Roman A. Polyak⋆

Nonlinear rescaling vs. smoothing technique in convex optimization

Abstract. We introduce an alternative to the smoothing technique approach for constrained optimization. As it turns out for any given smoothing function there exists a modification with particular properties. We use the modification for Nonlinear Rescaling (NR) the constraints of a given constrained optimization problem into an equivalent set of constraints.

The constraints transformation is scaled by a vector of positive parameters. The Lagrangian for the equivalent problems is to the correspondent Smoothing Penalty functions as Augmented Lagrangian to the Classical Penalty function or MBFs to the Barrier Functions. Moreover the Lagrangians for the equivalent problems combine the best properties of Quadratic and Nonquadratic Augmented Lagrangians and at the same time are free from their main drawbacks.

Sequential unconstrained minimization of the Lagrangian for the equivalent problem in primal space followed by both Lagrange multipliers and scaling parameters update leads to a new class of NR multipliers methods, which are equivalent to the Interior Quadratic Prox methods for the dual problem.

We proved convergence and estimate the rate of convergence of the NR multipliers method under very mild assumptions on the input data. We also estimate the rate of convergence under various assumptions on the input data.

In particular, under the standard second order optimality conditions the NR method converges with Q-linear rate without unbounded increase of the scaling parameters, which correspond to the active constraints.

We also established global quadratic convergence of the NR methods for Linear Programming with unique dual solution.

We provide numerical results, which strongly support the theory.

Key words. smoothing technique – nonlinear rescaling – multipliers method – Interior Prox method – Log-Sigmoid transformation – duality – Fermi-Dirac Entropy function

1. Introduction

The smoothing technique for constrained optimization employs a smooth approximation of $x_+ = \max\{0, x\}$ to transform a constrained optimization problem into a sequence of unconstrained optimization problems. The sequence of the unconstrained minimizers converges to the primal solution due to the unbounded increase of the scaling parameter. Therefore the smoothing technique is in fact a penalty type method with a smooth penalty function.

There are few well-known difficulties associated with penalty type approach: rather slow convergence, the Hessian of the penalty function becomes ill conditioned and the area where Newton method is "well defined" shrinks to a point when the scaling parameter infinitely grows.

R.A. Polyak: Department of SEOR & Mathematical Sciences Department, George Mason University, Fairfax VA 22030, USA, e-mail: rpolyak@gmu.edu

It motivates an alternative approach. It turns out that for any given smoothing function there exists a modification with particular properties. We use the modification to transform the constraints of a given convex optimization problem into an equivalent set of constraints. The constraints are scaled by a vector of positive scaling parameters.

Sequential unconstrained minimization of the Lagrangians for the equivalent problem in primal space followed by explicit formulas for both the Lagrange multipliers and the scaling parameters update leads to NR multipliers methods.

For the scaling parameters update we use the formula suggested by P. Tseng and D. Bertsekas in [37] for exponential multipliers method. We will show that NR multipliers methods with such "dynamic" scaling parameters update combine the best properties of Quadratic and Nonquadratic multipliers method and at the same time they are free from their most essential drawbacks.

In the first part of the paper we introduced the Log-Sigmoid Transformation (LST)

$$\psi(t) = 2 \ln 2 S(t, 1),$$

where $S(t, 1) = (1 + exp(-t))^{-1}$ is the Sigmoid Function, which is often used, in neural network literature, see e.g., [18]. We used LST to illustrate the basic ideas and techniques.

The transformation $\psi(t)$ is a modification of the Log-Sigmoid smoothing function, which has been recently used by C. Chen and O. Mangasarian [8] for solving convex inequalities and Linear Complementarity problems and by A. Auslender et al. [1] for constrained optimization.

The Lagrangian for the equivalent problem – Log-Sigmoid Lagrangian (LSL) is our main instrument in the first part of the paper.

There are four basic reasons for using LS transformation and the corresponding Lagrangian:

1) $\psi \in C^\infty$ on $(-\infty, \infty)$,
2) the LSL is a C^∞ function assuming the data possesses similar smoothing properties,
3) ψ'' are bounded on $(-\infty, \infty)$,
4) due to the properties of ψ and its Fenchel conjugate ψ^* the LSL enjoys the best qualities of both Quadratic and Nonquadratic Augmented Lagrangians and at the same time is free from their most essential drawbacks.

The convergence analysis for multipliers method with "dynamic" scaling parameters update has proven to be surprisingly difficult (see [37], p. 3). Just recently in [3] was proven that for a particular class of transformations ψ the correspondent method generates bounded primal and dual sequences that each of their limit points is a pair of optimal primal and dual solutions. The results in [3] was extended in [2] establishing that inexact proximal version of the multipliers method with "dynamic" scaling parameters update generates a bounded sequence with every limit point being an optimal solution. Moreover, in [2] the convergence of the Dual Interior Prox method with the second order φ-divergence kernel was proven and the rate of convergence was estimated for the regularized dual MBF kernel.

Our first contribution is a new convergence proof and estimation of the rate of convergence of the LS multipliers method with "dynamic" scaling parameters update under very mild assumptions on the input data.

The second contribution is the estimation of the rate of convergence under the strict complementarity condition. As it turns out LS method is equivalent to the Quadratic Prox method in the truncated rescaled dual space, which is defined by the dual variables that correspond to the active constraints. Therefore the Interior Prox method for the Dual problem has all the best qualities of the Quadratic Prox method [11], [20], [22], [24], [31], [32] and on the top of it keeps positive the Lagrange multipliers.

Under the standard second order optimality condition the LS multipliers method converges with Q-linear rate without unbounded increase of the scaling parameters which correspond to the active constraints.

The key ingredients of the convergence proof and the rate of convergence estimation are the equivalence of the LS multipliers method to the Interior Prox method with the second order Entropy-like φ–divergence distance function for the dual problem and the properties of the dual kernel $\varphi = -\psi^*$, where ψ^* is the Fenchel conjugate of the Log-Sigmoid function ψ, which happened to be Fermi-Dirac Entropy function. In particular, the strong concavity of the Fermi-Dirac Entropy function ψ^* is critical for both the convergence and the rate of convergence.

We show in this paper that the convergence and the rate of convergence under minimum assumptions on the input data for a wide class of constraints transformations including exponential, logarithmic, hyperbolic and parabolic MBF can be obtained by using the same considerations we used to prove the convergence of the LS multipliers method.

In the second part of the paper we extend our analysis on a broad class of smoothing functions. In particular, we introduce the modification of the Chen–Harker–Kanzow–Smale (CHKS) smoothing function and show that the basic LS results remain true for the CHKS's modification as well as for modifications of the exponential and barrier transformations, which lead to logarithmic, hyperbolic and parabolic MBFs.

Finally, we show that LS method for LP has global quadratic convergence if the dual LP has a unique solution.

The paper is organized as follows:

The problem formulation and the basic assumptions are given in the next section.

The LS transformation and its modification we consider in Sect. 3.

The Log-Sigmoid Lagrangian and its properties both global and local we consider in Sect. 4.

We prove the equivalence of the LS multipliers method to the Interior Prox method for the dual problem in Sect. 5.

In Sect. 6 we consider the convergence of the LS multipliers method and its dual equivalent.

In Sect. 7 we estimate the rate of convergence under different assumptions on the input data.

In Sect. 8 we extend the LS results for a general class of smoothing functions. In particular, we specify the results for the transformation, which corresponds to CHKS interior smoothing function.

In Sect. 9 we apply LS method for LP and prove global quadratic rate of convergence.

In the Sect. 10 we provide some numerical results, which strongly corroborate the theory.

We conclude the paper with some remarks related to the future research.

2. Statement of the problem and basic assumptions

Let $f : \mathbb{R}^n \to \mathbb{R}^1$ be convex and all $c_i : \mathbb{R}^n \to \mathbb{R}^1$, $i = 1, \ldots, p$ be concave and smooth functions. We consider a convex set $\Omega = \{x \in \mathbb{R}^n_+ : c_i(x) \geq 0, \ i = 1, \ldots, p\}$ and the following convex optimization problem

$$(\mathcal{P}) \qquad\qquad x^* \in X^* = Argmin\{f(x)|x \in \Omega\}$$

We assume that:

A. The optimal set X^* is not empty and bounded.
B. The Slater's condition holds, i.e. there exists $\hat{x} \in \mathbb{R}^n_{++} : c_i(\hat{x}) > 0, \ i = 1, \ldots, p$.

To simplify consideration we will include the nonnegativity constraints $x_i \geq 0$, $i = 1, \ldots, n$ into a set $c_i(x) \geq 0$, i.e.

$$\Omega = \big\{x \in \mathbb{R}^n : c_i(x) \geq 0, \ i = 1, \ldots, p,$$

$$c_{p+1}(x) = x_1 \geq 0, \ldots, c_{p+n}(x) = x_n \geq 0\big\}$$

$$= \big\{x \in \mathbb{R}^n : c_i(x) \geq 0, \ i = 1, \ldots, m\big\},$$

where $m = p + n$.

If **B** holds, then the Karush-Kuhn-Tucker's (K-K-T's) conditions hold true, i.e. there exists vector $\lambda^* = (\lambda_1^*, ..., \lambda_m^*) \in \mathbb{R}^m_+$ such that

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

$$\lambda_i^* c_i(x^*) = 0, \ i = 1, \ldots, m, \tag{2}$$

where $L(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i c_i(x)$ is the Lagrangian for the primal problem P. Also due to **B**, the K-K-T's set

$$L^* = \big\{\lambda \in \mathbb{R}^m_+ : \nabla f(x^*) - \sum_{i=1}^m \lambda_i \nabla c_i(x^*) = 0, \ x^* \in X^*\big\} \tag{3}$$

is bounded.

Let's assume that the active constraint set at x^* is $I^* = \{i : c_i(x^*) = 0\} = \{1, \ldots, r\}$. We consider the vectors functions $c^T(x) = (c_1(x), \ldots, c_m(x))$, $c_{(r)}^T(x) = (c_1(x), \ldots, c_r(x))$, and their Jacobians $\nabla c(x) = J(c(x))$ and $\nabla c_{(r)}(x) = J(c_{(r)}(x))$. The sufficient regularity conditions

$$rank \nabla c_{(r)}(x^*) = r, \lambda_i^* > 0, \ i \in I^* \tag{4}$$

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

$$\big(\nabla_{xx}^2 L(x^*, \lambda^*)y, y\big) \geq \rho(y, y), \ \rho > 0, \forall y \neq 0 : c_{(r)}(x^*)y = 0 \tag{5}$$

comprise the standard second order optimality conditions.

3. Log-Sigmoid transformation and its modification

The Log-Sigmoid transformation (LST) $\psi : \mathbb{R} \to (-\infty, 2\ln 2)$ we define by formula

$$\psi(t) = 2\ln 2 S(t, 1) = 2\ln 2(1 + e^{-t})^{-1} = 2(\ln 2 + t - \ln(1 + e^t)). \qquad (6)$$

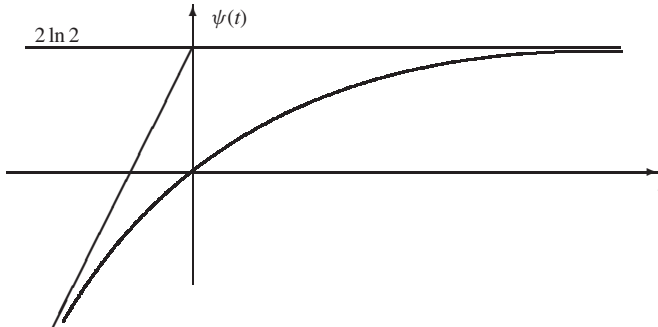The following proposition states the basic LST properties (see Fig. 1)

Fig. 1.

Proposition 1. *The LST ψ has the following properties:*

A1. $\psi(0) = 0$,
A2. $\psi'(t) > 0, \forall t \in (-\infty, +\infty)$ *and* $\psi'(0) = 1$,
A3. $\psi''(t) < 0, \forall t \in (-\infty, +\infty)$,
A4. $\lim_{t \to \infty} \psi'(t) = 0$.
A5. *a)* $0 < \psi'(t) < 2$; *b)* $-0.5 \leq \psi''(t) < 0, -\infty < t < +\infty$.

The substantial difference between $\psi(t)$ and the barrier or shifted barrier transformation (see [25]) is that $\psi(t)$ is defined on $(-\infty, +\infty)$ together with its derivatives of any order.

Property A5 distinguishes $\psi(t)$ from exponential see e.g. [6], [17], [37], barrier [21], [25], [27] and other transformations, which lead to classes of Nonquadratic Augmented Lagrangians P_I and $\hat{P}_I$ (see [6], p. 309) as well as transformations considered lately (see [3], [14], [27], [34], [35] and bibliography therein).

Also the LST possesses an extra important property.

To describe it we consider Fenchel conjugate $\psi^*(s)$. Due to A3 for any $s \in (0, 2)$ the equation $(st - \psi(t))'_t = 0$ can be solved for t, i.e. the inverse function ψ'^{-1} exists and $\psi'^{-1}(s) = t(s) = \ln(2s^{-1} - 1)$. The LS conjugate

$$\psi^*(s) = st(s) - \psi(t(s)) = (s - 2)\ln(2 - s) - s\ln s$$

is the Fermi-Dirac type Entropy type function [19].

Assuming that $t \ln t = 0$ for $t = 0$ we obtain that $-2 \ln 2 \le \psi^*(s) \le 0$ for $0 \le s \le 2$, and since $\psi^*(1) = 0$, we have $\max\{\psi^*(s) \,|\, 0 \le s \le 2\} = \psi^*(1) = 0$. Also we have (see [33], Sect. 24)

$$(0, 2) = ridom\, \psi^* \subset range\, \psi' \subset dom\, \psi^* = [0, 2].$$

Keeping in mind $t(s) = \psi'^{-1}(s) = \psi^{*'}(s)$ we obtain the following identity

$$s = \psi'(t) \equiv \psi'(\psi^{*'}(s)). \tag{7}$$

By differentiating the identity (7) we obtain $1 = \psi''(\psi^{*'}(s))\psi^{*''}(s)$ and again using $t = \psi^{*'}(s)$ we have

$$\psi^{*''}(s) = \frac{1}{\psi''(\psi^{*'}(s))} = \frac{1}{\psi''(t)}.$$

From A5, b) we obtain $\psi^{*''}(s) \le \psi^{*''}(1) = -2$, $\forall s \in (0, 2)$, or
   A6. $-\psi^{*''}(s) \ge 2$, $\forall s \in (0, 2)$.
   Let $\varphi : [0, 2] \to [0, 2\ln 2]$ be defined by formula

$$\varphi(s) = -\psi^*(s) = (2 - s)\ln(2 - s) + s\ln s. \tag{8}$$

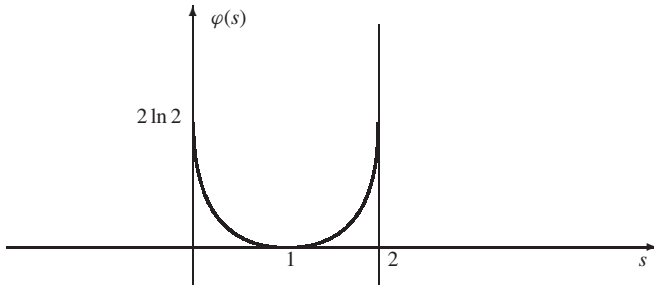We will call φ the Fermi-Dirac (FD) kernel (see Fig. 2).

Fig. 2.

The correspondent φ–divergence distance function, which will be introduced later, we will call Fermi-Dirac distance.

Proposition 2. *The FD kernel has the following properties.*
B1. *$\varphi(s)$ is nonnegative and strongly convex on (0,2);*
B2. *$\varphi(1) = \varphi'(1) = 0$;*
B3. *$\varphi'(s) = \ln s - \ln(2 - s)$, $\lim_{s \to 0^+} \varphi'(s) = -\infty$, $\lim_{s \to 2^-} \varphi'(s) = +\infty$;*
B4. *$\varphi''(s) = s^{-1} + (2 - s)^{-1} \ge 2$, $\forall s \in (0, 2)$;*

The properties of the FD kernel can be verified directly.

We conclude the section by introducing a modification of LST ψ. There are two main reasons for it. First of all from A2 we have

$$1 < \psi'(t) < 2, \ \forall t \in (-\infty, 0) \tag{9}$$

The critical element in the NR multipliers method, which we describe in Sect. 6 is the formulas for the Lagrange multipliers update $\lambda_i^{s+1} = \lambda_i^s \psi'\left(k_i^s c_i(x^{s+1})\right), i = 1, \ldots, m$. From (9) we have $\lambda_i^s < \lambda_i^{s+1} < 2\lambda_i^s$. It means that none of the Lagrange multipliers can be increased more that twice independent on the constraint violation and the value of the scaling parameter.

On the other hand due to A2 for any $i : c_i(x^{s+1}) > 0$ the Lagrange multiplier λ_i^s can be reduced practically to zero if $c_i(x^{s+1})$ or k_i^s or both are large enough. Such asymmetry can potentially compromise the convergence.

The second reason is also related to (9). We have $s = \psi'(t) \to 2^- \Rightarrow \psi''(t) \to 0$ and $\psi''(t) < 0$, so $s \to 2^- \Rightarrow \psi^{*''}(s) = \left(\psi''(t)\right)^{-1} \to -\infty$ or

$$s \to 2^- \Rightarrow \varphi''(s) \to \infty. \tag{10}$$

The property (10) can potentially compromise the convergence of the Dual Interior Prox method, which is equivalent to the NR multipliers method.

Therefore along with LST (6) we consider the following modification $\bar{\psi} : \mathbb{R} \to (-\infty, 2\ln 2)$, see Fig. 3, which is defined by formula

$$\bar{\psi}(t) = \begin{cases} \psi(t) = 2\ln 2 S(t, 1), \ t \geq -\ln 2 \\ q(t) = at^2 + bt + c, \ t \leq -\ln 2, \end{cases} \tag{11}$$
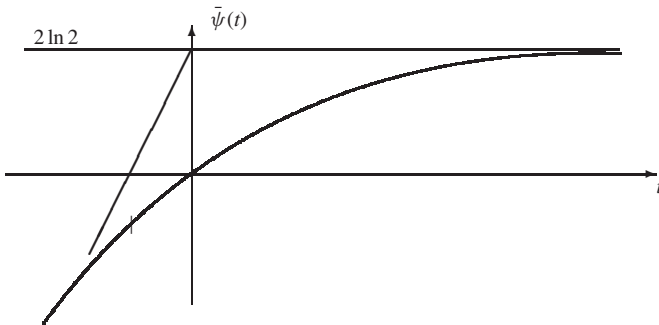

Fig. 3.

We find the parameters of the parabolic branch from $\psi(-\ln 2) = q(-\ln 2)$, $\psi'(-\ln 2) = q'(-\ln 2)$ and $\psi''(-\ln 2) = q''(-\ln 2)$, to ensure that $\bar{\psi} \in C^2$. By direct calculation we obtain $a = -2/9, b = \frac{4}{3}(1 - \frac{1}{3}\ln 2), c = \frac{10}{3}\ln 2 - \frac{2}{9}\ln^2 2 - 2\ln 3$.

So the properties A1–A4 and A5 b) remain true for the modification $\bar{\psi}$. At the same time

$$\lim_{t \to -\infty} \bar{\psi}'(t) = \infty \tag{12}$$

The Fenchel conjugate $\bar{\psi}^* : (0, \infty) \to (-\infty, 0)$ is defined as follows

$$\bar{\psi}^*(s) = \begin{cases} \psi^*(s), & 0 < s \le 4/3 = \psi'(-\ln 2) \\ q^*(s) = (4a)^{-1}(s-b)^2 - c, & 4/3 \le s < \infty. \end{cases}$$

Then instead of A6 we have

$$-\bar{\psi}^{*''}(s) \ge 2, \ \forall s \in (0, \infty). \tag{13}$$

We define the modified Fermi-Dirac kernel $\bar{\varphi} : (0, \infty) \to (0, \infty)$ by formula

$$\bar{\varphi}(s) = \begin{cases} -\psi^*(s), & 0 < s \le 4/3 \\ -q^*(s), & 4/3 \le s < \infty \end{cases}$$

and assume that $\bar{\varphi}(s) = +\infty$, for $s \in (-\infty, 0)$.

The modified FD kernel $\bar{\varphi}$ has the following properties, which correspond to B1–B4 (see Fig. 4):

$\bar{B}1.$ $\bar{\varphi}(s) \ge 0$, $s \in (0, \infty)$ is nonnegative and strongly convex on $(0, \infty)$;
$\bar{B}2.$ $\bar{\varphi}(1) = \bar{\varphi}'(1) = 0$;
$\bar{B}3.$ $\lim_{s \to 0^+} \bar{\varphi}'(s) = -\infty$, $\lim_{s \to \infty} \varphi'(s) = +\infty$;
$\bar{B}4.$ $\bar{\varphi}''(s) \ge 2$, $\forall s \in (0, \infty)$.
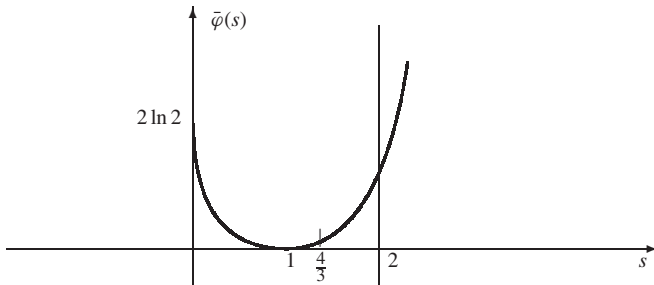
Fig. 4.

Also for $\bar{\varphi}(s)$ we have

$$\bar{\varphi}''(s) = \begin{cases} \varphi''(s), & 0 < s \le 4/3 \\ -(2a)^{-1} = 2.25, & 4/3 \le s < \infty. \end{cases}$$

So along with $\bar{B}1$–$\bar{B}4$ the modified FD kernel $\bar{\varphi}$ possesses the following extra property

$(\bar{B}5)$ $\qquad\qquad\qquad \bar{\varphi}''(s) = -\bar{\psi}^{*''} \le 2.25, \ \forall s \in [1, \infty).$

The transformation $\bar{\psi}$ relates to ψ as Penalty-Barrier [3] relates to the MBF transformation $\psi(t) = \ln(t+1)$. However the motivations for MBF and LS modifications are fundamentally different.

We shall see later that in LS multipliers methods the parabolic branch of $\bar{\psi}$ can be used only a finite number of steps. In fact for the scaling parameter $k > 0$ large enough it can happen only once. Hence from some point on only LS transformation governs the computational process in primal space and only FD kernel $\varphi = -\psi^*$ does it in the dual space.

Therefore we will retain ψ and φ as our notation in the rest of the paper keeping in mind that ψ possesses properties A1–A4 and A5(b) but instead of A5(a) has the property (12) and at the same time, φ along with the properties $\bar{B}1–\bar{B}4$ has the extra property $\bar{B}5$, i.e.

$$\psi(t) \equiv \begin{cases} \psi(t), \ t \geq -\ln 2 \\ q(t), \ t \leq -\ln 2, \end{cases}$$

$$\varphi(s) \equiv \begin{cases} -\psi^*(s), \ 0 < s \leq 4/3 \\ -q^*(s), \ 4/3 \leq s < \infty \end{cases}.$$

4. Equivalent problem and Log-Sigmoid Lagrangian

For any given vector $\mathbf{k} = (k_1, \dots, k_m) \in \mathbb{R}^m_{++}$ due to A1–A2 we have

$$c_i(x) \geq 0 \Leftrightarrow k_i^{-1}\psi(k_i c_i(x)) \geq 0, i = 1, \dots, m. \tag{14}$$

Therefore the problem

$$x^* \in X^* = Argmin\left\{f(x) \big| k_i^{-1}\psi(k_i c_i(x)) \geq 0, \ i = 1, \dots, m\right\} \tag{15}$$

is equivalent to the original problem $\mathcal{P}$.

The Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m_{++} \times \mathbb{R}^m_{++} \to \mathbb{R}$ for the equivalent problem (15) – Log-Sigmoid Lagrangian (LSL) is given by formula

$$\mathcal{L}(x, \lambda, \mathbf{k}) = f(x) - \sum_{i=1}^{m} \lambda_i k_i^{-1}\psi(k_i c_i(x)) \tag{16}$$

The following lemma establishes the basic LSL properties at any K-K-T's pair (x^*, λ^*).

Lemma 1. *For any* $\mathbf{k} \in \mathbb{R}^m_{++}$ *any K-K-T's pair* (x^*, λ^*) *the following properties hold true.*

1. $\mathcal{L}(x^*, \lambda^*, \mathbf{k}) = L(x^*, \lambda^*) + 2 \sum_{i=1}^{m} k_i^{-1}\lambda_i^* \ln \cosh(0.5 k_i c_i(x^*)) = L(x^*, \lambda^*) = f(x^*)$
2. $\nabla_x \mathcal{L}(x^*, \lambda^*, \mathbf{k}) = \nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0$
3. $\nabla^2_{xx}\mathcal{L}(x^*, \lambda^*, \mathbf{k}) = \nabla^2_{xx} L(x^*, \lambda^*) + 0.5\nabla c(x^*)^T K\Lambda^*\nabla c(x^*)$, *where* $K = diag(k_i)_{i=1}^m$, $\Lambda^* = diag(\lambda_i^*)_{i=1}^m$.

Proof. Property 1 follows from A1 and the complementary condition (2). The Property 2 follows from A2 and (2). Property 3 follows from A2 and A3.

Remark 1. In view of $\lambda_i^* = 0, i = r + 1, \dots, m$ Property 3 can be rewritten as follows

$$\nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*, \mathbf{k}) = \nabla_{xx}^2 L(x^*, \lambda^*) + 0.5 \nabla c_{(r)}(x^*)^T K_r \Lambda_r^* \nabla c_{(r)}(x^*),$$

where $K_r = diag(k_i)_{i=1}^r$, $\Lambda_r^* = diag(\lambda_i)_{i=1}^m$. Let $k > 0$ and $K_r = k\Lambda_r^{*-1}$, then

$$\nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*, \mathbf{k}) = \nabla_{xx}^2 L(x^*, \lambda^*) + 0.5 k \nabla c_{(r)}(x^*)^T \nabla c_{(r)}(x^*),$$

which is exactly the Hessian of the Quadratic Augmented Lagrangian [12], [29].

The similarity between LSL and Quadratic Augmented Lagrangian will play an important role in our future considerations.

If $k_i = \dots = k_m = k$, then

$$\mathcal{L}(x, \lambda, \mathbf{k}) \equiv \mathcal{L}(x, \lambda, k) = f(x) - k^{-1} \sum_{i=1}^m \lambda_i \psi(k c_i(x)). \tag{17}$$

Lemma 2. *If $f(x)$ and all $c_i(x) \in C^2$ then for any given pair $(\lambda, \mathbf{k}) \in \mathbb{R}_{++}^m \times \mathbb{R}_{++}^m$ the LSL Hessian is positive definite for any $x \in \mathbb{R}^n$, i.e. $\mathcal{L}(x, \lambda, \mathbf{k})$ is strictly convex in $\mathbb{R}^n$ and strongly convex on any bounded set in $\mathbb{R}^n$.*

Proof. The proof follows directly from the formula for the LSL Hessian.

$$\nabla_{xx}^2 \mathcal{L}(x, \lambda, \mathbf{k}) = \nabla_{xx}^2 L(x, \hat{\lambda}) - \nabla c_{(p)}^T(x) K_p \Lambda_p \Psi''\big(k_{(p)} c_{(p)}(x)\big) \nabla c_{(p)}(x)$$
$$- K_n \Lambda_n \Psi''\big(k_{(n)} c_{(n)}(x)\big),$$

where $\hat{\lambda} = \Psi'(k_{(p)} c_{(p)}(x))\lambda$, $\Psi'(k_{(p)} c_{(p)}(x)) = diag(\psi'(k_i c_i(x)))_{i=1}^p$, $\Psi''(k_{(p)} c_{(p)}(x)) = diag(\psi''(k_i c_i(x)))_{i=1}^p$, $\Psi''(k_{(n)} c_{(n)}(x)) = diag(\psi''(k_i x_i))_{i=1}^n$, $K_p = diag(k_i)_{i=1}^p$, $K_n = diag(k_{p+j})_{j=1}^n$, $\Lambda_p = diag(\lambda_i)_{i=1}^p$, $\Lambda_n = diag(\lambda_i)_{i=1}^n$, and the convexity of $f(x)$ and all $-c_i(x)$.

Lemma 3. *If X^* is bounded then for any pair $(\lambda, \mathbf{k}) \in \mathbb{R}_{++}^m \times \mathbb{R}_{++}^m$ there exists*

$$\hat{x} = \hat{x}(\lambda, \mathbf{k}) = argmin\{\mathcal{L}(x, \lambda, \mathbf{k}) \mid x \in \mathbb{R}^n\}$$

Proof. From boundness of X^* follows (see [1])

$$\begin{matrix} f_\infty(d) \le 0 \\ (-c_i)_\infty(d) \le 0, \ i = 1, \dots, m \end{matrix} \Rightarrow d = 0, \tag{18}$$

where f_∞ and $(-c_i)_\infty$ are recession functions of f and $-c_i$. Using considerations similar to those in [1] or [6] and keeping in mind the convexity of f, concavity of c_i the property (12) and (18) we obtain that for any fixed $\lambda \in \mathbb{R}_{++}^m$ and $\mathbf{k} \in \mathbb{R}_{++}^m$. the LSL $\mathcal{L}(x, \lambda, \mathbf{k})$ has no direction of recession in x, i.e. for any nontrivial direction $z \in \mathbb{R}^n$ we have

$$\lim_{t \to \infty} \mathcal{L}(x + tz, \lambda, \mathbf{k}) = \infty.$$

Therefore the set

$$\hat{X}(\lambda, \mathbf{k}) = \{\hat{x} \mid \mathcal{L}(\hat{x}, \lambda, \mathbf{k}) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mathbf{k})\}$$

is not empty and bounded ([33], Theorem 27.1d).

Moreover for $(\lambda, \mathbf{k}) \in \mathbb{R}^m_{++} \times \mathbb{R}^m_{++}$ due to Lemma 2 the set $\hat{X}(\lambda, \mathbf{k})$ contains only one point $\hat{x} = \hat{x}(\lambda, \mathbf{k}) = argmin\{\mathcal{L}(x, \lambda, \mathbf{k}) \mid x \in \mathbb{R}^n\}$. The uniqueness of the minimizer $\hat{x}(\lambda, \mathbf{k})$ means that in contrast to the dual function $d(\lambda) = \inf\{L(x, \lambda) \mid x \in \mathbb{R}^n\}$, which is based on the Lagrangian for the initial problem (P), the dual function $d_{\mathbf{k}}(\lambda) = \min\{\mathcal{L}(x, \lambda, \mathbf{k}) \mid x \in \mathbb{R}^n\}$ is smooth if f and c_i are smooth.

Besides, keeping in mind that $x(\cdot) = x(\lambda, \mathbf{k}) = argmin\{\mathcal{L}(x, \lambda, \mathbf{k}) \mid x \in \mathbb{R}^n\}$, i.e. $\nabla_x \mathcal{L}(x(\cdot), \cdot) = 0$ we obtain for the gradient of the dual function the following formula:

$$\nabla d_{\mathbf{k}}(\lambda) = \nabla_x \mathcal{L}(x(\cdot), \cdot) \nabla_\lambda x(\cdot) + \nabla_\lambda \mathcal{L}(x(\cdot), \cdot) = \nabla_\lambda \mathcal{L}(x(\cdot), \cdot)$$
$$= -\left(k_1^{-1} \psi(k_1 c_1(x(\cdot))), \ldots, k_m^{-1} \psi(k_m c_m(x(\cdot)))\right).$$

5. Log-Sigmoid multipliers methods – Interior Prox methods with Fermi-Dirac Entropy distance

In this section we consider the multipliers method, where along with the Lagrange multipliers the scaling vector $\mathbf{k} \in \mathbb{R}^m_{++}$ will be updated from step to step by formulas, which have been suggested by P. Tseng and D. Bertsekas in [37] for the exponential multipliers method.

Let $(\lambda, \mathbf{k}) \in \mathbb{R}^m_{++} \times \mathbb{R}^m_{++}, \lambda^0 = e = (1, \ldots, 1), \mathbf{k}^0 = (k_1^0, \ldots, k_m^0), k_i^0 = k/\lambda_i^0 = k$, $i = 1, \ldots, m$ and $k > 0$. The LS multipliers method generates three sequences $\{x^s\}$, $\{\lambda^s\}$ and $\{\mathbf{k}^s\}$ by formulas

$$x^{s+1} = argmin\{\mathcal{L}(x, \lambda^s, \mathbf{k}^s) \mid x \in \mathbb{R}^n\} \tag{19}$$

$$\lambda_i^{s+1} = \lambda_i^s \psi'\left(k_i^s c_i(x^{s+1})\right), \ i = 1, \ldots, m \tag{20}$$

$$k_i^{s+1} = k\left(\lambda_i^{s+1}\right)^{-1}, \ i = 1, \ldots, m. \tag{21}$$

From (20), $\psi'(0) = 1$ and the mean value formula we obtain

$$\lambda_i^{s+1} - \lambda_i^s = \lambda_i^s\left(\psi'\left(k_i^s c_i(x^{s+1})\right) - \psi'(0)\right) = \lambda_i^s k_i^s \psi''\left(\theta_i^s k_i^s c_i(x^{s+1})\right) c_i(x^{s+1}),$$

where $0 < \theta_i^s < 1$. Using (21) we can rewrite the multiplicative formula (20) for the Lagrange multipliers update in an additive form

$$\lambda_i^{s+1} = \lambda_i^s + k\psi''(\cdot)_{i,s} c_i(x^{s+1}), \ i = 1, \ldots, m.$$

So the method (19)–(21) one can view as an implicit rescaled subgradient method for dual problem, which guarantees $\lambda^s \in \mathbb{R}^m_{++}$ independent on $k > 0$.

It is equivalent to the following quadratic prox method for the dual problem in the rescaled dual space.

$$\lambda^{s+1} = argmax\Big\{d(\lambda) - 0.5k^{-1} \sum_{i=1}^m (-\psi''(\cdot)_{i,s})^{-1}\left(\lambda_i - \lambda_i^s\right)^2 \mid \lambda \in \mathbb{R}^m\Big\}.$$

Now we will show that (19)–(21) is equivalent to an Interior Prox method with second order φ–divergence distance function, which are based on FD kernel φ.

The dual problem associated with $\mathcal{P}$ consists in finding

$$(\mathcal{D}) \qquad\qquad \lambda^* \in L^* = Argmax\{d(\lambda) \,|\, \lambda \in \mathbb{R}^m_+\}.$$

The vector x^{s+1} is an unconstrained minimizer of LSL in x, i.e.

$$\nabla_x \mathcal{L}(x^{s+1}, \lambda^s, \mathbf{k}^s) = \nabla f(x^{s+1}) - \sum_{i=1}^m \lambda_i^s \psi'\big(k_i^s c_i(x^{s+1})\big) \nabla c_i(x^{s+1}) = 0. \qquad (22)$$

Using the formula (20) we can rewrite (22) as follows:

$$\nabla_x \mathcal{L}(x^{s+1}, \lambda^s, \mathbf{k}^s) = \nabla f(x^{s+1}) - \sum_{i=1}^m \lambda_i^{s+1} \nabla c_i(x^{s+1}) = \nabla_x L(x^{s+1}, \lambda^{s+1}) = 0. \quad (23)$$

It means that $d(\lambda^{s+1}) = L(x^{s+1}, \lambda^{s+1}) = \min_{x \in \mathbb{R}^n} L(x, \lambda^{s+1})$ and $\lambda^{s+1} \in \mathbb{R}^m_{++}$.

In other words λ^{s+1} is an interior point for the dual problem $(\mathcal{D})$, while the primal approximation x^{s+1} might be infeasible.

From the definition of the dual function we have

$$\big(-c_1(x^{s+1}), \dots, -c_m(x^{s+1})\big)^T \in \partial d(\lambda^{s+1}), \qquad (24)$$

where $\partial d(\lambda^{s+1})$ denotes the subdifferential of $d(\lambda)$ at λ^{s+1}. We can rewrite (24) in the following way

$$0 \in \partial d(\lambda^{s+1}) + \big(c_1(x^{s+1}), \dots, c_m(x^{s+1})\big)^T. \qquad (25)$$

Now let's consider the formula for the Lagrange multipliers update

$$\lambda_i^{s+1} = \lambda_i^s \psi'\big(k_i^s c_i(x^{s+1})\big), \ i = 1, \dots, m,$$

we obtain

$$c_i(x^{s+1}) = \big(k_i^s\big)^{-1} \psi'^{-1}\big(\lambda_i^{s+1}/\lambda_i^s\big), \ i = 1, \dots, m \qquad (26)$$

The existence of ψ'^{-1} is guaranteed by A3.

The inclusion (25) one can rewrite as follows:

$$0 \in \partial d(\lambda^{s+1}) + \left(\big(k_1^s\big)^{-1} \psi'^{-1}\big(\lambda_1^{s+1}/\lambda_1^s\big), \dots, \big(k_m^s\big)^{-1} \psi'^{-1}\big(\lambda_m^{s+1}/\lambda_m^s\big)\right)^T. \qquad (27)$$

Using again the formula $\psi'^{-1} = \psi^{*'}$ we can rewrite (26) as

$$c_i(x^{s+1}) = \big(k_i^s\big)^{-1} \psi^{*'}\big(\lambda_i^{s+1}/\lambda_i^s\big), \ i = 1, \dots, m \qquad (28)$$

and (27) as

$$0 \in \partial d(\lambda^{s+1}) + \left(\big(k_1^s\big)^{-1} \psi^{*'}\big(\lambda_1^{s+1}/\lambda_1^s\big), \dots, \big(k_m^s\big)^{-1} \psi^{*'}\big(\lambda_m^{s+1}/\lambda_m^s\big)\right)^T. \qquad (29)$$

Using $\varphi(t) = -\psi^*(t)$ we can rewrite (29) as follows

$$0 \in \partial d(\lambda^{s+1}) - \left(\left(k_1^s\right)^{-1} \varphi'\left(\lambda_1^{s+1}/\lambda_1^s\right), \dots, \left(k_m^s\right)^{-1} \varphi'\left(\lambda_m^{s+1}/\lambda_m^s\right) \right)^T. \qquad (30)$$

From B1, concavity of $d(\lambda)$ and (21) it follows that the inclusion (30) is the optimality criteria for $\lambda^{s+1} = (\lambda_1^{s+1}, \dots, \lambda_m^{s+1})$ to be a positive unconstrained maximizer in λ of the following function

$$H(\lambda, \lambda^s, \mathbf{k}^s) = d(\lambda) - \sum_{i=1}^{m} \left(k_i^s\right)^{-1} \lambda_i^s \varphi\left(\lambda_i/\lambda_i^s\right) = d(\lambda) - k^{-1} \sum_{i=1}^{m} \left(\lambda_i^s\right)^2 \varphi\left(\lambda_i/\lambda_i^s\right). \qquad (31)$$

The function $H(\lambda, \lambda^s, \mathbf{k}^s)$ is strongly concave in λ, so the unconstrained maximizer λ^{s+1} is unique.

The function $D_2 : \mathbb{R}_+^m \times \mathbb{R}_{++}^m \times \mathbb{R}_{++}^m \to \mathbb{R}_+$, which is defined by formula

$$D_2(u, v) = \sum_{i=1}^{m} v_i^2 \varphi(u_i/v_i) \qquad (32)$$

is the second order φ–divergence distance function, which is based on FD kernel $\varphi(t)$.

We will call $D_2(u, v)$ the second order Fermi-Dirac (FD) Entropy distance function. Due to $\varphi(t) = +\infty$, $t \in (-\infty, 0)$ and keeping in mind (30) we obtain

$$\lambda^{s+1} = argmax\left\{ d(\lambda) - k^{-1} \sum_{i=1}^{m} \left(\lambda_i^s\right)^2 \varphi\left(\lambda_i/\lambda_i^s\right) \,\big|\, \lambda \in \mathbb{R}^m \right\}$$

$$= argmax\left\{ d(\lambda) - k^{-1} D_2(\lambda, \lambda^s) \,\big|\, \lambda \in \mathbb{R}^m \right\} \qquad (33)$$

We proved the following theorem.

Theorem 1. *If $\mathcal{P}$ is a convex programming problem and the assumption A and B are satisfied, then for any $k > 0$ and $(\lambda^0, \mathbf{k}^0) \in \mathbb{R}_{++}^m \times \mathbb{R}_{++}^m$:*

1) *the LS method (19)–(21) is executable;*
2) *the LS method (19)–(21) is equivalent to the Interior Prox method (33) for the dual problem with the second order FD distance function.*
3) *both the primal $\{x^s\}_{s=0}^{\infty}$ and the dual $\{\lambda^s\}_{s=0}^{\infty}$ sequences are unique.*

If $k_1^s = \cdots = k_m^s = k$ then the method (19)–(21) turns into the following multipliers method

$$x^{s+1} = argmin\{\mathcal{L}(x, \lambda^s, k) \,|\, x \in \mathbb{R}^n\} \qquad (34)$$

$$\lambda_i^{s+1} = \lambda_i^s \psi'\left(kc_i(x^{s+1})\right), \ i = 1, \dots, m. \qquad (35)$$

By repeating the arguments of Theorem 1 it is easy to see that the multipliers method (34)–(35) is equivalent to the following Interior Prox method

$$\lambda^{s+1} = argmax\left\{ d(\lambda) - k^{-1} \sum_{i=1}^{m} \lambda_i^s \varphi\left(\lambda_i/\lambda_i^s\right) \,\big|\, \lambda \in \mathbb{R}^m \right\}$$

$$= argmax\left\{ d(\lambda) - k^{-1} D_1(\lambda, \lambda^s) \,\big|\, \lambda \in \mathbb{R}^m \right\}, \qquad (36)$$

where $D_1(u, v) = \sum_{i=1}^{m} v_i \varphi(u_i/v_i)$ is the first order FD distance.

Note that the first and the second order FD distance functions have all the characteristics of a distance function:

1. For any pair (u, v) such that $u \in \mathbb{R}_+^m$ and $v \in \mathbb{R}_{++}^m$ due to Property B1 from Assertion 3.2 we have

$$D_i(u, v) \geq 0$$

    and due to Properties B2 and B4 we have

2. $$D_i(u, v) = 0 \Leftrightarrow u = v, \; i = 1, 2$$

About φ–divergence functions and their applications see [34] and references therein.

A number of φ–divergence distance functions associated with exponential, barrier transformations as well as transformations that are based on quadratic extrapolation technique have been considered lately, see e.g. [2], [3], [14], [27], [35], [37], and references therein.

Unfortunately the FD kernel does not possess the properties, which were critical for the convergence proofs of the multipliers method (34)–(35), which leads to the Interior Prox Method with the first order φ divergence distance (see [14], [27], [35]).

On the other hand transformation $\bar{\psi}$ has an extra important property (13), which leads to strong convexity property $\bar{B}4$ for the modified FD kernel $\bar{\varphi}$ on the top to the standard properties of a dual kernel (see [3]).

The lack of strong convexity of the dual kernels, is one of the reasons for regularization of the dual kernels (see formula (2.12) in [2]). We will discuss the consequences of such regularization later. Now we would like to emphasize that the properties $\bar{B}4$ and $\bar{B}5$ make the second order FD distance $D_2(u, v)$ close to the Quadratic distance in the truncated dual space, which corresponds to the primal active constraints.

Also it makes the Interior Prox Method (33) close to the Martinet's Quadratic Prox method [20] and the mapping $v \to J_{\mathbf{k}}(v) = \max\{H(u, v, \mathbf{k}) \mid u \in \mathbb{R}^m\}$ close to Moreau proximal mapping [22] in the truncated dual space.

On the other hand, as we will see later, the parabolic branch of $\bar{\psi}$ can be used only a finite number of steps. For $k > 0$ large enough it can happen just once, therefore the multipliers method (19)–(21) has all the advantages of Nonquadratic Augmented Lagrangians for inequality constraints, namely the LSL remains as smooth as the initial functions and LS multipliers methods keep the Lagrange multipliers positive without any particular care.

So we are dealing with a new class of Nonquadratic Augmented Lagrangians and corresponding Prox methods, which combine the best properties of the Quadratic Prox and Interior Prox and at the same time are free from their drawbacks.

6. Convergence of the Log-Sigmoid multipliers method

The convergence analysis of exponential, barrier and other multipliers methods associated with Nonquadratic Augmented Lagrangians has proved surprisingly difficult. In particular, the exponential multipliers method was introduced as early as 1973, see [17]. Only in 1993 it was proven the ergodic convergence [37]. For the exponential multipliers

method with "dynamic" scaling parameters update by (21) even ergodic convergence was problematic (see [37], p. 3).

Partially the difficulties were overcome in [3] within more general interaction scheme between the scaling parameters and Lagrange multipliers. The main results in [3] are the boundness of the primal-dual sequence and that any limit point of the primal-dual sequence is the primal-dual solution. These results were extended and improved in [2], establishing that inexact proximal version generates bounded sequence with every limit point being an optimal solution.

Moreover in [2] the convergence results were strengthened for a particular second order dual Prox method type (33) with a regularized MBF kernel. By adding the quadratic term to the dual kernel, which corresponds to logarithmic MBF, the authors in [2] proved the global convergence of the dual sequence and established under assumptions A and B that the rate of convergence is $\mathcal{O}(ks)^{-1}$. Unfortunately such modification of the dual kernel leads to substantial difficulties, when it comes to finding the primal transformation, which is the Fenchel conjugate for the dual kernel. For example, in case of exponential transformation it leads to solving a transcendent equation. In case of the kernel, which corresponds to the Logarithmic MBF transformation the closed form solution of the corresponding equation is available but the primal transformation (see Sect. 7 in [2]) is substantially different from the original MBF transformation, which is proven to be very efficient numerically (see [4], [5], [23]).

So it seems that considering alternative ways for proving convergence and especially to estimate the rate of convergence of the NR method (19)–(21) for the original transformations under minimum assumption on the input data still remains an important issue.

In this section we present a new convergence proof and estimate the convergence rate of the LS multipliers method (19)–(21) and its dual equivalent (33) under very mild assumptions A and B on the input data.

The basic ingredients of the convergence proof are the equivalence of the LS multipliers method (19)–(21) to the Prox-method (33) with second order FD distance function and the properties $\bar{B}4$ and $\bar{B}5$.

Let $a = d(\lambda^*) - d(\lambda^0)$, $\Lambda_0 = \left\{ \lambda \in \mathbb{R}^m_+ : d(\lambda) \geq d(\lambda^0) \right\}$ is bounded due to concavity of $d(\lambda)$ and boundness of L^* (see Corollary 20 in [10]), $L_0 = \max \left\{ \max_{1 \leq i \leq m} \lambda_i : \lambda \in \Lambda_0 \right\}$, $I^-_{l+1} = \{i : c_i(x^{l+1}) < 0\}$. We consider the maximum constraints violation

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

and the estimation for the duality gap

$$d_l = \sum_{i=1}^m \lambda^l_i |c_i(x^l)| \geq \sum_{i=1}^m \lambda^l_i c_i(x^l).$$

Let $\bar{v}_s = \min_{1 \leq l \leq s} v_l$ and $\bar{d}_s = \min_{1 \leq l \leq s} d_l$.

Theorem 2. *If the assumption A and B are satisfied then*

1) *the dual sequence $\{\lambda^s\}$ is bounded and the following estimation*

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

holds true,

2) *the primal sequence $\{x^s\}$ is bounded and asymptotically feasible and the following estimation holds*

$$d(\lambda^{s+1}) - d(\lambda^s) \geq \frac{32}{81}k \sum_{i \in I_{s+1}^-} c_i^2(x^{s+1});$$

3) $\bar{v}_s \leq \frac{9\sqrt{a}}{4\sqrt{2}}(sk)^{-0.5}, \bar{d}_s \leq 2.25 L_0\sqrt{am/2}(ks)^{-0.5}$

4) *every limit point of the primal-dual sequence $\{x^s, \lambda^s\}$ is the primal-dual solution and*

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

Also $\lim_{s \to \infty} \rho(x^s, X^*) = 0, \lim_{s \to \infty} \rho(\lambda^s, L^*) = 0$, where $\rho(x, Y) = \min_{y \in Y} \|x - y\|$.

Proof. 1) We start by establishing the dual monotonicity. It follows immediately from (33) and $D_2(\lambda^s, \lambda^s) = 0$, i.e.

$$d(\lambda^{s+1}) - k^{-1}D_2(\lambda^{s+1}, \lambda^s) \geq d(\lambda^s) - k^{-1}D_2(\lambda^s, \lambda^s) = d(\lambda^s). \quad (37)$$

Taking into account $D(\lambda^{s+1}, \lambda^s) > 0$ for $\lambda^{s+1} \neq \lambda^s$, we obtain

$$d(\lambda^{s+1}) \geq d(\lambda^s) + k^{-1}D_2(\lambda^{s+1}, \lambda^s) > d(\lambda^s). \quad (38)$$

If $d(\lambda^{s+1}) = d(\lambda^s) \Rightarrow D_2(\lambda^{s+1}, \lambda^s) = 0 \Rightarrow \lambda^{s+1} = \lambda^s$. Due to the formula (20) for the Lagrange multipliers update the equality $\lambda_i^{s+1} = \lambda_i^s$ leads to $c_i(x^{s+1}) = 0$, $i = 1, \ldots, m$. So if $\lambda^{s+1} = \lambda^s \in \mathbb{R}_{++}^m$ then $c(x^{s+1}) = 0$ and for the pair (x^{s+1}, λ^{s+1}) the K-K-T's conditions hold, therefore $x^{s+1} = x^* \in X^*, \lambda^{s+1} = \lambda^* \in L^*$.

In other words we can either have $d(\lambda^{s+1}) > d(\lambda^s)$ or $\lambda^{s+1} = \lambda^s = \lambda^*$ and $x^{s+1} = x^*$.

Let's find the lower bound for $d(\lambda^{s+1}) - d(\lambda^s)$. From the concavity of $d(\lambda)$ and $-c(x^{s+1}) \in \partial d(\lambda^{s+1})$ we obtain

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

or

$$d(\lambda^{s+1}) - d(\lambda) \geq (c(x^{s+1}), \lambda - \lambda^{s+1}). \quad (39)$$

Keeping in mind A3 we conclude that ψ'^{-1} exists. Using the formula for the Lagrange multipliers update we have

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

then using $\psi'^{-1} = \psi^{*'}$, we obtain

$$c_i(x^{s+1}) = (k_i^s)^{-1}\psi^{*'}(\lambda_i^{s+1}/\lambda_i^s), \; i = 1, \ldots, m.$$

Using $\psi^{*'}(1) = \psi^{*'}(\lambda_i^s/\lambda_i^s) = 0$ and (39) we obtain

$$d(\lambda^{s+1}) - d(\lambda) \geq \sum_{i=1}^{m} (k_i^s)^{-1}\left(\psi^{*'}\left(\frac{\lambda_i^{s+1}}{\lambda_i^s}\right) - \psi^{*'}\left(\frac{\lambda_i^s}{\lambda_i^s}\right)\right)(\lambda_i - \lambda_i^{s+1}). \tag{40}$$

Using the mean value formula

$$\psi^{*'}\left(\frac{\lambda_i^{s+1}}{\lambda_i^s}\right) - \psi^{*'}\left(\frac{\lambda_i^s}{\lambda_i^s}\right) = -\psi^{*''}(\cdot)(\lambda_i^s)^{-1}(\lambda_i^s - \lambda_i^{s+1}),$$

the update formula (21), inequality (40) and (13) we obtain for $\lambda = \lambda^s$ the following inequality

$$d(\lambda^{s+1}) - d(\lambda^s) \geq 2k^{-1}\|\lambda^s - \lambda^{s+1}\|^2, \tag{41}$$

which is typical for the Quadratic Prox method (see [11], [20], [24]).

2) We remind that the set $\Lambda_0 = \{\lambda : d(\lambda) \geq d(\lambda^0)\}$ is bounded and so is the sequence $\{\lambda^s\} \subset \Lambda_0$.

Let $I_{s+1}^- = \{i : c_i(x^{s+1}) < 0\}$ be the set of indices of the constraints, which were violated at the step $s + 1$. Using $\psi^{*'}(1) = \psi^{*'}(\lambda_i^s/\lambda_i^s) = 0$, the mean value formula and (28) we obtain

$$-c_i(x^{s+1}) = (k_i^s)^{-1}\left[\psi^{*'}\left(\frac{\lambda_i^s}{\lambda_i^s}\right) - \psi^{*'}\left(\frac{\lambda_i^{s+1}}{\lambda_i^s}\right)\right]$$
$$= k^{-1}(-\psi^{*''}(\cdot))(\lambda_i^{s+1} - \lambda_i^s) \leq k^{-1}\varphi''(\cdot)|\lambda_i^{s+1} - \lambda_i^s|.$$

Using $\bar{B}5$ we have

$$\left|\lambda_i^{s+1} - \lambda_i^s\right| \geq (2.25)^{-1}k(-c_i(x^{s+1})), \; i \in I_{s+1}^-.$$

Combining the last estimation with (41) we obtain

$$d(\lambda^{s+1}) - d(\lambda^s) \geq \frac{32}{81}k \sum_{i \in I_{s+1}^-} c_i^2(x^{s+1}). \tag{42}$$

3) Let's consider $v_{l+1} = \max_{i \in I_{l+1}^-}(-c_i(x^{l+1}))$ — the maximum constraints violation at the step $l + 1$, then from (42) we have

$$d(\lambda^{l+1}) - d(\lambda^l) \geq \frac{32}{81}kv_{l+1}^2. \tag{43}$$

Summing up (43) from $l = 1$ to $l = s$ we obtain

$$a = d(\lambda^*) - d(\lambda^0) \geq d(\lambda^{s+1}) - d(\lambda^0) \geq \frac{32}{81} k \sum_{l=0}^{s} v_{l+1}^2,$$

therefore $v_s \to 0$. Let $\bar{v}_s = \min\{v_l \mid 1 \leq l \leq s\}$, then

$$\bar{v}_s \leq \frac{9\sqrt{a}}{4\sqrt{2}}(ks)^{-0.5} = \mathcal{O}\big((ks)^{-0.5}\big). \tag{44}$$

The primal asymptotic feasibility follows from $v_s \to 0$.
Now we will prove that similar to (44) estimation is taking place for the duality gap.
Using (20) and (21) we have

$$\lambda_i^{s+1} - \lambda_i^s = \psi'\big(k_i^s c_i(x^{s+1})\big)\lambda_i^s - \psi'(0)\lambda_i^s = k_i^s \lambda_i^s \psi''(\cdot)c_i(x^{s+1})$$

$$= k\psi''(\cdot)c_i(x^{s+1}) = \frac{k\psi''(\cdot)\lambda_i^{s+1}c_i(x^{s+1})}{\lambda_i^{s+1}}.$$

Hence

$$\big|\lambda_i^{s+1} - \lambda_i^s\big| = \frac{k|\psi''(\cdot)|\,\big|\lambda_i^{s+1}c_i(x^{s+1})\big|}{\lambda_i^{s+1}}, \quad i = 1, \ldots, m. \tag{45}$$

We remind that $L_0 = \max\Big\{\max_{1 \leq i \leq m} \lambda_i \,:\, \lambda \in \Lambda_0\Big\}$, therefore $\lambda_i^s \leq L_0, i = 1, \ldots, m$;
$s \geq 1$. In view of $\psi'''(t) = -2e^t(1-e^t)(1+e^t)^{-3} < 0, \forall t \in [-\ln 2, 0], \psi''(0) = -0.5$
and $\psi''(t) = -\frac{4}{9}, \forall t \in (-\infty, -\ln 2)$ we have $|\psi''(t)| \geq \frac{4}{9}, \forall t < 0$. Therefore from
(45) we have

$$\big|\lambda_i^{l+1} - \lambda_i^l\big| \geq \frac{4k}{9L_0}\big|\lambda_i^{l+1}c_i(x^{l+1})\big|, \quad i \in I_{l+1}^- = \big\{i \,:\, c_i(x^{l+1}) < 0\big\}. \tag{46}$$

Now we consider the set $I_{l+1}^+ = \{i \,:\, c_i(x^{l+1}) \geq 0\}$. Using (20) from (45) we obtain

$$\big|\lambda_i^{l+1} - \lambda_i^l\big| = \frac{k\,\big|\psi''\big(\theta_i^l k_i^l c_i(x^{l+1})\big)\big|\,\lambda_i^{l+1}c_i(x^{l+1})}{\lambda_i^l \psi'\big(k_i^l c_i(x^{l+1})\big)}, \quad i \in I_{l+1}^+, \tag{47}$$

where $0 < \theta_i^l < 1$.
For $0 < \theta < 1$ and $t \geq 0$ we have

$$\frac{|\psi''(\theta t)|}{\psi'(t)} = \frac{e^{\theta t}(1+e^t)}{(1+e^{\theta t})^2} \geq \frac{e^{\theta t}(1+e^{\theta t})}{(1+e^{\theta t})^2} \geq \frac{e^{\theta t}}{1+e^{\theta t}} \geq (1+e^{-\theta t})^{-1} \geq \frac{1}{2}.$$

Hence

$$\big|\lambda_i^{l+1} - \lambda_i^l\big| \geq \frac{k}{2L_0}\lambda_i^{l+1}c_i(x^{l+1}), \quad i \in I_{l+1}^+. \tag{48}$$

Combining (46) and (48) we obtain the inequalities

$$\left|\lambda_i^{l+1} - \lambda_i^l\right| \geq \frac{4}{9}\frac{k}{L_0}\left|\lambda_i^{l+1}c_i(x^{l+1})\right|, \, i = 1, \ldots, m; l \geq 0. \tag{49}$$

Using (41) and (49) we obtain

$$d(\lambda^{l+1}) - d(\lambda^l) \geq 2k^{-1}\sum_{i=1}^{m}\left(\lambda_i^{l+1} - \lambda_i^l\right)^2 \geq \frac{32k}{81L_0^2}\sum_{i=1}^{m}\left(\lambda_i^{l+1}c_i(x^{l+1})\right)^2. \tag{50}$$

Summing up (50) from $l = 0$ to $l = s$ we obtain

$$\sum_{l=0}^{s}\sum_{i=1}^{m}\left(\lambda_i^{l+1}c_i(x^{l+1})\right)^2 \leq \frac{81L_0^2}{32k}(d(\lambda^s) - d(\lambda^0))$$

$$\leq \frac{81L_0^2}{32k}(d(\lambda^*) - d(\lambda^0)) \leq \frac{81L_0^2a}{32k}. \tag{51}$$

Using the well known inequality

$$\frac{1}{m}\left(\sum_{i=1}^{m}t_i\right)^2 \leq \sum_{i=1}^{m}t_i^2$$

from (51) we obtain

$$\frac{1}{m}\sum_{l=0}^{s}\left(\sum_{i=1}^{m}\lambda_i^{l+1}c_i(x^{l+1})\right)^2 \leq \sum_{l=0}^{s}\sum_{i=1}^{m}\left(\lambda_i^{l+1}c_i(x^{l+1})\right)^2 \leq \frac{81L_0^2a}{32k}$$

or

$$\sum_{l=0}^{s}\left(\sum_{i=1}^{m}\lambda_i^{l+1}c_i(x^{l+1})\right)^2 \leq \frac{81mL_0^2a}{32k}.$$

So for the "best" in s steps duality gap

$$\bar{d}_s^2 = \min_{1\leq l\leq s}\left(\sum_{i=1}^{m}\lambda_i^{l+1}c_i(x^{l+1})\right)^2$$

we obtain

$$\bar{d}_s^2 \leq \frac{81mL_0^2a}{32ks}.$$

Or for the "best" in s steps duality gap we have the following estimation

$$\bar{d}_s \leq 2.25\sqrt{am/2}L_0(ks)^{-0.5} = \mathcal{O}((ks)^{-0.5}).$$

The boundness of $\{x^s\}$ follows from $v_s \to 0$ and boundness of Ω. The boundness of Ω one can assume without restricting the generality. It follows from the Corollary 20

in [10] and boundness of X^*, by adding one extra constraint $c_0(x) = M - f(x) \geq 0$ for $M > 0$ large enough.

4) Summing up (38) for $s = 1, \dots, N$ we obtain

$$k^{-1} \sum_{s=1}^{N} \sum_{i=1}^{m} (\lambda_i^s)^2 \varphi \left(\frac{\lambda_i^{s+1}}{\lambda_i^s} \right) \leq d(\lambda^{N+1}) - d(\lambda^0) \leq d(\lambda^*) - d(\lambda^0)$$

or

$$\lim_{s \to \infty} \sum_{i=1}^{m} (\lambda_i^s)^2 \varphi \left(\frac{\lambda_i^{s+1}}{\lambda_i^s} \right) = 0.$$

Keeping in mind $\lambda_i^s > 0$ and $\varphi \left(\frac{\lambda_i^{s+1}}{\lambda_i^s} \right) \geq 0$ we obtain

$$\lim_{s \to \infty} \left(\lambda_i^s \right)^2 \varphi \left(\frac{\lambda_i^{s+1}}{\lambda_i^s} \right) = 0, \ i = 1, \dots, m. \tag{52}$$

The bounded dual sequence $\{\lambda^s\}$ has a converging subsequence. Without restricting the generality let's assume that

$$\lim_{s \to \infty} \lambda^s = \bar{\lambda}.$$

We will consider two sets of indices $I_0 = \{i : \bar{\lambda}_i = 0\}$ and $I_+ = \{i : \bar{\lambda}_i > 0\}$. For $i \in I_+$ from (52) we have

$$\lim_{s \to \infty} \varphi \left(\frac{\lambda_i^{s+1}}{\lambda_i^s} \right) = 0 \Rightarrow \lim_{s \to \infty} \frac{\lambda_i^{s+1}}{\lambda_i^s} = 1.$$

Then using (20) we have

$$\lim_{s \to \infty} \psi' \left(k_i^s c_i(x^{s+1}) \right) = 1, \ i \in I_+. \tag{53}$$

In view of $\lim_{s \to \infty} \lambda_i^s = \bar{\lambda}_i > 0$, $i \in I_+$ we have $k_i^s \geq 0.5k\bar{\lambda}_i^{-1} > 0$, therefore from (46) we obtain $\lim_{s \to \infty} c_i(x^{s+1}) = 0$, $i \in I_+$.

For $i \in I_0$ due to the primal asymptotic feasibility, which follows from $v_s \to 0$, we obtain $\lim_{s \to \infty} \lambda_i^s c_i(x^s) = 0$, $i \in I_0$. Therefore,

$$\lim_{s \to \infty} \lambda_i^s c_i(x^s) = 0, \quad i = 1, \dots, m.$$

So we proved the asymptotic complementarity condition.

4) Passing to the limit in (23) we obtain

$$\nabla_x L(\bar{x}, \bar{\lambda}) = \nabla f(\bar{x}) - \sum_{i=1}^{m} \bar{\lambda}_i \nabla c_i(\bar{x}) = 0$$

for any limit point $(\bar{x}, \bar{\lambda})$ of $\{x^s, \lambda^s\}$, which together with

$$c_i(\bar{x}) \geq 0, \ \bar{\lambda}_i \geq 0, \ \bar{\lambda}_i c_i(\bar{x}) = 0, \ i = 1, \ldots, m$$

allow to conclude that $\bar{x} = x^*, \bar{\lambda} = \lambda^*$.

From the dual monotonicity we obtain

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

Invoking the asymptotic complementarity we obtain

$$d(\lambda^*) = \lim_{s \to \infty} d(\lambda^s) = \lim_{s \to \infty} L(x^s, \lambda^s) = \lim_{s \to \infty} f(x^s) = f(x^*). \tag{54}$$

Keeping in mind that $X^* = \{x \in \Omega : f(x) \leq f(x^*)\}$ from primal asymptotic feasibility and (54) we obtain $\lim_{s \to \infty} \rho(x^s, X^*) = 0$ (see [24] Lemma 11, Chap. 9, §1).

On the other hand, taking into account $L^* = \{\lambda \in \mathbb{R}^r_{++} : d(\lambda) \geq d(\lambda^*)\}$ and (54) we obtain as a consequence of the same Lemma 11 that $\lim_{s \to \infty} \rho(\lambda^s, L^*) = 0$.

Remark 2. It follows from (44) that for any given $\alpha > 0$ and any $i = 1, \ldots, m$ the inequality $c_i(x^s) \leq -\alpha$ is possible only for a finite number of steps. So the quadratic branch $(c_i(x^s) \leq -\ln 2)$ in the modification (12) can be used only a finite number of times, in fact, for $k > 0$ large enough, just once. Therefore in the asymptotic analysis we can assume that only the FD kernel is used in (33).

To the best of our knowledge the strongest so far result under the assumptions A and B for Interior Prox Method (33) was obtained in [2]. It was proven that for the regularized MBF kernel $\varphi(t) = 0.5\nu(t-1)^2 + \mu(t - \ln t - 1), \ \mu > 0, \ \nu > 0$ the method (33) produces a convergent sequence $\{\lambda^s\}$ and the rate of convergence in value is $\mathcal{O}(ks)^{-1}$.

In the next section we show that this estimation can be strengthened under some extra assumptions on input data by using the similarity between the Interior Prox method (33) and Quadratic Prox for the dual problem in the truncated rescaled dual space.

7. Rate of convergence

In this section we establish the rate of convergence for the LS multipliers method (19)–(21) and its dual equivalent (33) under some extra assumptions on the input data.

We will say that for the converging to (x^*, λ^*) primal-dual sequence $\{x^s, \lambda^s\}_{s=0}^{\infty}$ the complementarity condition is satisfied in the strict form if

$$\max\left\{\lambda_i^*, \ c_i(x^*)\right\} > 0, \ i = 1, \ldots, m. \tag{55}$$

Theorem 3. *If for the primal-dual sequence generated by (19)–(21) the complementarity condition is satisfied in strict form (55) then for any fixed $k > 0$ the following estimation holds true*

$$d(\lambda^*) - d(\lambda^s) = o(ks)^{-1}.$$

Proof. We remind that $I^* = \{i : c_i(x^*) = 0\} = \{1, \dots, r\}$ is the active constraint set, then $\min\{c_i(x^*)|i\bar{\in}I^*\} = \sigma > 0$. Therefore there is such a number s_0 that $c_i(x^s) \geq \sigma/2$, $s \geq s_0$, $i\bar{\in}I^*$. From (20) we have

a) $\lambda_i^{s+1} \leq 2\lambda_i^s(1 + e^{0.5k_i^s\sigma})^{-1} \leq \lambda_i^s e^{-0.5k_i^s\sigma} \to 0$ and   b) $k_i^s = k(\lambda_i^s)^{-1} \to \infty$, $i\bar{\in}I^*$.
$$\tag{56}$$

On the other hand,

$$\mathcal{L}(x, \lambda^s, \mathbf{k}^s) = f(x) - k^{-1}\sum_{i=1}^{r}(\lambda_i^s)^2\psi\left(k_i^s c_i(x)\right) - k^{-1}\sum_{i=r+1}^{m}(\lambda_i^s)^2\psi\left(k_i^s c_i(x)\right).$$

Keeping in mind $\psi(t) \leq 2\ln 2$ and (56) for $s > s_0$ the last term of $\mathcal{L}(x, \lambda^s, \mathbf{k}^s)$ can be estimated by $\mathcal{O}(\sum_{i=r+1}^{m}\lambda_i^{s_0}\exp(-0.5\sigma\sum_{j=0}^{s-s_0}k_i^{s_0+j}))$. So for s_0 large enough and any $s > s_0$ this term is negligibly small, and instead of $\mathcal{L}(x, \lambda, \mathbf{k})$ we can consider the truncated LSL $\mathcal{L}(x, \lambda, \mathbf{k}) := f(x) - \sum_{i=1}^{r}(k_i^s)^{-1}(\lambda_i^s)\psi\left(k_i^s c_i(x)\right)$ and the correspondent truncated Lagrangian $L(x, \lambda) := f(x) - \sum_{i=1}^{r}\lambda_i c_i(x)$. Accordingly, instead of the original dual function and the second order FD distance we consider the dual function $d(\lambda) := \inf_{x\in\mathbb{R}^r} L(x, \lambda)$ and the second order FD distance $D_2(u, v) := \sum_{i=1}^{r} v_i^2\varphi(u_i/v_i)$ in the truncated dual space $\mathbb{R}^r$.

For simplicity we retain the previous notations for the truncated LSL, truncated Lagrangian, correspondent dual function and FD distance.

Below we will assume that $\{\lambda^s\}_{s=1}^{\infty}$, is the truncated dual sequence, i.e. $\lambda^s = (\lambda_1^s, \dots, \lambda_r^s)$.

Let's consider the optimality criteria for the truncated Interior Prox method

$$\lambda^{s+1} = argmax\left\{d(\lambda) - k^{-1}\sum_{i=1}^{r}(\lambda_i^s)^2\varphi(\lambda_i/\lambda_i^s) \,\middle|\, \lambda \in \mathbb{R}^r\right\}.$$

We have

$$c(x^{s+1}) + k^{-1}\sum_{i=1}^{r}\lambda_i^s\varphi'(\lambda_i^{s+1}/\lambda_i^s)e_i = 0, \tag{57}$$

where $e_i = (0, \dots, 1, \dots, 0) \in \mathbb{R}^r$.

Using $\psi^{*'}(1) = \psi^{*'}(\lambda_i^s/\lambda_i^s) = 0$ we can rewrite (57) as follows

$$c(x^{s+1}) + k^{-1}\sum_{i=1}^{r}\lambda_i^s\left(\varphi'(\lambda_i^{s+1}/\lambda_i^s) - \varphi'(\lambda_i^s/\lambda_i^s)\right)e_i = 0.$$

Using the mean value formula we obtain

$$c(x^{s+1}) + k^{-1}\sum_{i=1}^{r}\varphi''\left(1 + \left(\frac{\lambda_i^{s+1}}{\lambda_i^s} - 1\right)\theta_i^s\right)(\lambda_i^{s+1} - \lambda_i^s)e_i = 0, \tag{58}$$

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

We remind that $-c(x^{s+1}) \in \partial d(\lambda^{s+1})$, so (58) is the optimality criteria for the following problem in the truncated dual space

$$\lambda^{s+1} = argmax\{d(\lambda) - 0.5k^{-1}\|\lambda - \lambda^s\|_{R_s}^2 \,|\, \lambda \in \mathbb{R}^r\}, \tag{59}$$

where $\|x\|_R = x^T R x$, $R_s = diag(r_i^s)_{i=1}^r$ and $r_i^s = \varphi''\left(1 + (\frac{\lambda_i^{s+1}}{\lambda_i^s} - 1)\theta_i^s\right) = \varphi''(\cdot)$, $i = 1, \ldots, r$.

Due to $\lambda_i^s \to \lambda_i^* > 0$ we have $\lim_{s\to\infty} \lambda_i^{s+1}/\lambda_i^s = 1, i = 1, \ldots, r$. Keeping in mind the continuity of $\varphi''(\cdot)$ and A6, we have

$$\lim_{s\to\infty} r_i^s = \lim_{s\to\infty} \varphi''\left(1 + \left(\frac{\lambda_i^{s+1}}{\lambda_i^s} - 1\right)\theta_i^s\right) = \varphi''(1) = \min_{0<s<\infty} \varphi''(s) = 2.$$

Therefore for $s_0 > 0$ large enough and any $s > s_0$ we have $2 \le \varphi''(\cdot) \le 3$.

It means that for the active constraints the Interior Prox Method (33) is equivalent to the Quadratic Prox method in the rescaled truncated dual space.

We will show now that the convergence analysis, which is typical for Quadratic Prox method (see [11], [24], [35] and references therein) can be extended for the Interior Prox Method (33) in the truncated dual space.

From (41) we have

$$d(\lambda^*) - d(\lambda^s) - (d(\lambda^*) - d(\lambda^{s+1})) \ge 2k^{-1}\|\lambda^s - \lambda^{s+1}\|^2$$

or

$$v_s - v_{s+1} \ge 2k^{-1}\|\lambda^s - \lambda^{s+1}\|^2, \tag{60}$$

where $v_s = d(\lambda^*) - d(\lambda^s) > 0$. Using the concavity of $d(\lambda)$ we obtain

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

or

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

Using (58) we obtain

$$d(\lambda^{s+1}) - d(\lambda) \ge -k^{-1}\left(R_s(\lambda^{s+1} - \lambda^s), \lambda - \lambda^{s+1}\right).$$

So for $\lambda = \lambda^*$ we have

$$k^{-1}\left(R_s(\lambda^{s+1} - \lambda^s), \lambda^* - \lambda^{s+1}\right) \ge d(\lambda^*) - d(\lambda^{s+1}) = v_{s+1}$$

or

$$k^{-1}\left(R_s(\lambda^{s+1} - \lambda^s), \lambda^* - \lambda^s\right) - k^{-1}\|\lambda^{s+1} - \lambda^s\|_{R_s}^2 \ge v_{s+1}.$$

Hence

$$\|R_s\| \cdot \|\lambda^{s+1} - \lambda^s\| \cdot \|\lambda^* - \lambda^s\| \ge kv_{s+1}$$

or

$$\|\lambda^{s+1} - \lambda^s\| \geq \frac{1}{3}kv_{s+1}\|\lambda^s - \lambda^*\|^{-1}. \tag{61}$$

From (60) and (61) it follows

$$v_s - v_{s+1} \geq \frac{2}{9}kv_{s+1}^2\|\lambda^s - \lambda^*\|^{-2}$$

or

$$v_s \geq v_{s+1}\left(1 + \frac{2}{9}kv_{s+1}\|\lambda^s - \lambda^*\|^{-2}\right).$$

By inverting the last inequality we obtain

$$v_s^{-1} \leq v_{s+1}^{-1}\left(1 + \frac{2}{9}kv_{s+1}\|\lambda^s - \lambda^*\|^{-2}\right)^{-1}. \tag{62}$$

Further,

$$d(\lambda^{s+1}) \geq d(\lambda^{s+1}) - k^{-1}\sum_{i=1}^{r}(\lambda_i^s)^2\varphi(\lambda_i^{s+1}/\lambda_i^s) \geq d(\lambda^*) - k^{-1}\sum_{i=1}^{r}(\lambda_i^s)^2\varphi(\lambda_i^*/\lambda_i^s)$$

or

$$k^{-1}\sum_{i=1}^{r}(\lambda_i^s)^2\varphi(\lambda_i^*/\lambda_i^s) \geq v_{s+1}.$$

Keeping in mind $\varphi(1) = \varphi(v_i^s/v_i^s) = \varphi'(1) = \varphi'(v_i^s/v_i^s) = 0$ and using twice the mean value formula from the last inequality we obtain

$$k^{-1}\sum_{i=1}^{r}\varphi''(\cdot)(\lambda_i^* - \lambda_i^s)^2 = k^{-1}\sum_{i=1}^{r}\varphi''\left(1 + \theta_i'\theta_i''(\lambda_i^* - \lambda_i^s)(\lambda_i^s)^{-1}\right)(\lambda_i^* - \lambda_i^s)^2$$

$$\geq k^{-1}\sum_{i=1}^{r}\varphi''\left(1 + \theta_i'\theta_i''(\lambda_i^* - \lambda_i^s)(\lambda_i^s)^{-1}\right)\theta_i'(\lambda_i^* - \lambda_i^s)^2$$

$$= k^{-1}\sum_{i=1}^{r}(\lambda_i^s)^2\varphi(\lambda_i^*/\lambda_i^s) \geq v_{s+1}$$

where $0 < \theta_i' < 1, 0 < \theta_i'' < 1$. Taking into account $\lim_{s\to\infty}\lambda_i^s = \lambda_i^* > 0, i = 1, \ldots, r$ we obtain $\varphi''(\cdot) \leq 3, s \geq s_0$. Therefore $3\|\lambda^s - \lambda^*\|^2 \geq kv_{s+1}$, i.e.

$$\frac{2}{9}kv_{s+1}\|\lambda^s - \lambda^*\|^{-2} \leq \frac{2}{3}.$$

Considering the function $(1 + t)^{-1}$ it is easy to see that $(1 + t)^{-1} \leq -\frac{3}{5}t + 1$, $0 \leq t \leq \frac{2}{3}$.

Using the last inequality for $t = \frac{2}{9}kv_{s+1}\|\lambda^s - \lambda^*\|^{-2}$ from (62) we obtain

$$v_s^{-1} \le v_{s+1}^{-1} \left(1 - \frac{3}{5} \cdot \frac{2}{9}kv_{s+1}\|\lambda^s - \lambda^*\|^{-2}\right)$$

or

$$v_i^{-1} \le v_{i+1}^{-1} - \frac{2}{15}k\|\lambda^{i+1} - \lambda^*\|^{-2}, \quad i = 0, \dots, s-1. \tag{63}$$

Summing (62) for $i = 0, \dots, s-1$ we obtain

$$v_s^{-1} \ge v_s^{-1} - v_0^{-1} \ge \frac{2}{15}k\sum_{i=0}^{s-1}\|\lambda^i - \lambda^*\|^{-2}$$

By inverting the last inequality we obtain

$$v_s = d(\lambda^*) - d(\lambda^s) \le \frac{15}{2k\sum_{i=0}^{s-1}\|\lambda^i - \lambda^*\|^{-2}}$$

or

$$ksv_s = \frac{15}{\frac{2}{s}\sum_{i=0}^{s-1}\|\lambda^i - \lambda^*\|^{-2}}$$

From $\|\lambda^s - \lambda^*\| \to 0$ follows $\|\lambda^s - \lambda^*\|^{-2} \to \infty$.

Using the Silverman–Toeplitz theorem [16] we have $\lim_{s\to\infty} s^{-1}\sum_{i=1}^{s}\|\lambda^i - \lambda^*\|^{-2} = \infty$. Therefore there exists $\alpha_s \to 0$ that

$$v_s = \frac{15}{2}\alpha_s\frac{1}{ks} = o\big((ks)^{-1}\big). \tag{64}$$

Remark 3. Although the assumption (55) looks restrictive it is always true for classes of Interior Point Methods and MBF in LP (see [30], [38]). In convex optimization it is true, for example, when the Lagrange multipliers for the active constraints are positive and the gradients are independent.

The estimation (64) can be strengthened. Under the standard second order optimality conditions the method (19)–(21) converges with Q-linear rate if $k > 0$ is fixed but large enough.

First of all due to the standard second order optimality conditions the primal–dual solution is unique, therefore the primal $\{x^s\}$ and dual $\{\lambda^s\}$ sequences converge to the primal–dual solution, for which the complementarity conditions are satisfied in a strict form (55).

From (20) we have $\lambda_i^{s+1} \le \lambda_i^s(1 + e^{0.5k_i^s\sigma})^{-1} \le \lambda_i^s e^{-0.5k_i^s\sigma}$. Using $e^x \ge x + 1$ and $k_i^s = k\lambda_i^{s-1}$, $i\bar{\in}I^*$ we obtain $\lambda_i^{s+1} \le \lambda_i^s(0.5k_i^s\sigma + 1)^{-1} \le \frac{2}{k\sigma}\lambda_i^{s2}$, $i\bar{\in}I^*$, i.e. the Lagrange multipliers for the passive constraints converge to zero quadratically.

From (21) we have $\lim_{s\to\infty} k_i^s = k(\lambda_i^*)^{-1}$, $i = 1, \dots, r$, i.e. the scaling parameters, which correspond to the active constraints, grow linearly with $k > 0$.

Therefore the methodology [25], [26] can be applied for the asymptotic analysis of the method (19)–(21).

For a given small enough $\delta > 0$ we define the extended neighborhood of λ^* as follows

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

Theorem 4. *It $f(x)$ and all $c_i(x) \in C^2$ and the standard second order optimality conditions hold, then there exists such a small $\delta > 0$ and large $k_0 > 0$ that for any $(\lambda, \mathbf{k}) \in D(\cdot)$*

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

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

   and

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

2. *for the pair $(\hat{x}, \hat{\lambda})$ the estimate*

$$\max\{\|\hat{x} - x^*\|, \, \|\hat{\lambda} - \lambda^*\|\} \leq ck^{-1}\|\lambda - \lambda^*\|. \tag{65}$$

   holds and $c > 0$ is independent on $k \geq k_0$.
3. *the LSL $\mathcal{L}(x, \lambda, \mathbf{k})$ is strongly convex in the neighborhood of $\hat{x}$.*

Theorem 4 can be proven by slight modification of the correspondent proof in [25] (see also [26]).

Corollary. *If the conditions of Theorem 4 are satisfied then for the primal-dual sequence $\{x^s, \lambda^s\}$ the following estimation is true*

$$\max\{\|x^{s+1} - x^*\|, \, \|\lambda^{s+1} - \lambda^*\|\} \leq \frac{c}{k}\|\lambda^s - \lambda^*\| \leq \left(\frac{c}{k}\right)^{s+1} \|\lambda^0 - \lambda^*\| \tag{66}$$

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

The numerical realization of the LS multipliers method requires to replace the unconstrained minimizer by its approximation. It leads to Newton LS method (see [28]) or to the Primal-Dual LS method [26]. In Sect. 10 we provide numerical results obtained by Newton LS and Primal-Dual LS multipliers method.

8. Generalization and extension

The results obtained for LS remain true for any smoothing function $\theta : \mathbb{R} \to \mathbb{R}_- = \{t : t \leq 0\}$, which is twice continuously differentiable, strictly increasing, strictly concave and satisfies

$$\text{a) } \lim_{t \to \infty} \theta(t) = 0, \quad \text{b) } \lim_{t \to -\infty} (\theta(t) - t) = 0. \tag{67}$$

Let $\sigma = (\theta'(0))^{-1} > 0$, we consider a function $\psi : \mathbb{R} \to (-\infty, -\sigma\theta(0))$ given by formula

$$\psi(t) = \sigma(\theta(t) - \theta(0)). \tag{68}$$

Along with $\psi(t)$ we consider the conjugate function $\psi^*(s) = \inf_t\{st - \psi(t)\}$.

To find $\psi^*(s)$ we have to solve the equation $s - \sigma\theta'(t) = 0$ for t. Due to $\theta''(t) < 0$ the inverse θ'^{-1} exists and $t = \theta'^{-1}(s/\sigma) = \theta^{*'}(s/\sigma)$.

By differentiating the identity $s = \sigma\theta'\left(\theta^{*'}(s/\sigma)\right)$ in s we obtain

$$\sigma\theta''\left(\theta^{*'}(s/\sigma)\right)\theta^{*''}(s/\sigma)\sigma^{-1} = 1.$$

Using again $t = \theta^{*'}(s/\sigma)$ we have

$$\theta^{*''}(s/\sigma) = \left(\theta''(t)\right)^{-1}. \tag{69}$$

Further, for the Fenchel conjugate $\psi^*(s)$ we obtain

$$\psi^*(s) = s\theta^{*'}(s/\sigma) - \psi\left(\theta^{*'}(s/\sigma)\right).$$

Then

$$\psi^{*'}(s) = \theta^{*'}\left(\frac{s}{\sigma}\right) + \frac{s}{\sigma}\theta^{*''}\left(\frac{s}{\sigma}\right) - \frac{1}{\sigma}\psi'\left(\theta^{*'}\left(\frac{s}{\sigma}\right)\right)\theta^{*''}\left(\frac{s}{\sigma}\right).$$

Keeping in mind $t = \theta^{*'}(s/\sigma)$ and $\psi'(t) = s$ we obtain

$$\psi^{*'}(s) = \theta^{*'}\left(\frac{s}{\sigma}\right).$$

Then using (69) we have

$$\psi^{*''}(s) = \sigma^{-1}\theta^{*''}\left(\frac{s}{\sigma}\right) = \frac{1}{\sigma\theta''(t)}. \tag{70}$$

Let $\theta''(t_0) = \min\{\theta''(t) \mid -\infty < t < \infty\}$, such minimum exists due to the continuity of $\theta''(t)$ and because $\lim_{t\to-\infty}\theta''(t) = \lim_{t\to\infty}\theta''(t) = 0$, which follows from (67).

The following assertion states the basic properties of the transformation ψ.

Proposition 3. *Let $\theta \in C^2$ be a strictly increasing and strictly concave function, which satisfies (67), then*

C1. $\psi(0) = 0$,
C2. *a)* $0 < \psi'(t) < \sigma$, $\forall t \in (-\infty, \infty)$ *and b)* $\psi'(0) = 1$,
C3. $\lim_{t\to-\infty}\psi'(t) = \sigma$; $\quad \lim_{t\to\infty}\psi'(t) = 0$,
C4. $\tau = \sigma\theta''(t_0) < \psi''(t) < 0$, $\forall t \in (-\infty, \infty)$,
C5. $\psi^{*''}(s) \leq \tau^{-1}$, $\forall s \in (0, \sigma)$.

The properties C1–C3 one can verify directly, properties C4–C5 follow from (70).

Now we consider the kernel $\varphi = -\psi^*$.

Proposition 4. *The kernel φ possesses the following properties.*

D1. $\varphi(s)$ is nonnegative and strongly convex on $(0, \sigma)$,
D2. $\varphi(1) = \varphi'(1) = 0$,
D3. $\lim_{s \to 0^+} \varphi'(s) = -\infty$,
D4. $\varphi''(s) \geq - \left(\sigma \theta''(t_0) \right)^{-1}$, $\forall s \in (0, \sigma)$.

Therefore for each smoothing function $\theta(t)$, which satisfies the conditions of Proposition 3, one can find a constrained transformation with properties A1–A6.

However due to $\lim_{t \to -\infty} \psi'(t) = \sigma$ we have $\lim_{t \to -\infty} \psi''(t) = 0$, therefore $\lim_{s \to \sigma^-} \psi^{*''}(s) = -\infty$ and $\lim_{s \to \sigma^-} \varphi''(s) = \infty$.

To avoid the complications, which we discussed in Sect. 4, we have to modify $\psi(t)$.

We will illustrate it using Chen-Harker-Kanzow-Smale (CHKS) smoothing function, which along with Log-Sigmoid function has been widely used for solving complementarity problems, see e.g. [7], [15] and references therein.

For a given $\eta > 0$ the following equation $z^2 - tz - \eta = 0$ has two roots $\theta_-(t) = 0.5 \left(t - \sqrt{t^2 + 4\eta} \right)$ and $\theta_+(t) = 0.5 \left(t + \sqrt{t^2 + 4\eta} \right)$. The function $\theta_- : (-\infty, \infty) \to (-\infty, 0)$ is CHKS interior smoothing function (see [7],[15]). We consider smoothing CHKS function $\theta \equiv \theta_- : (-\infty, \infty) \to (-\infty, 0)$ given by formula $\theta(t) = 0.5 \left(t - \sqrt{t^2 + 4\eta} \right)$, then

$$\theta''(0) = \min\{\theta''(t) = -4\eta(t^2 + 4\eta)^{-3/2} \mid -\infty < t < \infty\} = -0.5\eta^{-0.5},$$

$\sigma = (\theta'(0))^{-1} = 2$, and $\theta(0) = -\sqrt{\eta}$. The transformation $\psi : (-\infty, \infty)$, which is given by formula

$$\psi(t) = t - \sqrt{t^2 + 4\eta} + 2\sqrt{\eta}$$

we call the CHKS transformation.

It is easy to see that C1–C5 from Assertion 8.1 hold true for $\psi(t)$ given by the above formula with $\sigma = 2$ and $\tau = \min_t \psi''(t) = -\max_t 4\eta(t^2 + 4\eta)^{-\frac{3}{2}} = -(2\sqrt{\eta})^{-1}$.

The Fenchel conjugate $\psi^* : (0, 2) \to [0, -2\sqrt{\eta}]$ is defined by formula

$$\psi^*(s) = \inf_t \{st - \psi(t)\} = 2\sqrt{\eta} \left(\sqrt{(2 - s)s} - 1 \right).$$

Then $\psi^*(0) = \psi^*(2) = -2\sqrt{\eta}$ and

$$(0, 2) = ridom\ \psi^* \subset range\ \psi' \subset dom\ \psi^* = [0, 2].$$

The function $\varphi : [0, 2] \to [0, 2\sqrt{\eta}]$, which is given by formula $\varphi(s) = -\psi^*(s) = 2\sqrt{\eta} \left(1 - \sqrt{(2 - s)s} \right)$, we will call CHKS kernel.

The second order CHKS φ–divergence distance $D : \mathbb{R}_+^m \times \mathbb{R}_{++}^m \to \mathbb{R}_+$ we define by formula

$$D(u, v) = \sum_{i=1}^m v_i^2 \varphi(u_i/v_i).$$

The properties of the CHKS kernel $\varphi(t) = 2\sqrt{\eta}\left(1 - \sqrt{(2-t)t}\right)$ are similar to those of the FD kernel φ:

E1. $\varphi(t) \geq 0$, $t \in [0, 2]$, φ is strongly convex on $(0,2)$,
E2. $\varphi(1) = \varphi'(1) = 0$,
E3. $\varphi'(t) = 2\sqrt{\eta}(t - 1)\left(t(2 - t)\right)^{-0.5}$, $\lim\limits_{t \to 0^+} \varphi'(t) = -\infty$, $\lim_{t \to 2^-} \varphi'(t) = \infty$,
E4. $\varphi''(t) = 2\sqrt{\eta}\left(t(2 - t)\right)^{-1.5} \geq 2\sqrt{\eta}$, $\forall t \in (0, 2)$.

To avoid complications, which we discussed in Sect. 4 CHKS transformation $\psi(t)$ and the corresponding kernel $\varphi(t)$ can be modified the same way it was done for LS transformation in Sect. 4.

The modified CHKS transformation $\bar{\psi} : (-\infty, \infty) \to (-\infty, 2\sqrt{\eta})$ we define by formula

$$\bar{\psi}(t) = \begin{cases} \psi(t) = t - \sqrt{t^2 + 4\eta} + 2\sqrt{\eta}, & t \geq -\sqrt{\eta} \\ q(t) = at^2 + bt + c, & -\infty < t \leq -\sqrt{\eta}. \end{cases} \tag{71}$$

Parameters a, b and c we find from the system $\psi(-\sqrt{\eta}) = q(-\sqrt{\eta})$, $\psi'(-\sqrt{\eta}) = q'(-\sqrt{\eta})$, $\psi''(-\sqrt{\eta}) = 2a$ to ensure $\bar{\psi} \in C^2$. Then instead of C2 a) we obtain

$$0 < \bar{\psi}'(t) < \infty. \tag{72}$$

The Fenchel conjugate $\bar{\psi}^*$ is defined by

$$\bar{\psi}^*(s) = \begin{cases} \psi^*(s) = 2\sqrt{\eta}\left(\sqrt{(2-s)s} - 1\right), & 0 < s \leq \psi'(-\eta) = 1 + \frac{1}{\sqrt{5}} \\ q^*(s) = (4a)^{-1}(s - b)^2 - c, & 1 + \frac{1}{\sqrt{5}} < s < \infty. \end{cases}$$

The modified CHKS kernel $\bar{\varphi} : (0, \infty) \to (0, \infty)$ we define by formula

$$\bar{\varphi}(t) = \begin{cases} -\psi^*(s), & 0 < s \leq \psi'(-\eta) = 1 + \frac{1}{\sqrt{5}} \\ -q^*(s), & 1 + \frac{1}{\sqrt{5}} < s < \infty. \end{cases} \tag{73}$$

The following proposition states the properties of the modified CHKS kernel $\bar{\varphi}$.

Proposition 5. *The modified CHKS kernel $\bar{\varphi}$ possesses the following properties*

F1. $\bar{\varphi}(t) \geq 0$, $t \in (0, \infty)$,
F2. $\bar{\varphi}(1) = \bar{\varphi}'(1) = 0$,
F3. $\lim\limits_{s \to 0^+} \bar{\varphi}'(s) = -\infty$, $\lim_{s \to \infty} \bar{\varphi}'(t) = \infty$,
F4. $\bar{\varphi}''(s) \geq 2\sqrt{\eta}$, $s \in (0, \infty)$.
F5. $\bar{\varphi}''(s) \leq 1.25(5\eta)^{0.5}$, $1 \leq s < \infty$.

The multipliers method (19)–(21) with CHKS transformation is equivalent to the Interior Prox method (33) with the second order distance function, which is based on the CHKS kernel. By repeating the arguments from Sects. 6 and 7 it is easy to see that Theorems 2, 3, 4 remain true for methods (19)–(21) and (33) when modified LS is replaced by modified CHKS transformation.

In other words the following theorem holds.

Theorem 5. *For the primal-dual sequence* $\{x^s, \lambda^s\}$, *which is generated by method* (19)–(21) *with modified CHKS transformation given by* (71), *the following is true.*

1. *If* **A** *and* **B** *is satisfied then*

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

   and $\lim_{s \to \infty} \rho(x^s, X^*) = 0$, $\lim_{s \to \infty} \rho(\lambda^s, L^*) = 0$.
2. *If* (55) *is satisfied then*

$$d(\lambda^*) - d(\lambda^s) = o\big((ks)^{-1}\big).$$

3. *If* (4)–(5) *is satisfied, then for the primal-dual sequence generated by NR method with modified CHKS transformation the estimation* (65) *holds.*

For any transformation ψ given by (68) the corresponding modification with quadratic extrapolation at the point $t_0 < 0$ is given by the following formula

$$\bar{\psi}(t) = \begin{cases} \psi(t), & t \geq t_0, \\ q(t), & t \leq t_0, \end{cases}$$

where $q(t) = at^2 + bt + c$ and $a = 0.5\psi''(t_0)$, $b = \psi'(t_0) - t_0\psi''(t_0)$, $c = \psi(t_0) - t_0\psi'(t_0) + 0.5t_0^2\psi''(t_0)$.

Before concluding this section we would like to make a few comments about exponential multipliers method with "dynamic" scaling parameters update, which was introduced in [37].

As we mentioned already it was proven in [37] that for the exponential multipliers method with a fixed scaling parameter the ergodic convergence takes place.

At the same time the authors in [37] emphasized that the analysis for the "dynamic" scaling parameter update turns out to be very difficult: "we have been unable to show a correspondent result, ... even though the method in practice seems equally reliable" (see [37], p. 3).

It turns out that all results of the Theorem 2, 3 and 4 remain true for exponential multipliers method with a slightly modified exponential transformation $\psi(t) = 1 - e^{-t}$.

First of all let's consider the conjugate function $\psi^*(s) = \inf_t\{st - \psi(t)\} = -s \ln s + s - 1$. The kernel of the correspondent second order Entropy-like distance function is $\varphi(t) = t \ln t - t + 1$, then $\varphi'(t) = \ln t$, $\varphi''(t) = t^{-1}$ and $\lim_{t \to \infty} \varphi''(t) = 0$. Therefore the kernel φ does not satisfy the property $\bar{B}4$, which plays the key role in Theorem 2 and 3.

We modify the exponential transformation $\psi(t) = 1 - e^{-t}$ the same way it has been done for LS transformation. Let's consider

$$\bar{\psi}(t) = \begin{cases} \psi(t), & t \geq -1 \\ q(t) = at^2 + bt + c, & t \leq -1. \end{cases}$$

The parameters a, b and c we find from $\psi(-1) = q(-1)$, $\psi'(-1) = q'(-1)$ and $\psi''(-1) = q''(-1)$ to ensure that $\bar{\psi} \in C^2$. So $a = -0.5e$, $b = 0$, $c = 1 - 0.5e$ and

$$\bar{\psi}(t) = \begin{cases} 1 - e^{-t}, & t \geq -1 \\ q(t) = -0.5et^2 + 1 - 0.5e, & t \leq -1, \end{cases} \tag{74}$$

then

$$\bar{\psi}^*(s) = \begin{cases} \psi^*(s) = -s \ln s + s - 1, & s \leq e \\ q^*(s) = (4a)^{-1}(s-b)^2 - c = (-2e)^{-1}s^2 - 1 + 0.5e, & s \geq e. \end{cases}$$

Now the kernel $\bar{\varphi} : (0, \infty) \rightarrow (0, \infty)$ of the second order φ–divergence distance, which corresponds to the exponential multipliers method with "dynamic" scaling parameters update is given by formula

$$\bar{\varphi}(t) = \begin{cases} t \ln t - t + 1, & 0 \leq t \leq e \\ -q^*(s) = (2e)^{-1}t^2 + 1 - 0.5e, & t \geq e. \end{cases}$$

Then

$$\bar{\varphi}''(t) = \begin{cases} t^{-1}, & 0 < t \leq e \\ e^{-1}, & t \geq e. \end{cases}$$

So $\min\{\varphi''(t) \,|\, t > 0\} = e^{-1}$, i.e. for the modified exponential transformation the property type $\bar{B}4$ holds. At the same time $\bar{\varphi}''(t) \leq 1$, $\forall t \in [1, \infty]$, i.e. the property type $\bar{B}5$ holds as well. Therefore all results from Theorems 2, 3 and 4 remain true for this transformation. In other words the following proposition is taking place.

Proposition 6. *For the primal–dual sequence $\{x^s, \lambda^s\}$, which is generated by* (19)–(21) *with the modified exponential transformation* (74) *all statements of Theorem 5 remain true.*

Keeping in mind Remark 2 which remains true for the modified exponential transformation, we can conclude that only exponential branch of transformation $\bar{\psi}$ controls the computational process from some point on. Therefore the exponential multipliers method with dynamic scaling parameter update (see [37]) converges under very mild assumptions A and B. Under the strict complementarity conditions (55) it converges with $o\left((ks^{-1})\right)$ rate and under the standard second order optimality condition (4)–(5) it converges with Q-linear rate if $k > 0$ is fixed but large enough.

It confirms the observation in [37] that exponential method with fixed and "dynamic" scaling parameters update are equally reliable. It is worth to mention that the modification (71) prevents the exponential transformation and its derivatives from the exponential growth in case of constraints violation, which contributes to the numerical stability.

Remark 4. The logarithmic $\psi_1(t) = \ln(t + 1)$, hyperbolic $\psi_2(t) = t(t + 1)^{-1}$ and parabolic $\psi_3(t) = 2\left(\sqrt{t+1} - 1\right)$ transformations, which lead to the correspondent Modified Barrier Functions [25] can be modified the way it was done with the exponential transformation. Then all results of Theorems 2, 3, and 4 will remain true for the correspondent multipliers method with "dynamic" scaling parameter update.

In the next section we apply the LS method for Linear Programming. The convergence under very mild assumption follows from Theorem 2. Under the dual uniqueness we prove the global quadratic convergence. The key ingredients of the proof are the A. Hoffman type Lemma (see [13]), Theorem 1 and the properties of the FD kernel.

9. Log-Sigmoid Multipliers method for Linear Programming

Let $A : \mathbb{R}^n \to \mathbb{R}^m$, $a \in \mathbb{R}^n$, $b \in \mathbb{R}^m$. We assume that

$$X^* = Argmin\{(a, x) \,|\, c_i(x) = (Ax - b)_i = (a_i, x) - b_i \geq 0, \ i = 1, \ldots, m\} \quad (75)$$

and

$$L^* = Argmax\{(b, \lambda) \,|\, A^T \lambda - a = 0, \ \lambda_i \geq 0, \ i = 1, \ldots, m\} \quad (76)$$

are nonempty and bounded.

The LS method (19)–(21) being applied to (75) produces three sequences $\{x^s\}$, $\{\lambda^s\}$, and $\{\mathbf{k}^s\}$:

$$x^{s+1} : \nabla_x \mathcal{L}(x^{s+1}, \lambda^s, \mathbf{k}^s) = a - \sum_{i=1}^{m} \lambda_i^s \psi'\big(k_i^s c_i(x^{s+1})\big)a_i = 0, \quad (77)$$

$$\lambda^{s+1} : \lambda_i^{s+1} = \lambda_i^s \psi'\big(k_i^s c_i(x^{s+1})\big), \ i = 1, \ldots, m. \quad (78)$$

$$\mathbf{k}^{s+1} : k_i^{s+1} = k\big(\lambda_i^{s+1}\big)^{-1}, \ i = 1, \ldots, m. \quad (79)$$

If X^* and L^* are bounded then all statements of Theorem 2 are taking place for the primal-dual sequence $\{x^s, \lambda^s\}$ generated by (77)–(79). In particular

$$\lim_{s \to \infty} (a, x^s) = (a, x^*) = \lim_{s \to \infty} (b, \lambda^s) = (b, \lambda^*).$$

Using Lemma 5 (see [24], Chap. 10, §1) we can find such $\alpha > 0$ that

$$(b, \lambda^*) - (b, \lambda^s) \geq \alpha \rho(\lambda^s, L^*). \quad (80)$$

Therefore $\lim_{s \to \infty} \rho(\lambda^s, L^*) = 0$.

If λ^* is a unique dual solution then the same Lemma 5 guarantees the existence of such $\alpha > 0$ that

$$(b, \lambda^*) - (b, \lambda) = \alpha \|\lambda - \lambda^*\|. \quad (81)$$

holds true for $\forall \lambda \in L = \{\lambda : A^T \lambda = a, \lambda \in \mathbb{R}_+^m\}$.

Theorem 6. *If the dual problem* (76) *has a unique solution, then the dual sequence* $\{\lambda^s\}$ *converges in value quadratically, i.e. there is $c > 0$ independent on $k > 0$ that the following estimation*

$$(b, \lambda^*) - (b, \lambda^{s+1}) \leq ck^{-1}\big[(b, \lambda^*) - (b, \lambda^s)\big]^2. \quad (82)$$

holds true.

Proof. It follows from (77)–(79) that

$$\nabla_x \mathcal{L}(x^{s+1}, \lambda^s, k^s) = a - \sum_{i=1}^{m} \lambda_i^{s+1} a_i = a - A^T \lambda^{s+1} = 0 \qquad (83)$$

and $\lambda^{s+1} \in \mathbb{R}_{++}^m$. In other words the LS method generates dual interior point sequence $\{\lambda^s\}_{s=0}^{\infty}$.

From (83) we obtain

$$0 = (a - A^T \lambda^{s+1}, x^{s+1}) = (a, x^{s+1}) - \sum_{i=1}^{m} \lambda_i^{s+1} c_i(x^{s+1}) - (b, \lambda^{s+1})$$

or

$$(b, \lambda^{s+1}) = L(x^{s+1}, \lambda^{s+1}).$$

From Theorem 1 we obtain the equivalence of the multipliers method (77)–(79) to the following Interior Prox for the dual problem

$$\lambda^{s+1} = argmax \left\{ (b, \lambda) - k^{-1} \sum_{i=1}^{m} (\lambda_i^s)^2 \varphi \left(\frac{\lambda_i}{\lambda_i^s} \right) \,\middle|\, A^T \lambda - a = 0 \right\}. \qquad (84)$$

Keeping in mind Remark 2 we can assume without restricting the generality that only the Fermi-Dirac kernel $\varphi(t) = (2 - t) \ln(2 - t) + t \ln t$ is used in the method (84).

From (84) taking into account $\lambda^* \in \mathbb{R}_+^m$ and $A^T \lambda^* = a$ we obtain

$$(b, \lambda^{s+1}) - k^{-1} \sum_{i=1}^{m} (\lambda_i^s)^2 \varphi \left(\frac{\lambda_i^{s+1}}{\lambda_i^s} \right) \geq (b, \lambda^*) - k^{-1} \sum_{i=1}^{m} (\lambda_i^s)^2 \varphi \left(\frac{\lambda_i^*}{\lambda_i^s} \right).$$

Keeping in mind $k^{-1} \sum_{i=1}^{m} (\lambda_i^s)^2 \varphi \left(\frac{\lambda_i^{s+1}}{\lambda_i^s} \right) \geq 0$ we have

$$k^{-1} \sum_{i=1}^{m} (\lambda_i^s)^2 \varphi \left(\frac{\lambda_i^*}{\lambda_i^s} \right) \geq (b, \lambda^*) - (b, \lambda^{s+1}). \qquad (85)$$

Let's assume $\lambda_i^* > 0$, $i = 1, \ldots, r$; $\lambda_i^* = 0$, $i = r + 1, \ldots, m$. Then $\varphi \left(\frac{\lambda_i^*}{\lambda_i^s} \right) = 2 \ln 2$, $i = r + 1, \ldots, m$. Taking into account $\varphi(1) = \varphi'(1) = \varphi \left(\frac{\lambda_i^s}{\lambda_i^s} \right) = \varphi' \left(\frac{\lambda_i^s}{\lambda_i^s} \right) = 0$ and using the mean value formula twice we obtain

$$k^{-1} \sum_{i=1}^{m} (\lambda_i^s)^2 \varphi \left(\frac{\lambda_i^*}{\lambda_i^s} \right) = k^{-1} \left[\sum_{i=1}^{r} (\lambda_i^s)^2 \left(\varphi \left(\frac{\lambda_i^*}{\lambda_i^s} \right) - \varphi \left(\frac{\lambda_i^s}{\lambda_i^s} \right) \right) \right.$$

$$\left. + 2 \ln 2 \sum_{i=r+1}^{m} (\lambda_i^* - \lambda_i^s)^2 \right]$$

$$= k^{-1}\left[\sum_{i=1}^{r}\lambda_i^s\left[\varphi'\left(1+\theta_i'\frac{\lambda_i^*-\lambda_i^s}{\lambda_i^s}\right)-\varphi'\left(\frac{\lambda_i^s}{\lambda_i^s}\right)\right](\lambda_i^*-\lambda_i^s)\right.$$

$$\left.+2\ln 2\sum_{i=r+1}^{m}\left(\lambda_i^*-\lambda_i^s\right)^2\right]$$

$$= k^{-1}\left[\sum_{i=1}^{r}\varphi''\left(1+\theta_i''\theta_i'\frac{\lambda_i^*-\lambda_i^s}{\lambda_i^s}\right)(\lambda_i^*-\lambda_i^s)^2\theta_i'\right.$$

$$\left.+2\ln 2\sum_{i=r+1}^{m}\left(\lambda_i^*-\lambda_i^s\right)^2\right]$$

$$= k^{-1}\left[\sum_{i=1}^{r}\varphi''(\cdot)\left(\lambda_i^*-\lambda_i^s\right)^2\theta_i'+2\ln 2\sum_{i=r+1}^{m}\left(\lambda_i^*-\lambda_i^s\right)^2\right]$$

$$\leq k^{-1}\left[\sum_{i=1}^{r}\varphi''(\cdot)\left(\lambda_i^*-\lambda_i^s\right)^2+2\ln 2\sum_{i=r+1}^{m}\left(\lambda_i^*-\lambda_i^s\right)^2\right],$$

where $0<\theta_i'<1, 0<\theta_i''<1$.

Taking into account the dual uniqueness from (80) we obtain $\lim_{s\to\infty}\lambda_i^s=\lambda_i^*>0$, $i=1,\ldots,r$, therefore $\lim_{s\to\infty}\varphi''\left(1+\theta_i''\theta_i'\frac{\lambda_i^*-\lambda_i^s}{\lambda_i^s}\right)=\varphi''(1)=2, i=1,\ldots,r$. Hence there is such s_0 that for any $s\geq s_0$ we have $\varphi''(\cdot)\leq 3$. Hence

$$k^{-1}\sum_{i=1}^{m}(\lambda_i^s)^2\varphi\left(\frac{\lambda_i^*}{\lambda_i^s}\right)\leq 3k^{-1}\|\lambda^*-\lambda^s\|^2. \tag{86}$$

Combining (85) and (86) we have

$$3k^{-1}\|\lambda^*-\lambda^s\|^2\geq(b,\lambda^*)-(b,\lambda^{s+1}). \tag{87}$$

From (81) with $\lambda=\lambda^s$ we obtain

$$\|\lambda^s-\lambda^*\|=\alpha^{-1}\left[(b,\lambda^*)-(b,\lambda^s)\right].$$

Therefore the following estimation

$$(b,\lambda^*)-(b,\lambda^{s+1})\leq ck^{-1}\left[(b,\lambda^*)-(b,\lambda^s)\right]^2 \tag{88}$$

holds true with $c=3\alpha^{-2}$ for any $s\geq s_0$. It follows from Theorem 2 that by taking $k>0$ large enough one can make $s_0=1$.

Remark 5. Theorem 6 remains valid for the method (77)–(79) when the LS transformation is replaced by CHKS transformation given by (71) or the exponential transformation given by (74).

Remark 6. If the dual solution is not unique, it can be fixed by a slight change of vector b (see [24], p. 316).

10. Numerical results

The numerical realization of the LS method (19)–(21) requires to replace the unconstrained minimizer x^{s+1} by an approximation. Then the approximation is used for both Lagrange multipliers and scaling parameters update. To find an approximation $\bar{x}^{s+1}$ for x^{s+1} we applied Newton method with steplength for minimization $\mathcal{L}(x, \lambda^s, \mathbf{k}^s)$ in x. Such approach leads to the Newton LS method (see [28]). The other approach consists of using Newton method for solving nonlinear system of equations, which consists of the primal optimality criteria equations for LSL minimizer and equations for the Lagrange multipliers update. Such approach leads to the Primal-Dual LS method [26].

The numerical results obtained using both approaches allowed systematically observe the "hot start" phenomenon, which was described for MBF in [21], [25].

Practically speaking the "hot start" means that from some point on the primal-dual approximation will remain in the Newton area after each Lagrange multipliers and scaling parameters update.

It means that only few updates and very few (often one) Newton's steps per update requires to reduce the duality gap and primal infeasibility by several orders of magnitude.

We illustrate it on the following NLP and LP problems, where n is the number of variables, m is the number of equations and p is the number of inequalities.

For the NLP examples we show the norm of $\nabla_x \mathcal{L}$, duality gap, primal infeasibility. For LP calculations we show the duality gap, the primal and dual infeasibility.

Although Theorem 6 is proven under the assumption that the dual LP has a unique solution, we systematically observed the quadratic convergence for the constrained violation and the duality gap for all LP, which have been solved.

Table 1. Name: structure4; Objective: linear, Constraints: convex quadratic. $n = 1536$, $p = 720$

it	$\|\nabla_x \mathcal{L}\|$	gap	inf	# of steps
0	1.001756e+00	1.999851e+00	0.000000e+00	0
1	6.431239e+00	4.230525e-01	1.047209e-02	12
2	1.394687e+00	3.544221e-01	2.237616e-03	8
3	4.886293e-01	1.578621e-01	5.551414e-04	8
4	3.116503e-01	3.861605e-02	2.987699e-04	6
5	3.143736e-01	5.248152e-03	1.753116e-04	3
6	2.447404e-01	8.953510e-04	7.933149e-05	3
7	3.429093e-05	1.332356e-04	5.928732e-05	1
8	3.559139e-05	5.906833e-05	3.667183e-05	3
9	3.048857e-05	1.543989e-05	1.865640e-05	4
10	2.057229e-05	2.905176e-06	7.497569e-06	3
11	1.053395e-05	3.391419e-07	2.413493e-06	4
12	3.933137e-06	2.161987e-08	5.415852e-07	5
13	1.012385e-06	6.686092e-10	8.462800e-08	3
14	1.698804e-07	1.003167e-11	8.575809e-09	3
15	1.791739e-08	5.971439e-13	5.287028e-10	2
Total number of Newton steps				68

Table 2. Name: nnls; Nonnegative Least Squares Models; Objective: convex quadratic, Constraints: bounds. $n = 300$, $p = 300$

it	$\|\nabla_x \mathcal{L}\|$	gap	inf	# of steps
0	5.721662e+06	6.072350e+01	2.038436e+00	0
1	1.566275e-04	3.508411e-06	7.797084e-10	2
2	2.483283e-05	2.146990e-09	9.072215e-12	1
Total number of Newton steps				3

Table 3. Name: antenna; Objective: linear; Constraints: convex quadratic. $n = 49$, $m = 10$, $p = 156$

it	$\|\nabla_x \mathcal{L}\|$	gap	inf	# of steps
0	2.967835e+05	7.677044e-01	9.134264e-05	0
1	1.103439e+02	1.795130e-02	2.262034e-03	15
2	4.304839e+00	1.056079e-02	2.217783e-05	7
3	3.339791e+00	7.387677e-03	1.803251e-05	6
4	3.174754e+00	3.824959e-03	9.352697e-06	6
5	1.622338e+00	1.144959e-03	2.812697e-06	6
6	4.724850e-01	1.944959e-04	9.352697e-06	6
7	8.286059e-02	1.940084e-05	4.821745e-08	3
8	8.729887e-03	1.161975e-06	2.891458e-09	5
9	5.493163e-04	4.173671e-08	1.037934e-10	7
10	9.527629e-05	9.817914e-10	7.951639e-12	6
Total number of Newton steps				67

Table 4. Name: trafequil; Objective: convex nonlinear; Constraints: linear. $n = 722$, $m = 361$, $p = 722$

it	$\|\nabla_x \mathcal{L}\|$	gap	inf	# of steps
0	4.632893e+01	3.347643e+01	4.840000e+00	0
1	8.747497e+00	1.710541e-01	5.849002e-03	18
2	1.636448e-04	2.166123e-03	4.290428e-04	13
3	7.108554e-06	2.918467e-05	7.591514e-06	6
4	6.158934e-07	1.045222e-06	8.207374e-07	4
5	1.305577e-10	4.493538e-13	1.338086e-10	2
6	3.672736e-09	1.344654e-09	2.130356e-09	2
7	2.617049e-09	4.599113e-11	8.222189e-11	1
Total number of Newton steps				46

Table 5. Name: markowitz2; Objective: convex quadratic; Constraints: linear. $n = 1200, m = 201, p = 1000$

it	$\|\nabla_x \mathcal{L}\|$	gap	inf	# of steps
0	3.162834e+05	7.032438e+01	1.495739e+00	0
1	1.821451e-02	9.001130e-02	5.904234e-05	10
2	1.513762e-03	4.205607e-03	3.767383e-06	12
3	5.716835e-05	6.292277e-05	2.654451e-05	13
4	3.737993e-06	1.709659e-06	1.310097e-05	8
5	4.736475e-07	1.074959e-07	1.381697e-06	5
6	7.101903e-08	7.174959e-09	3.368086e-07	4
7	1.070573e-08	4.104959e-10	3.958086e-08	3
8	1.355662e-09	1.749759e-11	2.868086e-09	2
9	1.305577e-10	4.493538e-13	1.338086e-10	2
Total number of Newton steps				59

Table 6. Name: Israel. $n = 174, p = 316$

it	gap	primal inf	dual inf	# of steps
0	1.05e+10	1.21e+06	1.78e+04	0
1	6.40e+00	2.77e-09	7.53e-10	20
2	7.364041e-02	1.0729e-07	0.00e+00	14
3	7.497715e-07	4.2011e-12	0.00e+00	6
4	1.628188e-10	1.7764e-15	0.00e+00	3
Total number of Newton steps				43

Table 7. Name: AGG. $n = 488, p = 615$

it	gap	primal inf	dual inf	# of steps
0	4.29e+11	3.78e+07	2.49e+04	0
1	1.23e+00	3.01e-06	7.16e-10	19
2	2.942798e-04	2.7691e-10	0.00e+00	3
3	3.949872e-09	2.8421e-14	0.00e+00	3
Total number of Newton steps				25

Table 8. Name: AGG2. $n = 516, p = 758$

it	gap	primal inf	dual inf	# of steps
0	6.93e+10	7.41e+06	2.08e+04	0
1	6.07e+00	4.39e-07	5.41e-10	16
2	1.422620e-03	2.2625e-09	0.00e+00	3
3	2.630272e-10	7.1054e-15	0.00e+00	3
Total number of Newton steps				25

Table 9. Name: BNL1. $n = 516$, $p = 758$

it	gap	primal inf	dual inf	# of steps
0	3.98e+06	1.83e+04	8.41e+02	0
1	9.47e-05	7.13e-09	3.81e-12	25
2	2.645905e-07	3.8801e-10	0.00e+00	5
3	2.025197e-12	4.5938e-13	0.00e+00	4
Total number of Newton steps				34

11. Concluding remarks

We mentioned already that LS multipliers method is to the smoothing technique [1], [8] as Quadratic Augmented Lagrangian [12], [29] to the penalty method [9] for equalities constraints or MBF to Log-barrier SUMT [10] for inequalities constraints. In the dual space the Interior Prox method with Fermi–Dirac second order distance is to regularization method [1] as Quadratic-Prox [20], [32] to N. Tikhonov's regularization technique [36] or Interior Prox method with φ-divergence distance [27] to path following methods.

Although in Augmented Lagrangians, MBF or LS methods, there exists an exchange of the information between the primal and dual space, the calculations are always conducted sequentially: first is the primal minimization then the dual Lagrange multipliers update. On the other hand, it has become evident lately that the most efficient methods, which are based on path-following ideas are Primal-Dual methods for which the calculations are conducted simultaneously in the primal and dual spaces. In other words at least for LP the most efficient are the Primal-Dual Interior Point methods [38].

For each NR multipliers method there exists the Primal-Dual equivalent (see [26]). Our experiments with the Primal-Dual NR methods in general and with the Primal-Dual LS method in particular are very encouraging.

Currently both the theoretical and the experimental research on the Primal-Dual NR methods for NLP calculations are in the early stage. Global convergence, rate of convergence, "hot start" phenomenon, complexity issues, efficient numerical implementation of the Primal-Dual NR multipliers methods were left for further research.

References

1. Auslender, A., Cominetti, R., Haddou, M. (1997): Asymptotic analysis for penalty and barrier method in convex and linear programming. Math. Oper. Res. **22**, 43–62
2. Auslender, A., Teboulle, M., Ben-Tiba, S. (1999): Interior Proximal and Multipliers Methods based on second order Homogenous kernels. Math. Oper. Res. **24**(3), 645–668
3. Ben-Tal, A., Zibulevski, M. (1997): Penalty-Barrier Methods for convex programming problems. SIAM J. Optim. **7**, 347–366

4. Ben-Tal, A., Yuzefovich, B., Zibulevski, M. (1992): Penalty-Barrier Multipliers methods for minimax and constrained smooth convex optimization. Research Report 9-92, Optimization Laboratory, Technion, Israel
5. Breitfelt, M., Shanno, D. (1996): Experience with Modified Log-Barrier Method for Nonlinear Programming. Ann. Oper. Res. **62**, 439–464
6. Bertsekas, D. (1982): Constrained Optimization and Lagrange Multipliers Methods. Academic Press, New York
7. Chen, B., Harker, P.T. (1933): A non-interior point continuation method for Linear complementarity problems. SIAM J. Matrix Anal. Appl. **14**, 1168–1190
8. Chen, C., Mangasarian, O. (1995): Smoothing methods for convex inequalities and linear complementarity problems. Math. Program. **71**, 51–69
9. Courant, R. (1943): Variational Methods for the Solution of Problems of Equilibrium and Vibrations. Bull. Am. Math. Soc. **49**, 1–23
10. Fiacco, A., McCormick, G. (1990): Nonlinear Programming. Sequential Unconstrained Minimization Techniques. SIAM Classic Appl. Math., SIAM Philadelphia, PA
11. Guler, O. (1991): On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**, 403–419
12. Hestenes, M.R. (1969): Multipliers and gradient methods. J. Optim. Theory Appl. **4**(5), 303–320
13. Hoffman, A. (1952): On approximate solution of system of linear inequalities. J. Res. Nat. Bureau Standards **49**, 263–265
14. Iusem, A., Svaiter, B., Teboulle, M. (1994): Entropy-like proximal methods in convex programming. Math. Oper. Res. **19**, 790–814
15. Kanzow, C. (1996): Global convergence properties of some iterative methods for linear complementarity problems. SIAM Optim. **6**, 326–341
16. Knopp, K. (1956): Infinite sequence and series. Dover Publication Inc., New York
17. Kort, B.W., Bertsekas, D.P. (1973): Multiplier methods for convex Programming. Proceedings 1973, IEEE Conference on Decision and Control, San Diego, California, pp. 428–432
18. Mangasarian, O. (1993): Mathematical programming in Neural Networks. ORSA J. Comput. **5**(4), 349–360
19. Mathematical Encyclopedia, v5, p. 610. Soviet Encyclopedia, Moscow, 1985
20. Martinet, B. (1970): Regularisation d'inequations variotionnelles par approximations succesive. Revue Francaise d'Aufomatique et Informatique Rechershe Operationelle **4**, 154–159
21. Melman, A., Polyak, R. (1996): The Newton Modified Barrier method for Q Problems. Ann. Oper. Res. **62**, 465–519
22. Moreau, J. (1965): Proximite et dualite 'daus un espace Hilbertien. Bull. Soc. Math. France **93**, 273–299
23. Nash, S., Polyak, R., Sofer, A. (1994): A numerical comparison of Barrier and Modified Barrier Method for Large-Scale Bound-Constrained Optimization. In: Hager, W., Hearn, D., Pardalos, P. (eds), Large Scale Optimization, state of the art, pp. 319–338. Kluwer Academic Publisher
24. Polyak, B. (1987): Introduction to Optimization. Software Inc., NY
25. Polyak, R. (1992): Modified Barrier Functions. Math. Program. **54**, 177–222
26. Polyak, R. (2001): Log-Sigmoid Multipliers Method in Constrained Optimization. Ann. Oper. Res. **101**, 427–460
27. Polyak, R., Teboulle, M. (1997): Nonlinear Rescaling and Proximal-like Methods in convex optimization. Math. Program. **76**, 265–284
28. Polyak, R., Griva, I., Sobieski, J. (1998): The Newton Log-Sigmoid method in Constrained Optimization, A Collection of Technical Papers. 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization 3, pp. 2193–2201
29. Powell, M.J.D. (1969): A method for nonlinear constraints in minimization problems. In: Fletcher (ed.), Optimization, pp. 283–298. Acad. Press, London
30. Powell, M.J.D. (1995): Some convergence properties of the Modified Log Barrier Method for Linear Programming. SIAM J. Optim. **5**(4), 695–739
31. Rockafellar, R.T. (1976): Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898
32. Rockafellar, R.T. (1976): Augmented Lagrangians and applications of the proximal points algorithm in convex programming. Math. Oper. Res. **1**, 97–116
33. Rockafellar, R.T. (1970): Convex analysis. Princeton University Press, Princeton NJ
34. Teboulle, M. (1992): Entropic proximal mappings with application to nonlinear programming. Math. Oper. Res. **17**, 670–690
35. Teboulle, M. (1997): Convergence of Proximal-like Algorithms. SIAM J. Optim. **7**, 1069–1083
36. Tikhonov, A., Arseniev, V. (1974): Methodes de resolution de problemes mal poses. MIR
37. Tseng, P., Bertsekas, D. (1993): On the convergence of the exponential multipliers method for convex programming. Math. Program. **60**, 1–19
38. Wright, S. (1997): Primal-Dual Interior Points methods. SIAM