$$

The inverse $\psi^{-1}$ exists due to $3^0$. Using ${\psi'}^{-1} = {\psi^*}'$, formula (3.6) and assuming $\varphi= -\psi^*$, we can rewrite (3.12) as follows:

$$-c_i(x^{s+1}) = k^{-1}\lambda_i^s\varphi'(\lambda_i^{s+1}/\lambda_i^s). \tag{3.13}$$

Again, keeping in mind $-c(x^{s+1}) \in \partial d(\lambda^{s+1})$ from (3.13) we obtain

$$d(\lambda^{s+1}) = \max\{d(\lambda) - k^{-1}\sum_{i=1}^{q}(\lambda_i^s)^2\varphi(\lambda_i/\lambda_i^s) \mid \lambda \in \mathbb{R}^q\}. \tag{3.14}$$

Due to $\lim_{t\to\infty}\psi'(t) = 0$, which follows from $2^0$c), for the kernel $\varphi$ we obtain $\lim_{t\to 0_+}\varphi'(t) = -\infty$, hence $\lambda^{s+1} \in \mathbb{R}^q_{++}$.

In other words the LT method (3.4)–(3.6) is equivalent to the Interior Prox method

$$d(\lambda^{s+1}) = \max\{d(\lambda) - k^{-1}D(\lambda, \lambda^s) \mid \lambda\in\mathbb{R}^q\}, \tag{3.15}$$

where $D:\ \mathbb{R}^q_+ \times \mathbb{R}^q_{++} \to \mathbb{R}_+$ ,given by 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 $\varphi:\ \mathbb{R}_{++} \to \mathbb{R}_+$. □

The method (3.11) reminds the classical quadratic prox (see [6], [10], [13], [14], [25]). On the other hand, the LT method (3.4)–(3.6) is an exterior-interior point method. The LT method is an exterior-point method in the primal space and an interior-point mathod in the dual space. The Interior Prox method (3.15) has been studied in [2], [3], [20], [27].

The properties of the kernels $\varphi$ are induced by properties $1^0$–$5^0$ of the original transformation $\psi\in\Psi$. They were established in the following theorem.

**Theorem 3.2.** [20] *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) $\lim_{s\to 0+}\varphi'(s) = -\infty$; b) $\varphi'(s)$ is monotone increasing; c) $\varphi'(1) = 0$;*

*3) a) $\varphi''(s) \geq m > 0\ \forall s \in (0,\infty)$; b) $\varphi''(s) \leq M < \infty\ \forall s \in [1,\infty)$.*

Unfortunately several well known transformations including exponential $\psi_1(t) = 1 - e^{-t}$ [12], logarithmic $\psi_2(t) = \ln(t+1)$ and hyperbolic $\psi_3(t) = t(t+1)^{-1}$ MBF [17] as well as log-sigmoid $\psi_4(t) = 2(\ln 2 + t - \ln(1+e^t))$ and modified CHKS transformation $\psi_5(t) = t - \sqrt{t^2+4\eta} + 2\sqrt{\eta}$, $\eta>0$ ([20]) do not satisfy $1^0$–$5^0$. Transformations $\psi_1$–$\psi_3$ do not satisfy property $3^0$ ($m = 0$), while for $\psi_4$ and $\psi_5$ the property $4^0$ is violated ($M = \infty$). It can be fixed however by using the quadratic extrapolation idea, which first was introduced in [5] for the logarithmic MBF transformation $\psi_2$. For a given $-1 < \tau < 0$ we modified the transformations $\psi_1$–$\psi_5$ as follows:

$$\psi_j(t) := \begin{cases} \psi_j(t), & t \geq \tau, \\ q_j(t) & t \leq \tau, \end{cases} \qquad i = 1,\ldots,5, \tag{3.16}$$

where $q_j(t) = a_jt^2 + b_jt + c_j$ and the coefficients $a_j = 0.5\psi_j''(\tau)$, $b_j = \psi_j'(\tau) - \tau\psi_j''(\tau)$, $c_j = \psi_j(\tau) - \tau\psi_j'(\tau) + 0.5\tau^2\psi_j''(\tau)$ are defined from the system

$$\psi_j(\tau) = q_j(\tau), \qquad \psi_j'(\tau) = q_j'(\tau), \qquad \psi_j''(\tau) = q_j''(\tau).$$

One can check directly that $\psi_j$, $j = 1,\ldots,5$ given by (3.16) belong to $\Psi$. Their Fenchel conjugate are given by formula

$$\psi_j^*(s) := \begin{cases} \psi_i^*(s), & s \leq \psi'(\tau), \\ q^*(s), & s \geq \psi'(\tau), \end{cases} \qquad j = 1,\ldots,5.$$

Therefore due to Theorem 3.2 kernels $\varphi_i = -\psi_i^* \in \Phi$.

In [2] to guarantee 3a) the authors regularized the logarithmic MBF kernel $\psi_2(s) = s - \ln s - 1$. The regularized logarithmic MBF kernel $\bar{\varphi}_2(s) = 0.5\nu(s-1)^2 + \mu(s - \ln s - 1)$, $\mu > 0$, $\nu > 0$ has some very interesting properties allowing us to prove the global convergence of the dual sequence $\{\lambda^s\}$ generated by the Interior Prox method (3.15) to the dual solution $\lambda^*$ with $O((ks)^{-1})$ rate (see [2]). The Fenchel transform $\bar{\varphi}_2^*$ of the kernel $\bar{\varphi}_2$ leads to primal transformation $\bar{\psi}_2 = -\bar{\varphi}_2^*$, which satisfy properties $1^0$–$5^0$, therefore such transformation along with those given by (3.16) can be used in the framework of primal-dual LT method, which we develop in sections 5 and 6.

## 4. Convergence and rate of convergence of the LT method

In this section we present convergence results for the general LT method. The key component of the convergence proof is the equivalence of the LT method (3.4)–(3.6) to the Interior Quadratic Prox (3.11) for the dual problem.

Let $\lambda^0 \in \mathbb{R}^q_{++}$ be the initial vector of Lagrange multipliers and $d = d(\lambda^*) - d(\lambda^0)$. The dual level set $\Lambda_0 = \{\lambda \in \mathbb{R}^q_+ :\ d(\lambda) \geq d(\lambda^0)\}$ is bounded due to concavity of $d(\lambda)$ and boundedness of $L^*$ (see Corollary 20 in [8]). For $x \in \mathbb{R}^n$ we consider two sets of indices: $I^-(x) = \{i :\ c_i(x) < 0\}$ and $I^+(x) = \{i :\ c_i(x) \geq 0\}$. We introduce the maximum constraint violation

$$v_l = \max\{-c_i(x^l) \mid i \in I^-(x^l)\}$$

and the upper bound

$$d_l = \sum_{i=1}^q \lambda_i^l |c_i(x^l)| \geq \sum_{i=1}^q \lambda_i^l c_i(x^l)$$

for the duality gap at the step 1.

Let $\bar{v}_s = \min_{1 \leq l \leq s} v_l$, $\bar{d}_s = \min_{1 \leq l \leq s} d_l$. For a bounded close set $Y \subset \mathbb{R}^n$ and $y_0 \notin Y$ we consider the distance $\rho(y_0, Y) = \min\{\|y_0 - y\| \mid y \in Y\}$ from $y_0$ to $Y$.

**Theorem 4.1.** *If the assumptions* **A** *and* **B** *are satisfied then*

*1) the primal-dual sequence $\{x^s, \lambda^s\}$ is bounded, and the following estimations hold:*

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

*2) for the constraints violation and the duality gap the following bounds hold:*

$$\bar{v}_s = O((ks)^{-0.5}), \qquad \bar{d}_s = O((sk)^{-0.5});$$

*3) the primal-dual sequence $\{x^s, \lambda^s\}$ converges to the primal-dual solution in value, i. e.*

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

*and*

$$\lim_{s \to \infty} \rho(x^s, X^*) = 0, \quad \lim_{s \to \infty} \rho(\lambda^s, L^*) = 0;$$

*besides, any converging primal-dual subsequence has the primal-dual solution as a limit point.*

*Proof.* 1) From the concavity of $d(\lambda)$ and $-c(x^{s+1}) \in \partial d(\lambda^{s+1})$ we have

$$d(\lambda) - d(\lambda^{s+1}) \le (-c(x^{s+1}), \lambda - \lambda^{s+1})$$

or

$$d(\lambda^{s+1}) - d(\lambda) \ge (c(x^{s+1}), \lambda - \lambda^{s+1}). \tag{4.1}$$

Using (3.5) and (3.6) we have

$$c_i(x^{s+1}) = (k_i^s \lambda_i^s)^{-1} \psi'^{-1}(\lambda_i^{s+1}/\lambda_i^s) = k^{-1}\lambda_i^s \psi'^{-1}(\lambda_i^{s+1}/\lambda_i^s).$$

Using $\psi'^{-1} = \psi^{*\prime}$ we obtain

$$c_i(x^{s+1}) = k^{-1}\lambda_i^s \psi^{*\prime}(\lambda_i^{s+1}/\lambda_i^s), \quad i, \dots, q. \tag{4.2}$$

Keeping in mind $\psi^{*\prime}(1) = \psi^{*\prime}(\lambda_i^s/\lambda_i^s) = 0$ from (4.1) for $\lambda = \lambda^s$ we obtain

$$d(\lambda^{s+1}) - d(\lambda^s) \ge \sum_{i=1}^m k^{-1}\lambda_i^s(\psi^{*\prime}(\lambda_i^{s+1}/\lambda_i^s) - \psi^{*\prime}(\lambda_i^s/\lambda_i^s))(\lambda_i^s - \lambda_i^{s+1}). \tag{4.3}$$

Using the mean value formula we have

$$\psi^{*\prime}(\lambda_i^{s+1}/\lambda_i^s) - \psi^{*\prime}(\lambda_i^s/\lambda_i^s) = -\psi^{*\prime\prime}(\cdot)(\lambda_i^s)^{-1}(\lambda_i^s - \lambda_i^{s+1}) = \varphi''(\cdot)(\lambda_i^s)^{-1}(\lambda_i^s - \lambda_i^{s+1}).$$

Using 3(a) from Theorem 3.2 and (4.3) we obtain

$$d(\lambda^{s+1}) - d(\lambda^s) \ge mk^{-1}\|\lambda^s - \lambda^{s+1}\|^2. \tag{4.4}$$

Let $i \in I^-(x^{s+1}) = \{i : \ c_i(x^{s+1}) < 0\}$. Using $\psi^{*\prime}(1) = \psi^{*\prime}(\lambda_i^s/\lambda_i^s) = 0$, the mean value formula, the equation (4.2) and 3(b) from Theorem 3.2 we obtain

$$-c_i(x^{s+1}) = k^{-1}\lambda_i^s[\psi^{*\prime}(\lambda_i^s/\lambda_i^s) - \psi^{*\prime}(\lambda_i^{s+1}/\lambda_i^s)]$$
$$= k^{-1}(-\psi^{*\prime\prime}(\cdot))(\lambda_i^{s+1} - \lambda_i^s) \le k^{-1}\varphi''(\cdot)|\lambda_i^{s+1} - \lambda_i^s| \le k^{-1}M|\lambda_i^{s+1} - \lambda_i^s|,$$

or

$$|\lambda_i^{s+1} - \lambda_i^s| \ge kM^{-1}(-c_i(x^{s+1})), \quad i \in I^-(x^{s+1}).$$

Combining the last inequality with (4.4) we obtain

$$d(\lambda^{s+1}) - d(\lambda^s) \ge kmM^{-2} \sum_{i \in I^-(x^{s+1})} c_i^2(x^{s+1}). \tag{4.5}$$

2) From (4.5) we have

$$d(\lambda^{s+1}) - d(\lambda^s) \ge kmM^{-2}v_{s+1}^2. \tag{4.6}$$

Summing up (4.5) from $l = 0$ to $l = s$ we obtain

$$d = d(\lambda^*) - d(\lambda^0) \ge d(\lambda^{s+1}) - d(\lambda^0) \ge kmM^{-2}\sum_{l=0}^s v_{l+1}^2.$$

Keeping in mind that $\bar{v}_s = \min\{v_l \mid 1 \le l \le s\}$, we obtain

$$\bar{v}_s \le M\sqrt{dm^{-1}}(ks)^{-0.5} = O((ks)^{-0.5}). \tag{4.7}$$

The primal asymptotic feasibility follows from $v_l \to 0$.

The bound similar to (4.7) for the duality gap $\bar{d}_s$ can be established using arguments similar to those in [20].

3) From assumption **B** follows boundedness of $L^*$. From boundedness of $L^*$ and concavity of $d(\lambda)$ follows boundedness of $L_0 = \{\lambda \in \mathbb{R}_+^q : \ d(\lambda) \ge d(\lambda^0)\}$ (see Corollary 20 in [8]).

From (4.4) follows the dual monotonicity $d(\lambda^{s+1}) \geq d(\lambda^s) + mk^{-1}\|\lambda^{s+1} - \lambda^s\|^2$, $s \geq 0$. Also $d(\lambda^s) \leq f(x^*)$, $s \geq 0$, therefore there is $d(\lambda^\infty) = \lim_{s\to\infty} d(\lambda^s)$ and

$$\lim_{s\to\infty} (d(\lambda^{s+1}) - d(\lambda^s)) = 0. \tag{4.8}$$

The dual sequence $\{\lambda^s\}$ is bounded, therefore there is a converging subsequence $\{\lambda^{s_l}\}: \lim_{s\to\infty} \lambda^{s_l} = \bar{\lambda}$.

Due to (4.4) and (4.8) we have

$$\lim_{s\to\infty} \lambda^{s_l+1} = \bar{\lambda}. \tag{4.9}$$

The boundedness of the primal sequence $\{x^s\}$ follows from the boundedness of $X^*$ and $\lim_{l\to\infty} v_l = 0$ (see Corollary 20 in [8]).

From (3.4)–(3.5) we have

$$\nabla_x L(x^{s+1}, \lambda^s, k^s) = \nabla_x L(x^{s+1}, \lambda^{s+1}) = 0. \tag{4.10}$$

Without loss of generality we can assume that

$$\lim_{s\to\infty} x^{s_l+1} = \bar{x}.$$

By passing to limit (4.10) we obtain

$$\lim_{s_l\to\infty} \nabla_x L(x^{s_l+1}, \lambda^{s_l+1}) = \nabla_x L(\bar{x}, \bar{\lambda}) = 0.$$

We consider two sets of indices $I_+ = \{i: \bar{\lambda}_i > 0\}$ and $I_0 = \{i: \bar{\lambda}_i = 0\}$.

From (3.5)–(3.6) we have

$$c_i(x^{s_l+1}) = k^{-1}\lambda_i^{s_l}\psi'^{-1}(\lambda_i^{s_l+1}/\lambda_i^{s_l}), \quad i \in I_+. \tag{4.11}$$

By passing (4.11) to the limit and keeping in mind 2c) from Theorem 3.2 we obtain

$$c_i(\bar{x}) = k^{-1}\bar{\lambda}_i\varphi'(1) = 0, \quad i \in I_+.$$

Due to $v_s \to 0$ for any $i \in \bar{I}_0$ we have

$$\lim_{s_l\to\infty} c_i(x^{s_l+1}) = c_i(\bar{x}) \geq 0.$$

Therefore for the pair $(\bar{x}, \bar{\lambda})$ the KKT's conditions are satisfied, i. e.

$$\nabla_x L(\bar{x}, \bar{\lambda}) = 0, \quad \bar{\lambda} \in \mathbb{R}^q_+, \ c(\bar{x}) \in \mathbb{R}^q_+, \quad (\bar{\lambda}, c(\bar{x})) = 0,$$

therefore $\bar{x} = x^*$, $\bar{\lambda} = \lambda^*$. In view of the dual monotonicity we have

$$\lim_{s\to\infty} d(\lambda^s) = d(\bar{\lambda}) = d(\lambda^*).$$

Therefore,

$$d(\lambda^*) = \lim_{s\to\infty} d(\lambda^s) = \lim_{s\to\infty} (f(x^s) - \sum_{i=1}^q \lambda_i^s c_i(x^s)). \tag{4.12}$$

To complete the proof we have to show the asymptotic complementarity condition

$$\lim_{s\to\infty} \sum_{i=1}^q \lambda_i^s c_i(x^s) = 0. \tag{4.13}$$

From $\lambda^s \in \mathbb{R}^q_{++}$ and $v_s \to 0$ we have $\lim_{s\to\infty} \lambda_i^s c_i(x^s) \geq 0$. Let us assume the opposite to (4.13), i.e. that there is $i \in \{1,\ldots,q\}$ such that for a primal-dual subsequence $\{x^{s_l}, \lambda^{s_l}\}$ we have

$$\lim_{s\to\infty} \lambda_i^{s_l} c_i(x^{s_l}) = \rho > 0.$$

Without loosing the generality we can assume that $\lim_{s_l\to\infty} x^{s_l} = \bar{x}$, $\lim_{s_l\to\infty} \lambda^{s_l} = \bar{\lambda}_i > 0$. Then $\bar{\lambda}_i c_i(\bar{x}) = \rho > 0$, therefore $c_i(\bar{x}) > 0$ and $\bar{\lambda}_i = \rho (c_i(\bar{x}))^{-1} > 0$, which is impossible because for any $i \in I_+$ we have $c_i(\bar{x}) = 0$.

From (4.12)–(4.13) we have

$$d(\lambda^*) = \lim_{s\to\infty} d(\lambda^s) = \lim_{s\to\infty} (f(x^s) - \sum_{i=1}^q \lambda_i^s c_i(x^s)) = \lim_{s\to\infty} f(x^s) = f(x^*). \tag{4.14}$$

Keeping in mind that $X^* = \{x \in \Omega : \ f(x) \leq f(x^*)\}$, from primal asymptotic feasibility and (4.14) we obtain $\lim_{s\to\infty} \rho(x^s, X^*) = 0$ (see [19] Lemma 11, Chap. 9, §1).

Taking in to account $L^* = \{\lambda \in \mathbb{R}^r_{++} : d(\lambda) \geq d(\lambda^*)\}$ and using again(4.14) and Lemma 11 we obtain $\lim_{s\to\infty} \rho(\lambda^s, L^*) = 0$.

We completed the proof of the Theorem 4.1. □

**Remark 4.1.** *It follows from (4.5) that for any $\tau < 0$ and any $i = 1,\ldots,q$ the inequality $c_i(x^{s+1}) \leq \tau$ is possible only for a finite number of steps. Therefore from some point on only original transformations can be used in the LT method. In fact for $k > 0$ large enough the quadratic branch can be used just once. Therefore, the asymptotic analysis and the numerical performance of the LT method (3.4)–(3.6) and its dual equivalent (3.11) or (3.15) depends only on the properties of the original transformations $\psi_1$–$\psi_5$ and the correspondent 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 LT $\mathcal{L}(x, \lambda, \mathbf{k})$ if the input data has the corresponding property. It allows to use the Newton method for solving the Primal-Dual system, which is equivalent to (3.4)–(3.5). We will concentrate on it in section 5.*

Each transformation $\psi_j \in \Psi, j = 1,\ldots,5$ leads to a particular second order Entropy-like distance function $D_j(u,v) = \sum_{i=1}^q v_i^2 \varphi_j(u_i/v_i)$. Each distance function $D_j(u,v)$ leads to a corresponding Interior Prox method (3.15) for finding 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 method (3.15) (see [3]). However, when $d(\lambda)$ is the dual function for the dual problem $(D)$, such analysis can produce only limited results, because neither the primal nor the dual sequence controls the LT method. The LT method is rather governed by the PD system solving which is equivalent to LT step.

The PD system is defined by the primal-dual map similar to those we used to establish the rate of convergence of nonlinear rescaling methods(see [17]–[19]). Using the corresponding primal-dual map we can strengthen the convergence results of Theorem 4.1 by assuming the standard second order optimality conditions.

From (3.6) we have $\lim_{s\to\infty} k_i^s = k(\lambda_i^*)^{-2}$, $i = 1,\ldots,r$, i. e. the scaling parameters corresponding to the active constraints grow linearly with $k > 0$. Therefore the technique we used in [17], [19] can be applied for the asymptotic analysis of the method (3.4)–(3.6).

For a given small enough $\delta>0$, we define the following set:

$$D(\lambda^*,\mathbf{k},\delta)=\{(\lambda,\mathbf{k})\in\mathbb{R}^q_+\times\mathbb{R}^q_{++}:\ \lambda_i\geq\delta,\ |\lambda_i-\lambda_i^*|\leq\delta k,\quad i=1,\dots,r,$$
$$0\leq\lambda_i\leq k\delta,\ k\geq k_0,\quad i=r+1,\dots,q;\quad k_i=k\lambda_i^{-2},\quad i=1,\dots,q\}.$$

**Theorem 4.2.** *If $f(x)$, $c_i(x)\in C^2$ and the standard second order optimality conditions (2.3)–(2.4) hold, then there exists such small $\delta>0$ and large enough $k_0>0$ that for any $(\lambda,\mathbf{k})\in D(\cdot)$ we have:*

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

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

*and*

$$\widehat{\lambda}_i=\lambda_i\psi''(k(\lambda_i)^{-1}c_i(\widehat{x})),\quad \widehat{k}_i=k\widehat{\lambda}_i^{-2},\quad i=1,\dots,q.$$

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

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

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

*3) The LT $\mathcal{L}(x,\lambda,\mathbf{k})$ is strongly convex in the neighborhood of $\widehat{x}$.*

Theorem 4.2 can be proved by a slight modification of the correspondent proof of Theorem 1 in [17] (see also [19]).

**Remark 4.2.** *All results of Theorem 4.2 do not require convexity of $f(x)$ and all $-c_i(x)$, $i=1,\dots,q$. Therefore the LT method can be used for solving nonconvex optimization problems. In fact, it is enough to find $\widehat{x}$ just ones for $k>0$ large enough before the LT method starts finding minimums of strongly convex functions at each step. To find the first unconstrained minimizer for a wide class of nonconvex functions one can use very interesting cubic regularization of Newton's method recently developed in [15].*

Finding $x^{s+1}$ requires solving an unconstrained minimization problem (3.4), which is, generally speaking, an infinite procedure. The stopping criteria (see [19]–[21]) 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 $x\in\mathbb{R}^n$. If $\bar{x}^{s+1}$ is used in the formula (3.5) for the Lagrange multipliers update then bounds similar those established in 2) of Theorem 4.2 remain true.

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

$$\bar{x}^{s+1}:\ \|\nabla_x\mathcal{L}(\bar{x}^{s+1},\bar{\lambda}^s,\bar{\mathbf{k}}^s)\|\leq\sigma k^{-1}\|\Psi'(k(\bar{\lambda}^s)^{-1}c(\bar{x}^{s+1}))\bar{\lambda}^s-\bar{\lambda}^s\|, \tag{4.15}$$

$$\bar{\lambda}^{s+1}=\Psi'(k(\bar{\lambda}^s)^{-1}c(\bar{x}^{s+1}))\bar{\lambda}^s, \tag{4.16}$$

where

$$\Psi'(k(\bar{\lambda}^s)^{-1}c(\bar{x}^{s+1}))=\operatorname{diag}(\psi'(k(\bar{\lambda}^s_i)^{-1}c_i(\bar{x}^{s+1}))_{i=1}^q$$

and

$$\bar{\mathbf{k}}_s=(\bar{k}_i^s=k(\bar{\lambda}_i^s)^{-2},\quad i=1,\dots,q).$$

The following theorem can be proven the same way we proved Theorem 7.1 in [19].

**Theorem 4.3.** *If the standard second order optimality conditions (2.1)–(2.3) hold and the Hessians* $\nabla^2 f(x)$ *and* $\nabla^2 c_i(x),\ i=1,\dots,m$ *satisfy the Lipschitz conditions*

(4.17)
$$\|\nabla^2 f(x)-\nabla^2 f(y)\|\le L_0\|x-y\|,\quad \|\nabla^2 c_i(x)-\nabla^2 c_i(y)\|\le L_i\|x-y\|,\quad i=1,\dots,q,$$

*then there is* $k_0>0$ *large enough such that for the primal-dual sequence* $\{\bar{x}^s,\bar{\lambda}^s\}$ *generated by the formulas (4.15)–(4.16) the following estimations hold true and* $c>0$ *is independent from* $k\ge k_0$ *for* $s\ge 0$*:*

$$\|\bar{x}^{s+1}-x^*\|\le c(1+\sigma)k^{-1}\|\bar{\lambda}^s-\lambda^*\|,\quad \|\bar{\lambda}^{s+1}-\lambda^*\|\le c(1+\sigma)k^{-1}\|\bar{\lambda}^s-\lambda^*\|. \tag{4.18}$$

To find an approximation $\bar{x}^{s+1}$ one can use Newton's method with steplength for minimization $\mathcal{L}(x,\bar{\lambda}^s,\mathbf{k}^s)$ in $x$. It requires generally speaking several Newton steps to find $\bar{x}^{s+1}$. Then we update the vector of Lagrange multipliers $\bar{\lambda}^s$ and the scaling parameters vector $\bar{\mathbf{k}}^s$ using $\bar{x}^{s+1}$ instead of $x^{s+1}$ in (3.5) and $\bar{\lambda}^{s+1}$ instead of $\lambda^{s+1}$ in (3.6).

In the next section we develop a different approach. Instead of finding $\bar{x}^{s+1}$ and then updating the Lagrange multipliers we consider a primal-dual system, solving which is equivalent to finding $\bar{x}^{s+1}$ and $\bar{\lambda}^{s+1}$. Newton's method for solving the primal-dual system that is equivalent to (3.4)–(3.5) leads to the Primal-Dual LT method.

## 5. Local Primal-Dual LT method

In this section we describe the PDLT method and prove local quadratic rate of convergence of the primal-dual sequence to the primal-dual solution under the standard second order optimality condition. One step of the LT method (3.4)–(3.6) maps the given triple $(x,\lambda,\mathbf{k})\in\mathbb{R}^n\times\mathbb{R}^q_{++}\times\mathbb{R}^q_{++}$ into a triple $(\widehat{x},\widehat{\lambda},\widehat{\mathbf{k}})\in\mathbb{R}^n\times\mathbb{R}^q_{++}\times\mathbb{R}^q_{++}$ by formulas

$$\begin{aligned}\widehat{x}:\ \nabla_x\mathcal{L}(\widehat{x},\lambda,\mathbf{k})&=\nabla f(\widehat{x})-\sum\psi'(k_i\lambda_i c_i(\widehat{x}))\lambda_i\nabla c_i(x)\\&=\nabla f(\widehat{x})-\sum\widehat{\lambda}_i\nabla c_i(\widehat{x})=0,\end{aligned} \tag{5.1}$$

$$\widehat{\lambda}:\ \widehat{\lambda}_i=\lambda_i\psi'(k\lambda_i^{-1}c_i(\widehat{x})),\quad i=1,\dots,q, \tag{5.2}$$

$$\widehat{\mathbf{k}}:\ \widehat{k}_i=k\widehat{\lambda}_i^{-2},\quad \dots,q. \tag{5.3}$$

By removing the scaling vector update formula (5 3) from the system (5.1)–(5.3), we obtain the primal-dual LT system

$$\nabla_x L(\widehat{x},\widehat{\lambda})=\nabla f(\widehat{x})-\sum_{i=1}^{q}\widehat{\lambda}\nabla c_i(\widehat{x})=0, \tag{5.4}$$

$$\widehat{\lambda}=\Psi'(k\lambda^{-1}c(\widehat{x}))\lambda, \tag{5.5}$$

where $\Psi'(k\lambda^{-1}c(\widehat{x}))=\mathrm{diag}(\psi'(k_i\lambda_i^{-1}c_i(\widehat{x})))_{i=1}^q$.

From the standard second order optimality condition (2.3)–(2.4) follows the uniqueness of $x^*$ and $\lambda^*$. Also there is $\tau^*>0$ that

a) $\min\{c_i(x^*)\mid r+1\le i\le q\}\ge\tau^*$ and b) $\min\{\lambda_i^*\mid 1\le i\le r\}\ge\tau^*$.

Therefore due to (4.18) there is $k_0>0$ large enough that for any $k\ge k_0$ and $s\ge 1$

$$\text{a) } \min\{c_i(\bar{x}^s)\mid r+1\le i\le q\}\ge 0.5\tau^*\quad\text{and}\quad\text{b) } \min\{\bar{\lambda}_i^s\mid 1\le i\le r\}\ge 0.5\tau^*. \tag{5.6}$$

Using formula (3.5) and the property $2^0$c) we have

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

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

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

So for a given accuracy $0 < \varepsilon << 1$ 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 on they will be automatically ignored in the further calculations together with the part of LT related to the passive constraints. Therefore the primal-dual system (5.4)–(5.5) will be actually reduced to the following system for $\widehat{x}$ and $\widehat{\lambda} = (\widehat{\lambda}_1, \ldots, \widehat{\lambda}_r)$:

$$\nabla_x L(\widehat{x}, \widehat{\lambda}) = \nabla f(\widehat{x}) - \sum_{i=1}^{r} \widehat{\lambda}_i \nabla c_i(\widehat{x}) = 0, \tag{5.7}$$

$$\widehat{\lambda}_i = \psi'(k(\lambda_i)^{-1} c_i(\widehat{x}))\lambda_i, \quad i = 1, \ldots, r. \tag{5.8}$$

To simplify 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. $c^T(x) = (c_1(x), \ldots, c_r(x))$.

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

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

assuming that the input data is properly normalized. It follows from the KKT condition (2.1)–(2.2) that

$$\nu(x, \lambda) = 0 \quad \Leftrightarrow \quad x = x^*,\ \lambda = \lambda^*.$$

Later we will use the following lemma.

**Lemma 5.1.** [22] *Under the standard second order optimality condition (2.3)–(2.4) and Lipschitz condition (4.17) there exists* $0 < m_0 < M_0 < \infty$ *and* $\varepsilon_0 > 0$ *small enough that*

$$m_0 \|y - y^*\| \leq \nu(y) \leq M_0 \|y - y^*\| \quad \forall y \in \Omega_{\varepsilon_0}. \tag{5.9}$$

It follows from (5.9) that in the neigbourhood $\Omega_{\varepsilon_0}$ the merit function $\nu(y)$ is similar to $\|\nabla f(x)\|$ for unconstrained optimization problem $\min\{f(x) \mid x \in \mathbb{R}^n\}$. The merit function $\nu(y)$ 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 PDLT step;
4) to decide whether the primal or primal-dual direction has to be used at the current step.

First we consider Newton's method for solving system (5.7)–(5.8) and show its local quadratic convergence. To find the Newton direction $(\Delta x, \Delta \lambda)$ we have to linearize the system (5.7)–(5.8) at $y = (x, \lambda)$.

We start with system (5.8). Due to $3^0$ there exists the inverse $\psi'^{-1}$. Therefore using the identity $\psi'^{-1} = \psi^{*\prime}$ we can rewrite (5.8) as follows:

$$c_i(\widehat{x}) = k^{-1}\lambda_i\psi'^{-1}(\widehat{\lambda}_i/\lambda_i) = k^{-1}\lambda_i\psi^{*\prime}(\widehat{\lambda}_i/\lambda_i) = -k^{-1}\lambda_i\varphi'(\widehat{\lambda}_i/\lambda_i).$$

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

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

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 (5.7) 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 terms of the second and higher orders we obtain the following linearization of the PD system (5.7)–(5.8):

$$\nabla^2_{xx}L(x,\lambda)\Delta x - \nabla c^T(x)\Delta\lambda = -\nabla_x L(x,\lambda), \tag{5.10}$$

$$\nabla c(x)\Delta x + k^{-1}\varphi''(1)I^r\Delta\lambda = -c(x), \tag{5.11}$$

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

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

Then the system (5.10)–(5.11) can be written as follows:

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

The local PDLT method consists of the following operations:

1. Find the primal-dual Newton direction $\Delta y = (\Delta x, \Delta\lambda)$ from the system

$$N_k(\cdot)\begin{bmatrix} \Delta x \\ \Delta\lambda \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) \\ -c(\cdot) \end{bmatrix}. \tag{5.12}$$

2. Find the new primal-dual vector $\widehat{y} = (\widehat{x},\widehat{\lambda})$ by formulas

$$\widehat{x} := x + \Delta x, \quad \widehat{\lambda} := \lambda + \Delta\lambda. \tag{5.13}$$

3. Update the scaling parameter

$$\widehat{k} = (\nu(\widehat{y}))^{-1}. \tag{5.14}$$

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

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

We will use later the following technical Lemma.

**Lemma 5.2.** [22] *Let $A:\ \mathbb{R}^n \to \mathbb{R}^n$ be an invertible matrix and $\|A^{-1}\| \le c_0$ then for small enough $\varepsilon > 0$ any $B:\ \mathbb{R}^n - \mathbb{R}^n$ such that $\|A - B\| \le \varepsilon$ is invertible and the following bounds hold:*

$$\text{a)} \quad \|B^{-1}\| \le 2c_0 \quad \text{and} \quad \text{b)} \quad \|A^{-1} - B^{-1}\| \le 2c_0^2\varepsilon. \tag{5.15}$$

**Lemma 5.3.** *If the standard second order optimality conditions (2.3)–(2.4) and the Lipschitz conditions (4.17) are satisfied then there exists small enough* $\varepsilon_0 > 0$ *and large enough* $k_0 > 0$ *that both matrices* $N_\infty(y)$ *and* $N_k(y)$ *are non-singular and there is* $c_0 > 0$ *independent on* $y \in \Omega_{\varepsilon_0}$ *and* $k \geq k_0$ *that*

$$\max\{\|N_\infty^{-1}(y)\| \|N_k^{-1}(y)\|\} \leq 2c_0 \quad \forall y \in \Omega_{\varepsilon_0} \quad \text{and} \quad \forall k \geq k_0. \tag{5.16}$$

*Proof.* It is well known (see for example [16 ]) that under the standard second order optimality conditions (2.3)–(2.4) the matrix

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

is non-singular, hence there exists $c_0 > 0$ that $\|N_\infty^{-1}(y^*)\| \leq c_0$. Due to Lipschitz condition (4.17) there exists $L > 0$ 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 any small enough $\varepsilon > 0$ there exists such small $\varepsilon_0 > 0$ and large $k_0 > 0$ that

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

Applying Lemma 5.2 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 (5.16). □

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

**Theorem 5.1.** *If the standard second order optimality conditions (2.3)–(2.4) and the Lipschitz condition (4.17) are satisfied then there exists* $\varepsilon_0 > 0$ *small enough that for any primal-dual pair* $y = (x, \lambda) \in \Omega_{\varepsilon_0}$ *the PDLT methods (5.12)–(5.14) generates the primal-dual sequence that converges to the primal-dual solution with quadratic rate, i. e. , the following bound holds:*

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

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

*Proof.* The primal-dual Newton direction $\Delta y = (\Delta x, \Delta\lambda)$ we find from the system

$$N_k(y)\Delta y = b(y), \tag{5.17}$$

where

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

Along with the primal-dual system (5.7)–(5.8) we consider the Lagrange system of equations, which corresponds to the active constraints at the same point $y = (x, \lambda)$:

$$\nabla_x L(x, \lambda) = \nabla f(x) - \nabla c(x)^T \lambda = 0, \tag{5.18}$$

$$c(x) = 0. \tag{5.19}$$

We apply Newton method for solving (5.18)–(5.19) from the same starting point $y = (x, \lambda)$. The Newton directions $\Delta\bar{y} = (\Delta\bar{x}, \Delta\bar{\lambda})$ for (5.18)–(5.19) we find from the following system of linear equations:

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

The new approximation for the system (5.18)–(5.19) we obtain 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 (2.3)–(2.4) and the Lipschitz conditions (4.17) there is $c_1 > 0$ independent on $y \in \Omega_{\varepsilon_0}$ that the following bounds holds (see Theorem 9, Ch. 8 [16]):

$$\|\bar{y} - y^*\| \leq c_1 \|y - y^*\|^2. \tag{5.20}$$

Now we can prove the similar bound for $\|\widehat{y} - y^*\|$. We have

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

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 5.3 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 5.2 with $A = N_k(y), B = N_\infty(y)$ we obtain

$$\|\Delta y - \Delta\bar{y}\| \leq 2k^{-1}\varphi''(1)c_0^2\|b(y)\|. \tag{5.21}$$

In view of $\nabla_x L(x^*, \lambda^*) = 0, c(x^*) = 0$ and the Lipschitz condition (4.17) we have

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

Using (5.9), (5.14) and (5.21) we obtain

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

Therefore for $c_2 = 2\varphi''(1)c_0^2 M_0 L$, which is independent on $y \in \Omega_{\varepsilon_0}$, we have

$$\|\Delta y - \Delta\bar{y}\| \leq c_2\|y - y^*\|^2. \tag{5.22}$$

Using (5.20) and (5.22) for $c = 2\max\{c_1, c_2\}$ we obtain

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

and $c = \max\{c_1, c_2\} > 0$ is independent on $y \in \Omega_{\varepsilon_0}$. We completed the proof. □

## 6. Primal-Dual LT method

In this section we describe the globally convergent PDLT method.The globally convergent PDLT method roughly speaking works as the Newton LT multipliers method (4.15)–(4.16) in the initial phase and as the primal-dual LT method (5.12)–(5.14) in the final phase of the computational process.

Each step of PDLT consists of finding the primal-dual direction $\Delta y = (\Delta x, \Delta\lambda)$ by solving the linearized primal-dual system (5.4)–(5.5). Then we use either the primal-dual Newton direction $\Delta y$ to find a new primal-dual vector or the primal Newton direction $\Delta x$ for minimization of $\mathcal{L}(x, \lambda, \mathbf{k})$ in $x$.

The choice at each step depends on the merit function $\nu(y)$ value and how the value changes after one step. If the primal-dual step produces quadratic reduction of the merit function then the primal-dual 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 system (5.4)–(5.5) is linearized. Let us start with $y = (x, \lambda)$ and compute $\nu(y)$. By linearizing the system (5.4) we obtain

$$\nabla^2_{xx} L(x, \lambda)\Delta x - \nabla c^T(x)\Delta\lambda = -\nabla_x L(x, \lambda). \tag{6.1}$$

The system (5.5) we split into two sub-systems. The first is associated with the set $I_+(y) = \{i: \ \lambda_i > \nu(y)\}$ of "big" Lagrange multipliers, while the second is associated with the set $I_0(y) = \{i: \ \lambda_i \le \nu(y)\}$ of "small" Lagrange multipliers. Therefore, $I_+(y) \cap I_0(y) = \varnothing$ and $I_+(y) \cup I_0(y) = \{1, \dots, q\}$. We consider two sub-systems:

$$\widehat{\lambda}_i = \psi'(k\lambda_i^{-1} c_i(\widehat{x}))\lambda_i, \quad i \in I_+(y), \tag{6.2}$$

$$\widehat{\lambda}_i = \psi'(k\lambda_i^{-1} c_i(\widehat{x}))\lambda_i, \quad i \in I_0(y). \tag{6.3}$$

The equations (6.2) can be rewritten as follows:

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

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

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

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

$$c_i(x) + \nabla c_i(x)\Delta x = -k^{-1}\varphi''(1)\Delta\lambda_i, \quad i \in I_+(y). \tag{6.4}$$

Let $c_+(x)$ be the vector-function associated with "big" Lagrange multipliers, i. e. $c_+(x) = (c_i(x), \ i \in I_+(y))$, $\nabla c_+(x) = J(c_+(x))$ is the correspondent Jacobian and $\Delta\lambda_+ = (\Delta\lambda_i, \ i \in I_+(y))$ is the dual Newton direction associated with "big" Lagrange multipliers. Then the system (6.4) can be rewritten as follows:

$$\nabla c_+(x)\Delta x + k^{-1}\varphi''(1)\Delta\lambda_+ = -c_+(x). \tag{6.5}$$

Now let us linearize the system (6.3). Ignoring terms of second and higher order we obtain

$$\begin{aligned} \widehat{\lambda}_i = \lambda_i + \Delta\lambda_i &= \psi'(k\lambda_i^{-1}(c_i(x) + \nabla c_i(x)\Delta x))\lambda_i \\ &= \psi'(k\lambda_i^{-1} c_i(x))\lambda_i + k\psi''(k\lambda_i^{-1} c_i(x))\Delta c_i(x)\Delta x \\ &= \bar{\lambda}_i + k\psi''(k\lambda_i^{-1} c_i(x))\nabla c_i(x)\Delta x, \quad i \in I_0(y). \end{aligned} \tag{6.6}$$

Let $c_0(x)$ be the vector-function associated with "small" Lagrange multiplier, $\nabla c_0(x) = J(c_0(x))$ the correspondent Jacobian, $\lambda_0 = (\lambda_i, \ i \in I_0(y))$ is vector of "small" Lagrange multipliers and $\Delta\lambda_0 = (\lambda_i, \ i \in I_0(y))$ is the correspondent dual Newton direction. Then (6.6) can be rewritten as follows:

$$-k\Psi''(k\lambda_0^{-1} c_0(x))\Delta c_0(x)\Delta x + \Delta\lambda_0 = \bar{\lambda}_0 - \lambda_0, \tag{6.7}$$

where

$$\begin{aligned} \bar{\lambda}_0 &= \Psi'(k\lambda_0^{-1} c_0(x))\lambda_0, \\ \Psi'(k\lambda_0^{-1} c_0(x)) &= \mathrm{diag}(\psi'(k\lambda_i^{-1} c_i(x)))_{i \in I_0(y)}, \\ \Psi''(k\lambda_0^{-1} c_0(x)) &= \mathrm{diag}(\psi''(k\lambda_i^{-1} c_i(x)))_{i \in I_0(y)}. \end{aligned}$$

Combining (6.1), (6.6), (6.7) we obtain the following system for finding the primal-dual direction $\Delta y = (\Delta x, \Delta\lambda)$, where $\Delta\lambda = (\Delta\lambda_+, \Delta\lambda_0)$ and $I_B$ and $I_S$ are identical matrices in spaces of "big" and "small" Lagrange multipliers:

$$(6.8)\quad M(x,\lambda)\Delta y = \begin{bmatrix} \nabla^2_{xx}L(x,\lambda) & -\nabla c_+^T(x) & -\nabla c_0^T(x) \\ \nabla c_+(x) & k^{-1}\varphi''(1)I_B & 0 \\ -k\Psi''(k\lambda_0^{-1}c_0(x))\nabla c_0(x) & 0 & I_S \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda_+ \\ \Delta\lambda_0 \end{bmatrix} = \begin{bmatrix} -\nabla_x L(x,\lambda) \\ -c_+(x) \\ \bar{\lambda}_0 - \lambda_0 \end{bmatrix}.$$

To guarantee the existence of the primal-dual LT direction $\Delta y$ for any $(x,\lambda)\in \mathbb{R}^n\times\mathbb{R}^q_+$ we replace the system (6.8) by the following regularized system where $I^n$ identical matrix in $\mathbb{R}^n$:

$$(6.9)\quad M_k(x,\lambda)\Delta y = \begin{bmatrix} \nabla^2_{xx}L(x,\lambda)+k^{-1}I^n & -\nabla c_+^T(x) & -\nabla c_0^T(x) \\ \nabla c_+(x) & k^{-1}\varphi''(1)I_B & 0 \\ -k\Psi''(k\lambda_0^{-1}c_0(x))\nabla c_0(x) & 0 & I_S \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda_+ \\ \Delta\lambda_0 \end{bmatrix} = \begin{bmatrix} -\nabla_x L(x,\lambda) \\ -c_+(x) \\ \bar{\lambda}_0 - \lambda_0 \end{bmatrix}.$$

Finding the primal-dual direction $\delta y$ from the system (6.9) we call $PDLTD(x,\lambda)$ procedure.

Now we are ready to describe the PDLT method.

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

**Step 2:** If $\nu\le\varepsilon$ then stop. **Output:** $x,\lambda$.

**Step 3:** Find direction: $(\Delta x,\Delta\lambda):=\mathrm{PDLTD}(x,\lambda_c)$, $\lambda:=\lambda_c$. Set $\widehat{x}:=x+\Delta x$, $\widehat{\lambda}:=\lambda+\Delta\lambda$.

**Step 4:** If $\nu(\widehat{x},\widehat{\lambda})\le\min\{\nu^{2-\theta},1-\theta\}$, set $x:=\widehat{x}$, $\lambda:=\widehat{\lambda}$, $\nu:=\nu(x,\lambda)$, $k:=\max\{\nu^{-1},k\}$, Goto Step 2.

**Step 5:** Decrease $t\le1$ until $\mathcal{L}(x+t\Delta x,\lambda_c,\mathbf{k})-\mathcal{L}(x,\lambda_c,\mathbf{k})\le\eta t(\Delta\mathcal{L}(x,\lambda_c,\mathbf{k}),\Delta x)$.

**Step 6:** Set $x:=x+t\Delta x$, $\widehat{\lambda}:=\Psi'(k\lambda_c^{-1}c(x))\lambda_c$.

**Step 7:** If $\|\Delta_x\mathcal{L}(x,\lambda_c,\mathbf{k})\|\le\frac{\sigma}{k}\|\widehat{\lambda}-\lambda_c\|$, Goto Step 9.

**Step 8:** Find direction: $(\Delta x,\Delta\lambda):=\mathrm{PDLTD}(x,\lambda_c)$, Goto Step 5.

**Step 9:** If $\nu(x,\widehat{\lambda})\le\nu^{2-\theta}$, set $\lambda_c:=\widehat{\lambda}$, $\lambda:=\lambda_c$, $\nu:=\nu(x,\lambda)$, $k:=\max\{\nu^{-1},k\}$, $\mathbf{k}:=(k_i=k\lambda_i^{-2},\ i=1,\dots,q)$, Goto Step 2.

**Step 10:** Set $k:=k\alpha$, Goto step 8.

The matrix $M_k(y)\equiv M_k(x,\lambda)$ is nonsingular 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 LT direction $\Delta y$ can be found for any $y=(x,\lambda)\in\mathbb{R}^n\times\mathbb{R}^q_+$ and any $k>0$.

It follows from (5.9) that for $\forall y \notin \Omega_{\varepsilon_0}$ there is $\tau > 0$ that $\nu(y) \geq \tau$, therefore from (5.14) we have $k^{-1} = \nu(y) \geq \tau$.

After finding $\Delta\lambda_+$ and $\Delta\lambda_0$ from the second and third system in (6.9) and substituting their value into the first system we obtain

$$P_k(y)\Delta x \equiv P_k(x,\lambda)\Delta x = -\nabla_x L(x,\bar{\lambda}) = -\nabla_x \mathcal{L}(x,\lambda,k), \tag{6.10}$$

where

$$P_k(y) = \nabla^2_{xx} L(x,\lambda) + k^{-1} I^n$$
$$+ k(\varphi''(1))^{-1} \nabla c_+^T(x) \nabla c_+(x) - k \nabla c_0^T(x) \Psi''(k\lambda_0^{-1} c_0(x)) \nabla c_0(x)$$

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

Using the arguments similar to those we used in case of $M_k(y)$ we conclude that the symmetric matrix $P_k(y)$ is positive definite. Moreover due to $k^{-1} \geq \tau > 0$ the matrix $P_k(y)$ has uniformly bounded from below mineigval $P_k(y) \geq \tau > 0$ $\forall y \notin \Omega_{\varepsilon_0}$. For any $y \in \Omega_{\varepsilon_0}$ the mineigvalue $P_k(y) \geq \rho > 0$ due to Debreu's lemma [7], the standard second order optimality condition (2.2)–(2.4) and the Lipschitz condition (4.17). Therefore the primal Newton direction $\Delta x$ defined by (6.9) or (6.10) is a descent direction for minimization $\mathcal{L}(x,\lambda,\mathbf{k})$ in x. Therefore for $0 < \eta \leq 0.5$ we can find $t \geq t_0 > 0$ that

$$\mathcal{L}(x + t\Delta x, \lambda, \mathbf{k}) - \mathcal{L}(x,\lambda,\mathbf{k}) \leq \eta t(\nabla \mathcal{L}(x,\lambda,\mathbf{k}), \Delta x) \leq -\tau t\eta \|\Delta x\|_2^2. \tag{6.11}$$

On the other hand, due to the boundedness of the primal-dual sequence and the Lipschitz conditions (4.17) there exists such $\bar{M} > 0$ that $||\nabla_{xx} \mathcal{L}(x,\lambda,\mathbf{k})|| \leq \bar{M}$. Hence, the primal sequence generated by Newton's method $x := x + t\Delta x$ with $t > 0$ defined from (6. 11) converges to $\widehat{x} = \widehat{x}(\lambda,\mathbf{k}) :\ \nabla_x \mathcal{L}(\widehat{x},\lambda,\mathbf{k}) = \nabla_x L(\widehat{x},\widehat{\lambda}) = 0$.

Under the standard second order optimality condition according to Theorem 4.3 we can find $\bar{x}^{s+1}$ from (4.15) in finite number of Newton steps and update the Lagrange multipliers by (4.16). Due to (4.18) after $s_0 = O(\ln \varepsilon_0^{-1})$ updates we find the primal-dual approximation $y \in \Omega_{\varepsilon_0}$.

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

Keeping in mind properties $2^0$c), $2^0$d) of the transformation $\psi \in \Psi$ as well as (5.6), (5.9) and (5.14) after $s_1 = O(\ln \ln \varepsilon^{-1})$ updates we obtain

$$\max\{||k\Psi''(k\lambda_0^{-1} c_0(x))||, ||\bar{\lambda}_0 - \lambda_0||\} = o(\varepsilon^2), \quad i \in I_0. \tag{6.12}$$

For any $y \in \Omega_{\varepsilon_0}$ the term $\|\nabla c_0(x)\|$ is bounded. The boundedness of $\|\Delta x\|$ follows from boundedness of $\|\nabla_x L(x,\bar{\lambda})\|$ and the fact that $P_k(y)$ has a mineigenvalue bounded from below by a positive number uniformly in $y$.

Let us consider the third part of the system (6.9), that is associated with the "small" Lagrange multipliers

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

It follows from (6.12) that $\|\Delta\lambda_0\| = o(\varepsilon^2)$. It means that after $s = \max\{s_0, s_1\}$ updates the part of the system (6.9) associated with "small" Lagrange multipliers became irrelevant for the calculation of a Newton direction from (6.9). In fact, the system (6.9) reduces into the following system:

$$\bar{M}_k(x,\lambda)\Delta\bar{y} = \bar{b}(x,\lambda), \tag{6.13}$$

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

$$\bar{M}_k(x,\lambda) = \begin{bmatrix} \nabla^2_{xx} L(x,\lambda) + k^{-1} I^n & -\nabla c^T_{(+)}(x) \\ \nabla c_{(+)}(x) & k^{-1}\varphi''_{(1)} I^+ \end{bmatrix}.$$

At this point we have $y{\in}\Omega_{\varepsilon_0}$, therefore it follows from (5.9) that $\nu(y){\le}M_0\varepsilon_0$. Hence for small enough $\varepsilon_0 > 0$ from $|\lambda_i - \lambda^*_i|{\le}\varepsilon_0$ we obtain $\lambda_i{\ge}\nu(y)$, $i{\in}I^*$. On the other hand we have $\nu(y) > \lambda_i = O(\varepsilon^2)$, $i{\in}I_0$, otherwise we obtain $\nu(y){\le}O(\varepsilon^2)$ and from (5.9) follows $\|y - y^*\| = o(\varepsilon^2)$. So, if after $s = \max\{s_0, s_1\}$ Lagrange multipliers updates we have not solve the problem with a given accuracy $\varepsilon > 0$ then $I_+(y) = I^*$ and $I_0(y) = I^*_0 = \{r+1, \dots, q\}$ and we continue to perform the PDLT (5.12)–(5.14) using

$$\bar{M}_k(x,\lambda) = \begin{bmatrix} \nabla^2_{xx} L(x,\lambda) + k^{-1} I^n & -\nabla c^T_{(r)}(x) \\ \nabla c_{(r)}(x) & k^{-1}\varphi''(1) I^r \end{bmatrix}$$

instead of $N_k(\cdot)$.

Therefore we have

$$||\Delta y - \Delta\bar{y}|| = ||(\bar{M}^{-1}_k(y) - N^{-1}_\infty(y))b(y)|| \le ||\bar{M}^{-1}_k(y) - N^{-1}_\infty(y)||||b(y)||.$$

On the other hand $\|\bar{M}_k(y) - N_\infty(y)\| \le k^{-1}(1 + \varphi''(1))$. From Lemma 5.3 we have $\max\{\|\bar{M}^{-1}_k(y)\|, \|N^{-1}_\infty(y)\|\}{\le}2c_0$. Keeping in mind(5.9), (5.14) (4.17) we obtain the following estimation:

$$\begin{aligned} (6.14)\quad & \|\Delta y - \Delta\bar{y}\|{\le}2{c_0}^2 k^{-1}(1 + \varphi''(1))\|\bar{b}(y)\| \\ & = 2c_0^2\nu(y)(1 + \varphi''(1))\|\bar{b}(y)\|{\le}2(1 + \varphi''(1))c_0^2 M_0 L\|y - y^*\|^2 = c_3\|y - y^*\|^2, \end{aligned}$$

where $c_3 > 0$ is independent on $y{\in}S(y^*, \varepsilon_0)$.

In other words, by finding $\Delta y$ from (6.13) instead of (5.13) we do not compromise a quadratic rate of convergence of the PDLT method. Therefore the following theorem takes place.

**Theorem 6.1.** *If the standard second order optimality conditions are satisfied and the Hessian $\nabla^2 f(x)$ and all Hessians $\nabla^2 c_i(x)$, $i = 1, \dots, q$ satisfy Lipschitz condition, then PDLT method generates globally convergent primal-dual sequence that converges to the primal-dual solution with asymptotic quadratic rate.*

## 7. Concluding remarks

The PDLT method (5.12)–(5.14) is fundamentally different from the Newton LT multipliers method. The distinct characteristics of the PDLT method is its global convergence with an asymptotic quadratic rate. The PDLT method combines the best features of the Newton LT method and the Newton method for the Lagrange system of equations corresponding to the active constraints. At the same time the PDLT method is free from their critical drawbacks. In the intial phase the PDLT method performs similar to Newton's method for LT minimization followed by Lagrange multipliers and scaling parameters update. Such method converges under a fixed penalty parameter. This allows us to reach the $\varepsilon_0$-neighborhood of the primal-dual solution in $O(ln\varepsilon_0^{-1})$ Lagrange multipliers update without compromising the condition number of the LT Hessian.

On the other hand, once in the neighborhood of the primal-dual 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 uncounstrained minimization the PDLT solves the primal-dual LT system. Moreover, the primal-dual direction $\Delta y$ becomes very close to the Newton direction (see(6.14)) for the Lagrange system of equations corresponding to the active constraints. This guarantees the asymptotic quadratic convergence.

The situation reminds the one of Newton's method with steplength for unconstrained smooth optimization.

The way we regularize the Lagrangian Hessian (see (6.9)) allows us on the one hand to guarantee the global convergence, on the other hand to avoid compromising an asymptotic quadratic rate of convergence.

Several issues remain for future research. First, the neighborhood of the primal-dual solution where the quadratic rate of convergence occurs needs to be characterized using parameters that measure the nondegeneracy of the constrained optimization problems.

Second, the value of the scaling parameter $k_0 > 0$ is *a priori* unknown and depends on the condition number (see [17]) measuring the nondegeneracy of a constrained optimization problem. This number could be expressed using parameters of a constrained optimization problem at the solution, which are obviously unknown. Therefore it is important to find an efficient way to change the penalty parameter $k > 0$ using the merit function value.

Third, it is important to understand to what extent the PDLT method can be used in the non-convex case. In this regard recent results from [15] together with local convexity properties of the LT that follows from Debreu's lemma [7] may play an important role.

Fourth, numerical experiments using various versions of the primal-dual NR methods produce very encouraging results (see [9], [21] and [21]). On the other hand PDLT method has certain specific features that requires more numerical work to understand better its practical efficiency.

## References


1. A. Auslender, R. Cominetti and M. Haddou. 1997. " Asymptotic analysis of penalty and barrier methods in convex and linear programming. " Math. Oper. Res. , vol. 22, pp. 43–62.
2. A. Auslender, M. Tebboule and S. Ben-Tiba. 1999. "Interior proximal and multipliers methods based on second-order homogeneous kernels. " Math. Oper. Res. , vol. 24, no. 3, pp. 645–668.
3. A. Auslender, M. Tebboule. 2004. "Interior Gradient and Epsilon-Subgradient Descent Methods for constrained Convex Minimization" Math. Oper. Res. Vol 29, no1, pp1-26.
4. A. Ben-Tal and M. Zibulevsky. 1997. "Penalty-barrier methods for convex programming problems. " SIAM J. Optim. , vol. 7, pp. 347–366.
5. A. Ben-Tal, B. Yuzefovich and M. Zibulevsky. 1992. "Penalty-barrier multipliers methods for minimax and constrained smooth convex optimization. " Optimization Laboratory, Technion, Israel. Research Report 9-92.
6. D. Bertsekas. 1982. Constrained Optimization and Lagrange Multipliers Methods. Academic Press: New York.
7. G. Debreu. 1952. "Definite and Semidefinite Quadratic Forms, Econometrica Vol. 20, P295-300.

8. A. Fiacco and G. McCormick. 1990. Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Classics in Applied Mathematics. SIAM: Philadelphia, PA.
9. O. Guler. 1991. “On the convergence of the proximal point algorithm for convex minimization. ” SIAM J. Control Optim. , vol. 29, pp. 403–419.
10. I. Griva. 2004. “Numerical experiments with an interior-exterior point method for Nonlinear Programming”. Computational Optimization and Applications, vol. 29 (2), pp. 173–195.
11. M. R. Hestenes. 1969. “Multipliers and gradient methods. ” J. Optim. Theory Appl. , vol. 4, pp. 303–320.
12. B. W. Kort and D. P. Bertsekas. 1973. “Multiplier methods for convex programming, ” in Proc. IEEE Conf. on Decision and Control. San Diego, CA. , pp. 428–432.
13. B. Martinet. 1970. “Regularization d’inequations variationelles par approximations successive. ” Revue Francaise d’Automatique et Informatique Rechershe Operationelle, vol. 4, pp. 154–159.
14. J. Moreau. 1965. “Proximité et dualité dans un espace Hilbertien. ” Bull. Soc. Math. France, vol. 93, pp. 273–299.
15. Y. Nesterov, B. Polyak. 2003. Cubic regularization of a Newton scheme and its global performance. Discussion paper UCL, center for operations research and econometrics. universite catholique de LovAin .
16. B. Polyak, B. 1987. Introduction to Optimization. Software Inc. : NY.
17. R. Polyak. 1992. “Modified barrier functions. ” Math. Programming, vol. 54, pp. 177–222.
18. P. Polyak. 1988. ”Smooth optimization methods for minmax problems” SIAM Journal on control and optimization vol 26 no 6, pp 1274-1286.
19. R. Polyak. 2001. “Log-sigmoid multipliers method in constrained optimization. ” Ann. Oper. Res. , vol. 101, pp. 427–460.
20. R. Polyak. 2002. “Nonlinear rescaling vs. smoothing technique in convex optimization. ” Math. Programming, vol. 92, 197-235.
21. R. Polyak and I. Griva. 2004. “Primal-dual nonlinear rescaling methods for convex optimization, ”JOTA:vol 122, N 1 PP 111-156.
22. I. Griva, R. Polyak ”Primal-Dual nonlinear rescaling method with dynamic scaling parameter update” Math. Programming, to appear 2004
23. M. J. D. Powell. 1969. “A method for nonlinear constraints in minimization problems, ” in Fletcher, ed. , Optimization, London Academic Press, pp. 283–298.
24. R. T. Rockafellar. 1973. “A dual approach to solving nonlinear programming problems by unconstrainted minimization. ” Math. Programming, vol. 5, pp. 354–373.
25. R. T. Rockafellar. 1976. “Augmented Lagrangians and applications of the proximal points algorithms in convex programming. ” Math. Oper. Res. , vol. 1, pp. 97–116.
26. D. F. Shanno and R. J. Vanderbei. 1999. “An interior point algorithm for nonconvex nonlinear programming. ” COAP, vol. 13, pp. 231–252.

27. P. Tseng and D. Bertsekas. 1993. “On the convergence of the exponential multipliers method for convex programming. ” Math. Programming, vol. 60, pp. 1–19.

Department of SEOR and Mathematical Sciences Department, George Mason University

*E-mail address*: rpolyak@gmu.edu