Lagrangian Transformation and Interior Ellipsoid Methods in Convex Optimization

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

In memory of Ilya I. Dikin.

Ilya Dikin passed away on February 2008 at the age of 71.
One can only guess how the field of Optimization would look
if the importance of I. Dikin's work would had been recognized
not in the late 80s, but in late 60s when he introduced the
affine scaling method for LP calculation.

November 16, 2012

Abstract

The rediscovery of the affine scaling method in the late 80s was one of the turning points which led to a new chapter in Modern Optimization - the interior point methods (IPMs). Simultaneously and independently the theory of exterior point methods for convex optimization arose. The two seemingly unconnected fields turned out to be intrinsically connected.

The purpose of this paper is to show the connections between primal exterior and dual interior point methods.

Our main tool is the Lagrangian Transformation (LT) which for inequality constrained has the best features of the classical augmented Lagrangian. We show that the primal exterior LT method is equivalent to the dual interior ellipsoid method. Using the equivalence we prove convergence, estimate the convergence rate and establish the complexity bound for the interior ellipsoid method under minimum assumptions on the input data.

We show that application of the LT method with modified barrier (MBF) transformation for linear programming (LP) leads to Dikin's affine scaling method for the dual LP.

Keywords: Lagrangian Transformation; Interior Point Methods; Duality; Bregman Distance; Augmented Lagrangian; Interior Ellipsoid Method.

2000 Mathematics Subject Classifications: 90C25; 90C26

1 Introduction

The affine scaling method for solving linear and quadratic programming problems was introduced in 1967 by Ilya Dikin in his paper [7] published in Soviet Mathematics Doklady -the leading scientific journal of the former Soviet Union. Unfortunately for almost two decades I. Dikin's results were practically unknown not only to the Western but to the Eastern Optimization community as well.

The method was independently rediscovered in 1986 by E.Barnes [2] and by R.Vanderbei, M.Meketon and B.Freedman [37] as a version of N.Karmarkar's method [15],which together with P.Gill et. al [10], as well as C. Gonzaga and J.Renegar's results (see [9],[31]) marked the starting point of a new chapter in Modern Optimization-interior point methods (see a nice survey by F.Potra and S.Wright [29]).

The remarkable Self-Concordance (SC) theory was developed by Yu. Nesterov and A.Nemirovsky soon after. The SC theory allowed understanding the complexity of the IPMs for well structured convex optimization problems from a unique and general point of view (see [20],[21]).

The exterior point methods (EPMs) (see [1],[3], [23]-[25], [35],[36]) attracted much less attention over the years mainly due to the lack of polynomial complexity.

On the other hand the EPMs were used in instances where SC theory can't be applied and, what is more important, some of them, in particular the nonlinear rescaling method with truncated MBF transformation turned out to be very efficient for solving difficult large scale NLP (see for example [11], [16], [25], [27] and references therein).

The EPMs along with an exterior primal generate an interior dual sequence as well. The connection between EPMs and interior proximal point methods for the dual problem were well known for quite some time (see [24]-[27], [35], [36]). Practically, however, exterior and interior point methods were disconnected for all these years.

Only very recently, L. Matioli and C. Gonzaga (see [17]) found that the LT scheme with truncated MBF transformation leads to the so-called exterior M^2BF method, which is equivalent to the interior ellipsoid method with Dikin's ellipsoids for the dual problem.

In this paper we systematically study the intimate connections between the primal EPMs and dual IPMs generated by the Lagrangian Transformation scheme (see [17], [26]).

Let us mention some of the results obtained.

First, we show that the LT multiplier method is equivalent to an interior proximal point (interior prox) method with Bregman type distance, which, in turn, is equivalent to an interior quadratic prox (IQP) for the dual problem in the from step to step rescaled dual space.

Second, we show that the IQP is equivalent to an interior ellipsoid method for solving the dual problem. The equivalence is used to prove the convergence in value of the dual sequence to the dual solution for any Bregman type distance function with well defined kernel. LT with MBF transformation leads to dual proximal point method with Bregman distance but the MBF kernel is not well defined. The MBF kernel, however, is a self -concordant function and the correspondent dual prox-method is an interior ellipsoid method with Dikin's ellipses.

Third, the rescaling technique is used to establish the convergence rate of the IEMs and to estimate the the upper bound for the number steps required for finding an $\epsilon-$approximation to the optimal solution.

All these results are true for convex optimization problems including those for which the SC theory can't be applied.

Finally, it was shown that the LT method with truncated MBF transformation for LP leads to I. Dikin's affine scaling type method for the dual LP problem.

The paper is organized as follows. The problem formulation and basic assumptions are introduced in the following Section. The LT is introduced in Section 3. The LT method and its equivalence to the IEM for the dual problem are considered in Section 4. The convergence analysis of the interior ellipsoid method is given in Section 5. In Section 6 the LT method with MBF transformation is applied for LP. We conclude the paper with some remarks concerning future research.

2 Problem Formulation and Basic Assumptions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and all $c_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, q$ be concave and continuously differentiable. The convex optimization problem consists of finding

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

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

We assume:

$A1$. The optimal set $X^* = \{x \in \Omega : f(x) = f(x^*)\} \neq \emptyset$ is bounded.

$A2$. The Slater condition holds:

$$\exists \quad x^0 \in \Omega : c_i(x^0) > 0, i = 1, \ldots, q.$$

Let's consider the Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^q_+ \to \mathbb{R}$ associated with the primal problem (P).

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

It follows from A2 that Karush–Kuhn-Tucker's (KKT's) conditions hold:

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

b) $\lambda_i^* c_i(x^*) = 0, c_i(x^*) \geq 0, \lambda_i^* \geq 0, \quad i = 1, \ldots, q,$

Let's consider the dual function $d : \mathbb{R}^q_+ \to \mathbb{R}$

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

The dual problem consists of finding

$$(D) \qquad d(\lambda^*) = \max\{d(\lambda) | \lambda \in \mathbb{R}^q_+\}.$$

Again due to the assumption A2 the dual optimal set

$$L^* = \text{Argmax}\{d(\lambda) | \lambda \in \mathbb{R}^q_+\}$$

is bounded.

Keeping in mind A1 we can assume boundedness of Ω, because if it is not so by adding one extra constraint $f(x) \leq N$ we obtain a bounded feasible set. For large enough N the extra constraint does not effect the optimal solution.

3 Lagrangian Transformation

Let's consider a class Ψ of twice continuous differentiable functions $\psi : R \to R$ with the following properties

1. $\psi(0) = 0$
2. a) $\psi'(t) > 0$, b) $\psi'(0) = 1$, $\psi'(t) \leq at^{-1}, a > 0, t > 0$
3. $-m^{-1} \leq \psi''(t) < 0, \forall t \in (-\infty, \infty)$
4. $\psi''(t) \leq -M^{-1}, \forall t \in (-\infty, 0), 0 < m < M < \infty.$

We start with some well known transformations, which unfortunately do not belong to Ψ.

- Exponential transformation [3]: $\hat{\psi}_1(t) = 1 - e^{-t}$
- Logarithmic MBF transformation [23]: $\hat{\psi}_2(t) = \ln(t + 1)$
- Hyperbolic MBF transformation [23]: $\hat{\psi}_3(t) = t/(t + 1)$
- Log–Sigmoid transformation [25]: $\hat{\psi}_4(t) = 2(\ln 2 + t - \ln(1 + e^t))$
- Modified Chen-Harker-Kanzow-Smale (CMKS) transformation [25]: $\hat{\psi}_5(t) = t - \sqrt{t^2 + 4\eta} + 2\sqrt{\eta}, \quad \eta > 0$

The transformations $\hat{\psi}_1 - \hat{\psi}_3$ do not satisfy 3. ($m = 0$), while transformations $\hat{\psi}_4$ and $\hat{\psi}_5$ do not satisfy 4. ($M = \infty$). A slight modification of $\hat{\psi}_i, i = 1, \ldots, 5$, however, leads to $\psi_i \in \Psi$.

Let $-1 < \tau < 0$, the modified transformations $\psi_i : \mathbb{R} \to \mathbb{R}$ are defined as follows

$$\psi_i(t) := \begin{cases} \hat{\psi}_i(t), \infty > t \geq \tau \\ q_i(t), -\infty < t \leq \tau \end{cases} \tag{1}$$

where $q_i(t) = a_i t^2 + b_i t + c_i$ and $a_i = 0.5\hat{\psi}_i''(\tau)$, $b_i = \hat{\psi}_i'(\tau) - \tau\hat{\psi}''(\tau)$,
$c_i = \hat{\psi}_i'(\tau) - \tau\hat{\psi}_i'(\tau) + 0.5\tau^2\hat{\psi}_i''(\tau)$. It is easy to check that all $\psi_i \in \Psi$, *i.e.* for all ψ_i the properties 1.-4. hold.

For a given $\psi_i \in \Psi$, let's consider the conjugate

$$\psi_i^*(s) := \begin{cases} \hat{\psi}_i^*(s), & s \leq \hat{\psi}_i'(\tau) \\ q_i^*(s) = (4a_i)^{-1}(s - b_i)^2 - c_i, & s \geq \hat{\psi}_i'(\tau), i = 1, \ldots, 5, \end{cases} \tag{2}$$

where $\hat{\psi^*}_i(s) = \inf_t\{st - \hat{\psi}_i(t)\}$ is the Fenchel conjugate of $\hat{\psi}_i$. With the class of transformations Ψ we associate the class of kernels $\varphi \in \Phi = \{\varphi = -\psi^* : \psi \in \Psi\}$

Using properties 2. and 4. one can find $0 < \theta_i < 1$ that

$$\hat{\psi}_i'(\tau) - \hat{\psi}_i'(0) = -\hat{\psi}_i''(\tau\theta_i)(-\tau) \geq -\tau M^{-1}, i = 1, \ldots, 5$$

or

$$\hat{\psi}_i'(\tau) \geq 1 - \tau M^{-1} = 1 + |\tau| M^{-1}. \tag{3}$$

Therefore using the definition (2) for any $0 < s \leq 1 + |\tau| M^{-1}$ we have

$$\varphi_i(s) = \hat{\varphi}_i(s) = -\hat{\psi}_i^*(s) = \inf_t\{st - \hat{\psi}_i(t)\}.$$

The corresponding kernels

- Exponential kernel $\hat{\varphi}_1(s) = s \ln s - s + 1, \hat{\varphi}_1(0) = 1$;
- MBF kernel $\hat{\varphi}_2(s) = -\ln s + s - 1$;
- Hyperbolic MBF kernel $\hat{\varphi}_3(s) = -2\sqrt{s} + s + 1$;
- Fermi-Dirac kernel $\hat{\varphi}_4(s) = (2 - s)\ln(2 - s) + s \ln s, \hat{\varphi}_4(0) = 2\ln 2$;
- CMKS kernel $\hat{\varphi}_5(s) = -2\sqrt{\eta}(\sqrt{(2 - s)s} - 1)$

are infinitely differentiable on $(0, 1 + |\tau| M^{-1})$.

Definition 1. *A kernel $\varphi \in \Phi$ is well defined if $\varphi(0) < \infty$.*

So $\varphi_1, \varphi_3 - \varphi_5$ are well defined. The MBF kernel $\hat{\varphi}_2(s)$ is not well defined, but it is the only self-concordant function on $\mathbb{R}_+$ among φ_1-φ_5, while the MBF transformation $\psi_2(t) = \ln(t+1)$ is the only self-concordant barrier among $\psi_1 - \psi_5$ (see [20]-[21]).

The properties of kernels $\varphi \in \Phi$ are induced by the properties 1.–4. of transformations $\psi \in \Psi$and are given by the following Lemma.

Lemma 1. *[27]. The kernels $\varphi \in \Phi$ are strongly convex on $\mathbb{R}^q_+$, twice continuously differentiable and possess the following properties*

1) $\varphi(s) \geq 0,\ \forall s \in (0, \infty)$ *and* $\min\limits_{s \geq 0} \varphi(s) = \varphi(1) = 0$;

2) *a)* $\lim\limits_{s \to 0^+} \varphi'(s) = -\infty$, *b)* $\varphi'(s)$ *is monotone increasing and*

   c) $\varphi'(1) = 0$;

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

Let $Q \subset \mathbb{R}^q$ be an open convex set, $\hat{Q}$ the closure of Q and $\varphi : Q \to \mathbb{R}$ be a strictly convex and continuously differentiable function on Q; then the Bregman distance $\mathbb{B}_\varphi : Q \times Q \to R_+$ induced by φ is given by the following formula [4] (see also [5],[6],[14])

$$\mathbb{B}_\varphi(x, y) = \varphi(x) - \varphi(y) - (\nabla \varphi(y), x - y).$$

Let $\varphi \in \Phi$, the function $B_\varphi : \mathbb{R}^q_+ \times \mathbb{R}^q_{++} \to \mathbb{R}_+$, which is defined by the formula

$$B_\varphi(u, v) = \sum_{i=1}^q \varphi(u_i/v_i),$$

we call the Bregman type distance induced by the kernel φ. Due to $\varphi(1) = \varphi'(1) = 0$ for any $\varphi \in \Phi$, we have

$$\varphi(t) = \varphi(t) - \varphi(1) - \varphi'(1)(t - 1), \tag{4}$$

which means that $\varphi(t) : \mathbb{R}_{++} \to \mathbb{R}_{++}$ is Bregman distance between $t > 0$ and 1. By taking $t_i = \frac{u_i}{v_i}$ from (4) we obtain

$$B_\varphi(u, v) = B_\varphi(u, v) - B_\varphi(v, v) - (\nabla_u B_\varphi(v, v), u - v),$$

which justifies the definition of the Bregman type distance.

Lemma 2. *The Bregman type distance satisfies the following properties:*

1) $B_\varphi(u, v) \geq 0, \forall u \in \mathbb{R}^q_+, v \in \mathbb{R}^q_{++}$;

2) $B_\varphi(u, v) \geq \frac{1}{2} m \sum_{i=1}^q (\frac{u_i}{v_i} - 1)^2, \forall u \in \mathbb{R}^q_+, v \in \mathbb{R}^q_{++}$;

3) $B_\varphi(u, v) = 0 \Leftrightarrow u = v,\ \forall v \in \mathbb{R}^q_{++}$;

4) *for any fixed* $v \in \mathbb{R}^q_{++}$ *the gradient* $\nabla_u B_\varphi(u, v)$ *is a barrier function in* $u \in \mathbb{R}^q_{++}$, *i.e.*

$$\lim_{u_i \to 0_+} \frac{\partial}{\partial u_i} B_\varphi(u, v) = -\infty, i = 1, \ldots, q;$$

5) $B_\varphi(u, v) \leq \frac{1}{2} M \sum_{i=1}^q (\frac{u_i}{v_i} - 1)^2, \forall u \in \mathbb{R}^q_+, u \geq v > 0$.

Proof. It follows from 1) of Lemma 1 that $\varphi(s) \geq 0,\ \forall s > 0$, therefore

1) $B_\varphi(u,v) \geq 0, \forall u \in \mathbb{R}^q_+, v \in \mathbb{R}^q_{++}$.

2) It follows from $\varphi(1) = \varphi'(1) = 0$ that

$$
\begin{aligned}
B_\varphi(u,v) &= \sum_{i=1}^q \varphi\left(\frac{u_i}{v_i}\right) = \sum_{i=1}^q \varphi\left(\frac{v_i}{v_i}\right) \\
&+ \sum_{i=1}^q \varphi'\left(\frac{v_i}{v_i}\right)\left(\frac{u_i}{v_i} - 1\right) + \frac{1}{2}\sum_{i=1}^q \varphi''\left(1 + \theta_i\left(\frac{u_i}{v_i} - 1\right)\right)(\frac{u_i}{v_i} - 1)^2 \\
&= \frac{1}{2}\sum_{i=1}^q \varphi''\left(1 + \theta_i\left(\frac{u_i}{v_i} - 1\right)\right)\left(\frac{u_i}{v_i} - 1\right)^2,
\end{aligned}
$$

where $0 < \theta_i < 1$. Using 3a) from Lemma 1 we obtain

$$
B_\varphi(u,v) \geq \frac{1}{2} m \sum_{i=1}^q \left(\frac{u_i}{v_i} - 1\right)^2. \tag{5}
$$

3) It follows from (5) and 1) of Lemma 1 that

$$
B_\varphi(u,v) = 0 \Leftrightarrow u = v,\ \forall v \in \mathbb{R}^q_{++}.
$$

4) It follows from 2a) of Lemma 1 that $\lim_{s \to 0^+} \varphi'(s) = -\infty$, therefore for any $v \in \mathbb{R}^q_{++}$ we have

$$
\lim_{u_i \to 0_+} \frac{\partial}{\partial u_i} B_\varphi(u,v) = -\infty, i = 1, \ldots, q.
$$

5) Using 3b) from Lemma 1 we obtain

$$
B_\varphi(u,v) \leq \frac{1}{2} M \sum_{i=1}^q \left(\frac{u_i}{v_i} - 1\right)^2, \forall u \geq v > 0. \tag{6}
$$

$\square$

For a given $\psi \in \Psi$ the Lagrangian Transformation $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^q_+ \times \mathbb{R}_{++} \to \mathbb{R}$ with a fixed scaling parameter $k > 0$ is defined by the formula

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

It follows from property 3. of ψ and definition (7) that $\mathcal{L}(x, \lambda, k)$ is convex in x for any fixed $\lambda \in \mathbb{R}^q_{++}$ and any $k > 0$. It follows from properties 1. and 2. of $\psi \in \Psi$ that for any KKT pair (x^*, λ^*) and any $k > 0$ we have

1^0 $\mathcal{L}(x^*, \lambda^*, k) = f(x^*)$;

2^0 $\nabla_x \mathcal{L}(x^*, \lambda^*, k) = \nabla_x L(x^*, \lambda^*)$;

3^0 $\nabla^2\mathcal{L}(x^*,\lambda^*,k)=\nabla^2_{xx}L(x^*,\lambda^*)-k\psi''(0)\nabla c(x^*)^T\Lambda^{*2}\nabla c(x^*)$,

where $\nabla c(x)=J(c(x))$– the Jacobian of vector–function $c(x)^T=(c_1(x),\dots,c_q(x))$ and $\Lambda^*=\mathtt{diag}(\lambda_i^*)_{i=1}^q$.

The properties 1^0–3^0 are similar to the Augmented Lagrangian (AL) properties at the primal–dual solution (x^*,λ^*) for equality constrained optimization (see [3],[13],[22],[30],[32]-[34]).

The fundamental difference, however, between equality constrained optimization and problem (P) is that the active constraints set in (P) is unknown *a priori*.

One of the purposes of the paper is to show that the LT approach leads to an Interior Quadratic Prox for the dual problem, which retains the basic properties typical of the classical Quadratic Prox in spite of the fact that the active constraints set of (P) is unknown *a priori*.

4 LT Multipliers Method

Let $\psi\in\Psi$, $\lambda^0\in\mathbb{R}^q_{++}$ and $k>0$ are given. The LT method generates a primal–dual sequence $\{x^s,\lambda^s\}_{s=1}^\infty$ by formulas

$$x^{s+1}:\nabla_x\mathcal{L}(x^{s+1},\lambda^s,k)=0 \tag{8}$$

$$\lambda_i^{s+1}=\lambda_i^s\psi'(k\lambda_i^sc_i(x^{s+1})),i=1,\dots,q. \tag{9}$$

It follows from assumptions A1, convexity of f, concavity of $c_i, i=1,\dots,q$ and property 4. of $\psi\in\Psi$ that for any $\lambda^s\in\mathbb{R}^q_{++}$ and $k>0$ the level set $\{x:\mathcal{L}(x,\lambda^s,k)\le\mathcal{L}(x^s,\lambda^s,k)\}$ is bounded for all $s\ge1$ (see for example [1]). Therefore the minimizer x^s exists for all $s\ge1$. It follows from property 2. a) of $\psi\in\Psi$ and (9) that $\lambda^s\in\mathbb{R}^q_{++}\Rightarrow\lambda^{s+1}\in\mathbb{R}^q_{++}$. Therefore the LT method (8)– (9) is well defined.

Also from (8) we have

$$\nabla_x\mathcal{L}(x^{s+1},\lambda^s,k)=\nabla f(x^{s+1})-\sum_{i=1}^q\lambda_i^s\psi'(k\lambda_i^sc_i(x^{s+1}))\nabla c_i(x^{s+1}))=0. \tag{10}$$

Using (9) and (10) we obtain

$$\nabla_x\mathcal{L}(x^{s+1},\lambda^s,k)=\nabla f(x^{s+1})-\sum_{i=1}^q\lambda_i^{s+1}\nabla c_i(x^{s+1})=\nabla_xL(x^{s+1},\lambda^{s+1})=0,$$

i.e

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

and

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

Theorem 1. *1) The LT method (8)– (9) is equivalent to the following Prox method*

$$\lambda^{s+1}=\arg\max\{d(\lambda)-k^{-1}B_\varphi(\lambda,\lambda^s)|\lambda\in\mathbb{R}^q_{++}\}, \tag{11}$$

where $B_\varphi(u,v)$ is Bregman type interior distance function.

2) There exists a positive sequence $\{\epsilon_s\}_{s=0}^{\infty} : \lim_{s\to\infty} \epsilon_s = 0$, so that for $\forall s \geq s_0$ the following bound holds

$$\max_{1\leq i\leq q} |\frac{\lambda_i^{s+1}}{\lambda_i^s} - 1| \leq (2m^{-1})^{\frac{1}{2}}\epsilon_s. \tag{12}$$

.

Proof. 1) It follows from (9) that

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

In view of property 3. for any $\psi \in \Psi$ there exists an inverse function ψ'^{-1}, therefore

$$c_i(x^{s+1}) = k^{-1}(\lambda_i^s)^{-1}\psi'^{-1}(\lambda_i^{s+1}/\lambda_i^s), i = 1, \ldots, q. \tag{13}$$

Using the well known identity $\psi'^{-1} = \psi^{*\prime}$ we can rewrite (13) as follows

$$-c_i(x^{s+1}) = -k^{-1}(\lambda_i^s)^{-1}\psi^{*\prime}(\lambda_i^{s+1}/\lambda_i^s),\ \ i = 1, \ldots, q. \tag{14}$$

By introducing $\varphi = -\psi^*$ and keeping in mind $-c(x^{s+1}) = g^{s+1} \in \partial d(\lambda^{s+1})$, where $c(x)^T = (c_1(x), \ldots, c_q(x))$ and $\partial d(\lambda^{s+1})$ is the subdifferential of $d(\lambda)$ at $\lambda = \lambda^{s+1}$, we can rewrite (14) as follows

$$g^{s+1} - k^{-1}\sum_{i=1}^{q}(\lambda_i^s)^{-1}\varphi'(\lambda_i^{s+1}/\lambda_i^s)e_i = 0,$$

where $e_i = (0, \ldots, 1, \ldots, 0)$.

The last system is the optimality criteria for λ^{s+1} in (11), where $B_\varphi(u,v) = \sum_{i=1}^{q}\varphi(u_i/v_i)$ is the Bregman type interior distance function, therefore the LT method (8)-(9) is equivalent to the Prox method(11).

2) It follows from (11) and 3) of Lemma 2 that

$$d(\lambda^{s+1}) - k^{-1}B_\varphi(\lambda^{s+1}, \lambda^s) \geq d(\lambda^s) - k^{-1}B_\varphi(\lambda^s, \lambda^s) = d(\lambda^s)$$

Therefore

$$d(\lambda^{s+1}) \geq d(\lambda^s) + k^{-1}B_\varphi(\lambda^{s+1}, \lambda^s), \forall s \geq 0$$

Summing up the last inequality from $s = 0$ to $s = N$, we obtain,

$$d(\lambda^*) - d(\lambda^0) \geq d(\lambda^{N+1}) - d(\lambda^0) \geq k^{-1}\sum_{s=0}^{N} B_\varphi(\lambda^{s+1}, \lambda^s) \tag{15}$$

i.e.

$$\lim_{s\to\infty} B(\lambda^{s+1}, \lambda^s) = \lim_{s\to\infty}\sum_{i=1}^{q}\varphi(\frac{\lambda_i^{s+1}}{\lambda_i^s}) = 0.$$

Therefore in view of 2) of Lemma 2 there exists a sequence $\{\epsilon_s > 0\}_{s=0}^{\infty} : \lim_{s\to\infty} \epsilon_s = 0$ that

$$\epsilon_s^2 = B_{\varphi}(\lambda^{s+1}, \lambda^s) \geq \frac{1}{2} m \sum_{i=1}^{q} \left(\frac{\lambda_i^{s+1}}{\lambda_i^s} - 1 \right)^2 ,$$

i.e. the bound (12) holds.

□

Corollary It follows from (3) and (12) that there is $s_0 > 0$ so that for any $s \geq s_0$, only kernels $\hat{\varphi}_i$, which correspond to the original transformations $\hat{\psi}_i$ $i = 1, \ldots, 5$ are used in the Prox method (11), *i.e.* the quadratic branch of transformations in (1) is irrelevant from some point on. It is also true for any $s > 1$ if $k_0 > 0$ is large enough.

On the other hand, for all original transformations $\hat{\psi}_1 - \hat{\psi}_5$ either $m = 0$ or $M = \infty$, which make the existence of x^{s+1} in (8) as well as Corollary problematic. Also it makes statement 3) of Lemma 1 as well as statements 2) and 5) of Lemma 2 trivial and useless.

Remark 1. *The LT method (8)– (9) with the truncated MBF transformation $\psi_2(t)$ and its dual equivalent (11) were considered by L. Matioli and C. Conzaga in [17]. Their basic result (see Theorem 3.2 and Theorem 4.6 in [17]) is convergence of the dual sequence in value, i.e.* $\lim_{s\to\infty} d(\lambda^s) = d(\lambda^*)$. *They also show that any cluster point of the dual sequence $\{\lambda^s\}_{s=0}^{\infty}$is a dual solution.*

It follows from the Corollary that for $\forall s \geq s_0$ only the MBF kernel $\varphi_2(\tau) = -\ln \tau + \tau - 1$, is used in (11). Therefore the following Bregman distance

$$\mathbb{B}_{\varphi_2}(u, v) = \sum_{i=1}^{q} \left(-\ln \frac{u_i}{v_i} + \frac{u_i}{v_i} - 1 \right) \tag{16}$$

controls the computational process in (11) from some point on. The Bregman distance $\mathbb{B}_{\varphi_2}(u, v)$is induced by the standard log-barrier function $\varphi(\tau) = -\ln \tau$, *i.e.*

$$\begin{aligned} \mathbb{B}_{\varphi_2}(u, v) &= \sum_{i=1}^{q} \left(-\ln u_i + \ln v_i + \frac{1}{v_i} (u_i - v_i) \right) \\ &= \sum_{i=1}^{q} \left(-\ln \frac{u_i}{v_i} + \frac{u_i}{v_i} - 1 \right). \end{aligned}$$

For a Prox method (11) with Bregman distance one can expect results similar to those in [6], which are stronger than Theorem 3.2 in [17].

In particular, it follows from Theorem 3.4 in [6] that the entire dual sequence $\{\lambda^s\}_{s=0}^{\infty}$ converges to λ^* and the following bound

$$d(\lambda^*) - d(\lambda^s) \leq (ks)^{-1} D(\lambda^*, \lambda^0) \tag{17}$$

holds, where $D(u, v)$ is the Bregman distance.

The MBF kernel $\varphi_2(\tau) = -\ln \tau + \tau - 1$, however, is not well defined (see Definition 1), therefore, generally speaking, the results in [6] cannot be applied to the Prox (11) with Bregman distance (16).

In particular $\mathbb{B}_{\varphi_2}(\lambda^*, \lambda^0) = \infty$, which makes the bound (17) trivial and useless.

Remark 2. *The MBF kernel $\varphi_2(\tau)$ is a self-concordant function on R_{++}, hence the Bregman distance $\mathbb{B}_{\varphi_2}(u,v)$ is a self-concordant function in $u \in \mathbb{R}^q_{++}$ under any fixed $v \in \mathbb{R}^q_{++}$, therefore for LP and QP calculations Newton's method for solving(11) can be very efficient (see [21], p 191,192).*

For convex optimization problems, which are not well structured (see [21] Ch 4), *i.e.* the constraints or/and the objective function epigraph can't be equipped with a self-concordant barrier the SC theory, generally speaking, does not work and polynomial complexity of IPM becomes problematic.

The results of the following theorem are independent of the structure of the convex optimization problem. It establishes the equivalence of the Prox method (11) to an Interior Quadratic Prox (IQP) in the step by step rescaled dual space, which in turn, is equivalent to an Interior Ellipsoid Method (IEM). The equivalence will be used later for convergence analysis of the Prox method (11).

In the case of the MBF transformation the corresponding IEM is based on the self-concordant MBF kernel $\hat{\varphi}_2(\tau)$, therefore the corresponding interior ellipsoids are Dikin's ellipsoids (see [21] p. 182 and [17]).

Theorem 2. *1) For a given $\varphi \in \Phi$ there exists a diagonal matrix $H_\varphi = \mathrm{diag}(h^i_\varphi)^q_{i=1}$ with $h^i_\varphi > 0, i = 1, \dots, q$, so that: $B_\varphi(u,v) = \frac{1}{2}\|u-v\|^2_{H_\varphi}$, where $\|w\|^2_{H_\varphi} = w^T H_\varphi w$;*

2) The Interior Prox method (11) is equivalent to an IQP method in the step by step rescaled dual space, i.e.

$$\lambda^{s+1} = \arg\max\{d(\lambda) - \frac{1}{2k}\|\lambda - \lambda^s\|^2_{H^s_\varphi} | \lambda \in \mathbb{R}^q_+\} \tag{18}$$

where $H^s_\varphi = \mathrm{diag}(h^{i,s}_\varphi) = \mathrm{diag}(2\varphi''(1+\theta^s_i(\lambda^{s+1}_i/\lambda^s_i - 1))(\lambda^s_i)^{-2})$ and $0 < \theta^s_i < 1$;

3) The IQP is equivalent to an IEM for the dual problem;

4) There exists a converging to zero sequence $\{r_s > 0\}^\infty_{s=0}$ and step $s_0 > 0$ such that for $\forall s \geq s_0$ the LT method (8)– (9) with the truncated MBF transformation $\psi_2(t)$ is equivalent to the following IEM for the dual problem

$$\lambda^{s+1} = \arg\max\{d(\lambda) | \lambda \in E(\lambda^s, r_s)\}, \tag{19}$$

where $H_s = \mathrm{diag}(\lambda^s_i)^q_{i=1}$ and $E(\lambda^s, r_s) = \{\lambda : (\lambda - \lambda^s)^T H_s^{-2}(\lambda - \lambda^s) \leq r_s^2\}$ is Dikin's ellipsoid associated with the standard log–barrier function $F(\lambda) = -\sum_{i=1}^q \ln \lambda_i$, for the dual feasible set $\mathbb{R}^q_+$.

Proof. 1) It follows from $\varphi(1) = \varphi'(1) = 0$ that

$$B_\varphi(u,v) = \frac{1}{2}\sum_{i=1}^q \varphi''(1 + \theta_i(\frac{u_i}{v_i} - 1))(\frac{u_i}{v_i} - 1)^2, \tag{20}$$

where $0 < \theta_i < 1$, $i = 1, \dots, q$.

Due to 3a) of Lemma 1 we have $\varphi''(1+\theta_i(\frac{u_i}{v_i}-1)) \geq m > 0$ and due to property 2a) of $\psi \in \Psi$ we have $v \in \mathbb{R}^q_{++}$, therefore $h^i_\varphi = 2\varphi''(1+\theta_i(\frac{u_i}{v_i}-1))v_i^{-2} > 0, i = 1, \ldots, q$.

We consider the diagonal matrix $H_\varphi = \mathrm{diag}(h^i_\varphi)^q_{i=1}$, then from (20) we have

$$B_\varphi(u,v) = \frac{1}{2}\|u-v\|^2_{H_\varphi}. \tag{21}$$

2) By taking $u = \lambda$, $v = \lambda^s$ and $H_\varphi = H^s_\varphi$ from (11) and (21) we obtain (18), i.e. (11) is equivalent to an Interior Quadratic Prox method (18) for the dual problem.

3) Let's consider the optimality criteria for the problem (18). Keeping in mind $\lambda^{s+1} \in \mathbb{R}^q_{++}$ we conclude that λ^{s+1} is an unconstrained maximizer in (18). Therefore one can find $g^{s+1} \in \partial d(\lambda^{s+1})$ so that
$$g^{s+1} - k^{-1}H^s_\varphi(\lambda^{s+1} - \lambda^s) = 0. \tag{22}$$
Let $r_s = \|\lambda^{s+1} - \lambda^s\|_{H^s_\varphi}$, we consider an ellipsoid $E(\lambda^s, r_s) = \{\lambda : (\lambda - \lambda^s)^T H^s_\varphi(\lambda - \lambda^s) \leq r^2_s\}$ with center $\lambda^s \in \mathbb{R}^q_{++}$ and radius r_s. It follows from 4) of Lemma 2 that $E(\lambda^s, r_s)$ is an interior ellipsoid in $\mathbb{R}^q_{++}$, *i.e.* $E(\lambda^s, r_s) \subset \mathbb{R}^q_{++}$.

Also $\lambda^{s+1} \in \partial E(\lambda^s, r_s) = \{\lambda : (\lambda - \lambda^s)^T H^s_\varphi(\lambda - \lambda^s) = r^2_s\}$, therefore (22) is the optimality criterion for the following optimization problem

$$d(\lambda^{s+1}) = \max\{d(\lambda) | \lambda \in E(\lambda^s, r_s)\} \tag{23}$$

and $(2k)^{-1}$ is the optimal Lagrange multiplier for the only constraint in (23).

In other words the Interior Prox method (11) is equivalent to an Interior Ellipsoid Method (23).

4) Let's consider the LT method (8)-(9) with MBF transformation $\psi_2(t) = ln(t+1)$. The correspondent problem (11) is a proximal point method with Bregman distance

$$B_{\varphi_2}(\lambda, \lambda^s) = \sum_{i=1}^q (-ln\frac{\lambda_i}{\lambda^s_i} + \frac{\lambda_i}{\lambda^s_i} - 1).$$

We have
$$\nabla^2_{\lambda\lambda} B(\lambda, \lambda^s)|_{\lambda=\lambda^s} = H^{-2}_s. \tag{24}$$
In view of $B_{\varphi_2}(\lambda^s, \lambda^s) = 0$ and $\nabla_\lambda B_{\varphi_2}(\lambda^s, \lambda^s) = 0^q$, we obtain

$$B_{\varphi_2}(\lambda, \lambda^s) = \frac{1}{2}(\lambda - \lambda^s)^T H^{-2}_s(\lambda - \lambda^s) + o(\|\lambda - \lambda^s\|^2) = Q_{\varphi_2}(\lambda, \lambda^s) + o(\|\lambda - \lambda^s\|^2).$$

It follows from (12) that for $s_0 > 0$ large enough and any $s \geq s_0$ one can ignore the last term, i.e. we can replace $B_{\varphi_2}(\lambda, \lambda^2)$ by $Q_{\varphi_2}(\lambda, \lambda^s)$ in (11).Then the correspondent optimality criteria (14) can be rewritten as follows

$$g^{s+1} - k^{-1}H^{-2}_s(\lambda^{s+1} - \lambda^s) = 0.$$

Therefore (see item 3))

$$d(\lambda^{s+1}) = \max\{d(\lambda)|\lambda \in E_{\varphi_2}(\lambda^s, r_s)\},$$

where $r_s^2 = Q_{\varphi_2}(\lambda^{s+1}, \lambda^s)$ and

$$E_{\varphi_2}(\lambda^s, r_s) = \{\lambda : (\lambda - \lambda^s) H_s^{-2} (\lambda - \lambda^s) = r_s^2\}$$

is Dikin's ellipsoid. It follows from the Corollary that for $k > 0$ large enough item 4) holds true for any $s \geq 1$.

$\square$

The item 4) of Theorem 2 is based on (24), which is not true for any Bregman type distance $B_{\varphi_i}(\lambda, \lambda^s), i = 1, 3, 4, 5$.

Remark 3. *The second and third derivatives of MBF kernel $\varphi_2(t) = -lnt + t - 1$ are identical to corespondent derivatives of the classical log-barrier $\varphi(t) = -lnt$, therefore $\varphi_2(t)$ is a self-concordant function (see definition 4.1.1. in [21]). Unfortunately $\varphi_2(t)$ is not a self-concordant barrier (see definition 4.2.2 in [21]).*

In spite of similarities between IEM (23) with Dikin's ellipsoid and classical IPMs, the rate of convergence of Prox Method (11) with Bregman distance even for well defined kernels is only arithmetic (see [6]).

In order to get the polynomial complexity typical for IPM, one has to replace the fixed scaling parameter $k > 0$ by a special monotone increasing sequence $\{k_s > 0\}_{s=0}^{\infty}$ and to control the size of Dikin's ellipsoids properly.

If epi $d(\lambda)$ can't be equipped with a self-concordant barrier then the complexity analysis requires means other than SC theory.

Let's consider the LT method (8)– (9) with a scaling parameter update from step to step for a general convex optimization problem.

Let $\{k_s\}_{s=0}^{\infty}$ be a positive, non-decreasing sequence, $k_0 > 0$ sufficiently large, $\lambda^0 := e = (1, \ldots, 1) \in \mathbb{R}^q_{++}$ and $\psi \in \Psi$ is such that the corresponding kernel $\varphi = -\psi^* \in \Phi$ is well defined.

The LT method with scaling parameter update generates the primal-dual sequence $\{x^s, \lambda^s\}_{s=0}^{\infty}$ by formulas

$$x^{s+1} : \nabla_x \mathcal{L}(x^{s+1}, \lambda^s, k_s) = 0 \tag{25}$$

$$\lambda_i^{s+1} = \lambda_i^s \psi'(k_s \lambda_i^s c_i(x^{s+1})), \;\; i = 1 \ldots q. \tag{26}$$

It follows from Theorem 1 that the LT method (25)– (26) is equivalent to the following Interior Prox method

$$\lambda^{s+1} = \text{argmax}\{d(\lambda) - k_s^{-1} B_{\varphi}(\lambda, \lambda^s)|\lambda \in \mathbb{R}^q_+\}, \tag{27}$$

where $B_{\varphi}(\lambda, \lambda^s)$ is the Bregman type distance. Using Theorem 2, we can rewrite the Prox method (27) as the following IQP

$$\lambda^{s+1} = \text{argmax}\{d(\lambda) - (2k_s)^{-1} \|\lambda - \lambda^s\|^2_{H^s_{\varphi}} |\lambda \in \mathbb{R}^q_+\}. \tag{28}$$

5 Convergence Analysis

In this section we establish convergence properties of the Interior Quadratic Prox method (28). The equivalence of (25)-(26) and IQP (28) is the key component of our analysis.

By introducing $\Lambda = (H^s_\varphi)^{\frac{1}{2}}\lambda$ one can rewrite the IQP (28) as the following Quadratic Prox in the rescaled dual space

$$\Lambda^{s+1} = \arg\max\{\mathcal{D}_s(\Lambda) - k_s^{-1}\|\Lambda - \Lambda^s\|^2 | \Lambda \in \mathbb{R}^q\},$$

where $\mathcal{D}_s(\Lambda) = d((H^s_\varphi)^{-1/2}\Lambda)$.

Note that the Euclidean Distance $E(U,V) = \|U - V\|^2$ is a Bregman distance induced by the kernel $\varphi(w) = \frac{1}{2}\|w\|^2$ because $\nabla\varphi(w) = w$ and

$$E(U,V) = \mathbb{B}_\varphi(U,V) = \frac{1}{2}U^2 - \frac{1}{2}V^2 - (U-V,V) = \frac{1}{2}\|U-V\|^2.$$

The "three points identity" is the basic element for convergence analysis of the classical Quadratic Prox (see [6], [12] and references therein). The basic ingredient of our analysis is the "three points identity" in the rescaled dual space.

Let $H = \operatorname{diag}(h_i)_{i=1}^q$ be a diagonal matrix with $h_i > 0, i = 1,\dots,q$ and a,b,c are three vectors in $\mathbb{R}^q$. The following three points identity in a rescaled $\mathbb{R}^q$

$$\frac{1}{2}\|a-b\|_H^2 + \frac{1}{2}\|b-c\|_H^2 - \frac{1}{2}\|a-c\|_H^2 = (a-b,c-b)_H = (a-b)^T H(c-b) \qquad (29)$$

follows immediately from the standard three point identity

$$\frac{1}{2}\|A-B\|^2 + \frac{1}{2}\|B-C\|^2 - \frac{1}{2}\|A-C\|^2 = (A-B,C-B)$$

by taking $A = H^{1/2}a, B = H^{1/2}b, C = H^{1/2}c$.

In what follows the Hausdorff distance between two compact sets in $\mathbb{R}^q_+$ will play an important role.

Let X and Y be two bounded closed sets in $\mathbb{R}^n$ and $d(x,y) = \|x-y\|$ the Euclidean distance between $x \in X, y \in Y$. Then the Hausdorff distance between X and Y is defined as follows

$$d_H(X,Y) = \max\{\max_{x\in X}\min_{y\in Y} d(x,y), \max_{y\in Y}\min_{x\in X} d(x,y)\}.$$

For any pair of compact sets X and $Y \subset \mathbb{R}^n$

$$d_H(X,Y) = 0 \Leftrightarrow X = Y$$

Let $Q \subset \mathbb{R}^q_{++}$ be a compact set, $\hat{Q} = \mathbb{R}^q_{++} \setminus Q$, $S(u,\epsilon) = \{v \in \mathbb{R}^q_+ : \|u - v\| \le \epsilon\}$ and

$$\partial Q = \{u \in Q | \exists v \in Q : v \in S(u,\epsilon), \exists \hat{v} \in \hat{Q} : \hat{v} \in S(u,\epsilon)\}, \forall \epsilon > 0$$

be the boundary of Q.

Let $A \subset B \subset C$ be convex compact sets in $\mathbb{R}^q_+$. The following inequality follows from the definition of Hausdorff distance.

$$d_H(A,\partial B) < d_H(A,\partial C) \qquad (30)$$

We consider the dual sequence $\{\lambda^s\}_{s=0}^{\infty}$ generated by IQP (28), the corresponding convex and bounded level sets $L_s = \{\lambda \in \mathbb{R}^q_+ : d(\lambda) \geq d(\lambda^s)\}$ and their boundaries $\partial L_s = \{\lambda \in L_s : d(\lambda) = d(\lambda^s)\}$.

Let $e = (1, \ldots, 1) \in \mathbb{R}^q_{++}$,

$$d(\lambda^0) - k_0^{-1} B_\varphi(\lambda^0, e) = \max\{d(\lambda) - k_0^{-1} B_\varphi(\lambda, e) | \lambda \in \mathbb{R}^q_+\} \tag{31}$$

and $\lambda^0 \neq \lambda^*$.

It follows from concavity of $d(\lambda)$ and boundedness of L^* that $L_0 = \{\lambda : d(\lambda) \geq d(\lambda^0)\}$ is bounded, therefore

$$\max\{\max_{1 \leq i \leq q} \lambda_i | \lambda \in L_0\} = L < \infty \tag{32}$$

Theorem 3. *1) If $\varphi \in \Phi$ is well defined then the following bound holds*

$$d(\lambda^*) - d(\lambda^0) < k_0^{-1} \sum_{i=1}^{q} \varphi(\lambda_i^*); \tag{33}$$

2) the sequence $\{\lambda^s\}_{s=0}^{\infty}$ is monotone increasing in value, i.e. $d(\lambda^s) > d(\lambda^{s-1}), s \geq 1$ and

$$d(\lambda^*) - d(\lambda^s) \leq d(\lambda^*) - d(\lambda^{s-1}) - (2k_s)^{-1} \frac{m}{L^2} \|\lambda^s - \lambda^{s-1}\|^2; \tag{34}$$

3) the Hausdorff distance between the optimal set L^* and ∂L_s is monotone decreasing, i.e.

$$d_H(L^*, \partial L_s) < d_H(L^*, \partial L_{s-1}), \forall s \geq 1; \tag{35}$$

4) for $k_0 > 0$ large enough and any nondecreasing sequence $\{k_s\}_{s=0}^{\infty}$ we have

$$\lim_{s \to \infty} d_H(L^*, \partial L_s) = 0. \tag{36}$$

Proof. 1) For any well defined $\varphi \in \Phi$ from (31) we have $d(\lambda^0) - k_0^{-1} B_\varphi(\lambda^0, e) \geq d(\lambda^*) - k_0^{-1} B_\varphi(\lambda^*, e) = d(\lambda^*) - k_0^{-1} \sum_{i=1}^{q} \varphi(\lambda_i^*)$. Keeping in mind $B_\varphi(\lambda^0, e) > 0$ we obtain (33).

2) From $B_\varphi(\lambda^{s-1}, \lambda^{s-1}) = 0$ follows $d(\lambda^s) \geq d(\lambda^{s-1}) + (2k_s)^{-1} B_\varphi(\lambda^s, \lambda^{s-1})$. Therefore $d(\lambda^*) - d(\lambda^s) \leq d(\lambda^*) - d(\lambda^{s-1}) - (2k_s)^{-1} B_\varphi(\lambda^s, \lambda^{s-1})$. From 2) of Lemma 2 we have

$$d(\lambda^*) - d(\lambda^s) \leq d(\lambda^*) - d(\lambda^{s-1}) - (2k_s)^{-1} m \sum \frac{(\lambda_i^s - \lambda_i^{s-1})^2}{(\lambda_i^{s-1})^2} \tag{37}$$

The bound (34) follows from (32) and (37).

3) From concavity, continuity of $d(\lambda)$ and boundedness of L^* for any $s > 0$ follows that level sets L_s are convex compacts in $\mathbb{R}^q_+$. From (34) we have $L^* \subset L_{s+1} \subset L_s, s \geq 0$. Therefore from (30) follows (35).

4) It follows from (28) that

$$d(\lambda^s) - d(\lambda^{s-1}) \geq (2k_s)^{-1} \|\lambda^s - \lambda^{s-1}\|^2_{H_\varphi^{s-1}}. \tag{38}$$

The sequence $\{d(\lambda^s)\}_{s=0}^{\infty}$ is monotone increasing and bounded by $d(\lambda^*)$, therefore there exists $\lim\limits_{s\to\infty} d(\lambda^s) = \bar{d}$.

First, we show that $\bar{d} = d(\lambda^*)$. Let $g^s \in \partial d(\lambda^s)$ be subgradient of $d(\lambda)$ at $\lambda = \lambda^s$. The optimality criteria

$$(\lambda - \lambda^s, k_s g^s - H_\varphi^{s-1}(\lambda^s - \lambda^{s-1})) \leq 0, \quad \forall \lambda \in \mathbb{R}^q_+$$

for the maximizer λ^s in (28) we can rewrite as follows

$$k_s(\lambda - \lambda^s, g^s) \leq (\lambda - \lambda^s)^T H_\varphi^{s-1}(\lambda^s - \lambda^{s-1}), \forall \lambda \in \mathbb{R}^q_+. \tag{39}$$

Using concavity of $d(\lambda)$ and (39) for $\lambda \in \mathbb{R}^q_+$ we obtain

$$k_s(d(\lambda^s) - d(\lambda)) \geq k_s(g^s, \lambda^s - \lambda) \geq (\lambda - \lambda^s)^T H_\varphi^{s-1}(\lambda^{s-1} - \lambda^s). \tag{40}$$

Using the three point identity (29) with $a = \lambda, b = \lambda^{s-1}, c = \lambda^s$ and $H = H_\varphi^{s-1}$ from (40) we have

$$\begin{aligned} k_s(d(\lambda^s) - d(\lambda)) \geq (\lambda - \lambda^s)^T H_\varphi^{s-1}(\lambda^{s-1} - \lambda^s) = \\ \frac{1}{2}(\lambda - \lambda^s)^T H_\varphi^{s-1}(\lambda - \lambda^s) + \frac{1}{2}(\lambda^s - \lambda^{s-1})^T H_\varphi^{s-1}(\lambda^s - \lambda^{s-1}) - \\ \frac{1}{2}(\lambda - \lambda^{s-1})^T H_\varphi^{s-1}(\lambda - \lambda^{s-1}), \quad \forall \lambda \in \mathbb{R}^q_+. \end{aligned} \tag{41}$$

It follows from (41) that for any $\lambda \in L^*$

$$\begin{aligned} 0 \geq k_s(d(\lambda^s) - d(\lambda)) \geq \\ \frac{1}{2}\|\lambda - \lambda^s\|^2_{H_\varphi^{s-1}} + \frac{1}{2}\|\lambda^s - \lambda^{s-1}\|^2_{H_\varphi^{s-1}} - \frac{1}{2}\|\lambda - \lambda^{s-1}\|^2_{H_\varphi^{s-1}}, \forall s \;\; \geq \;\; 1. \end{aligned} \tag{42}$$

For $h_\varphi^{i,s} = 2\varphi''(1 + \theta_i^s(\frac{\lambda_i^s}{\lambda_i^{s-1}} - 1))(\lambda_i^{s-1})^{-2}, i = 1, \ldots, q$ and $s_0 > 0$ large enough and any $s \geq s_0$ due to (12) there is $\delta_s > 0$ small enough that

$$2(\varphi''(1) - \delta_s)(\lambda_i^{s-1})^{-2} \leq h_\varphi^{i,s} \leq 2(\varphi''(1) + \delta_s)(\lambda_i^{s-1})^{-2}, \qquad i - 1, \ldots, q \tag{43}$$

and $\lim\limits_{s\to\infty} \delta_s = 0$.

We can rewrite the inequality (42) as follows

$$\|\lambda^* - \lambda^s\|^2_{H_\varphi^{s-1}} - \|\lambda^* - \lambda^{s-1}\|^2_{H_\varphi^{s-1}} \leq k_s(d(\lambda^s) - d(\lambda^*)). \tag{44}$$

Let us assume $\varphi''(1) = 1$ (it is true for φ_1 and φ_2). Then, using (43), we can rewrite (44) as follows

$$\Delta_s = (1+\delta_s)\sum_{i=1}^{q}(\frac{\lambda_i^* - \lambda_i^{s-1}}{\lambda_i^{s-1}})^2 - (1-\delta_s)\sum_{i=1}^{q}(\frac{\lambda_i^* - \lambda_i^{s}}{\lambda_i^{s-1}})^2$$
$$= \|\lambda^* - \lambda^{s-1}\|^2_{H_{s-1}} - \|\lambda^* - \lambda^s\|^2_{H_{s-1}} + \tag{45}$$
$$\delta_s(\|\lambda^* - \lambda^{s-1}\|^2_{H_{s-1}} + \|\lambda^* - \lambda^s\|^2_{H_{s-1}})$$
$$\geq k_s(d(\lambda^*) - d(\lambda^s)). \tag{46}$$

For $k_0 > 0$ large enough and any nondecreasing sequence $\{k_s\}_{s=0}^\infty$, it follows from (12) that $\|\lambda^* - \lambda^{s-1}\|^2_{H_{s-1}} - \|\lambda^* - \lambda^s\|^2_{H_{s-1}} \to 0$, which together with boundedness of $(\|\lambda^* - \lambda^{s-1}\|^2_{H_{s-1}} + \|\lambda^* - \lambda^s\|^2_{H_{s-1}})$ and $\delta_s \to 0$ leads to $\Delta_s \to 0$.

On the other hand, if $\bar{d} = \lim_{s\to\infty} d(\lambda^s) < d(\lambda^*)$ then $k_s(d(\lambda^*) - d(\lambda^s)) \geq k_0(d(\lambda^*) - d(\bar{\lambda}))$, which makes (45) impossible for $k_0 > 0$ large enough and any $s \geq s_0$, *i.e.* $\lim_{s\to\infty} d(\lambda^s) = d(\lambda^*)$.

Convergence property (36) follows from the definition of the Hausdorff distance and $\lim_{s\to\infty} d(\lambda^s) = d(\lambda^*)$. □

Remark 4. *It follows from (33) that for any well defined kernel $\varphi \in \Phi$ and $\kappa_0 > 0$ large enough the properties (43) hold for $s > 1$.*

Now we will establish the upper bound for the number of steps required by the IQP method (28) for finding an $\epsilon-$approximation for $d(\lambda^*)$. We consider the dual sequence $\{\lambda^s\}_{s=0}^\infty$ generated by IQP (28). The following lemma plays the key role in establishing the upper bound.

Lemma 3. *For any well defined kernel $\varphi \in \Phi$, any nondecreasing sequence $\{k_s\}_{s=0}^\infty$ and $k_0 > 0$ large enough, the following bound holds*

$$d(\lambda^*) - d(\lambda^s) \leq (1 - \frac{\|\lambda^s - \lambda^{s-1}\|_{H_\varphi^{s-1}}}{2\sqrt{q}})(d(\lambda^*) - d(\lambda^{s-1})), \forall s \geq 1. \tag{47}$$

Proof. Without restricting generality we can assume $d(\lambda^{s-1}) = 0$. Let $\|\lambda^* - \lambda^{s-1}\| = \min\{\|\lambda - \lambda^{s-1}\| | \lambda \in L^*\}$

Consider $[\lambda^*, \lambda^{s-1}] = \{\lambda : \lambda = (1-t)\lambda^{s-1} + t\lambda^* | 0 \leq t \leq 1\}$ and let $\bar{\lambda}^s = [\lambda^*, \lambda^{s-1}] \bigcap \partial E(\lambda^{s-1}, r_{s-1})$ where $r_{s-1} = \|\lambda^s - \lambda^{s-1}\|_{H_\varphi^{s-1}}$ and $\partial E(\lambda^{s-1}, r_{s-1}) = \{\lambda : (\lambda - \lambda^{s-1})^T H_\varphi^{s-1}(\lambda - \lambda^{s-1}) = r_{s-1}^2\}$ is the boundary of the interior ellipsoid $E(\lambda^{s-1}, r_{s-1})$ (see Fig 1).

It follows from (23) that $d(\lambda^s) \geq d(\bar{\lambda}^s)$ and $\|\bar{\lambda}^s - \lambda^{s-1}\|_{H_\varphi^{s-1}} = \|\lambda^s - \lambda^{s-1}\|_{H_\varphi^{s-1}} = r_{s-1}$. From concavity of $d(\lambda)$ follows $d(\bar{\lambda}^s) \geq \hat{d}(\bar{\lambda}^s)$. Therefore (see Figure 1) we have

$$\frac{\hat{d}(\bar{\lambda}^s)}{d(\lambda^*)} = \frac{\|\bar{\lambda}^s - \lambda^{s-1}\|}{\|\lambda^* - \lambda^{s-1}\|} = \frac{\|\bar{\lambda}^s - \lambda^{s-1}\|_{H_\varphi^{s-1}}}{\|\lambda^* - \lambda^{s-1}\|_{H_\varphi^{s-1}}} = \frac{\|\lambda^s - \lambda^{s-1}\|_{H_\varphi^{s-1}}}{\|\lambda^* - \lambda^{s-1}\|_{H_\varphi^{s-1}}}. \tag{48}$$

Therefore from (43) and (48) we obtain

$$\frac{\hat{d}(\bar{\lambda}^s)}{d(\lambda^*)} = \frac{\|\lambda^s - \lambda^{s-1}\|_{H_\varphi^{s-1}}}{\|\lambda^* - \lambda^{s-1}\|_{H_\varphi^{s-1}}}. \tag{49}$$

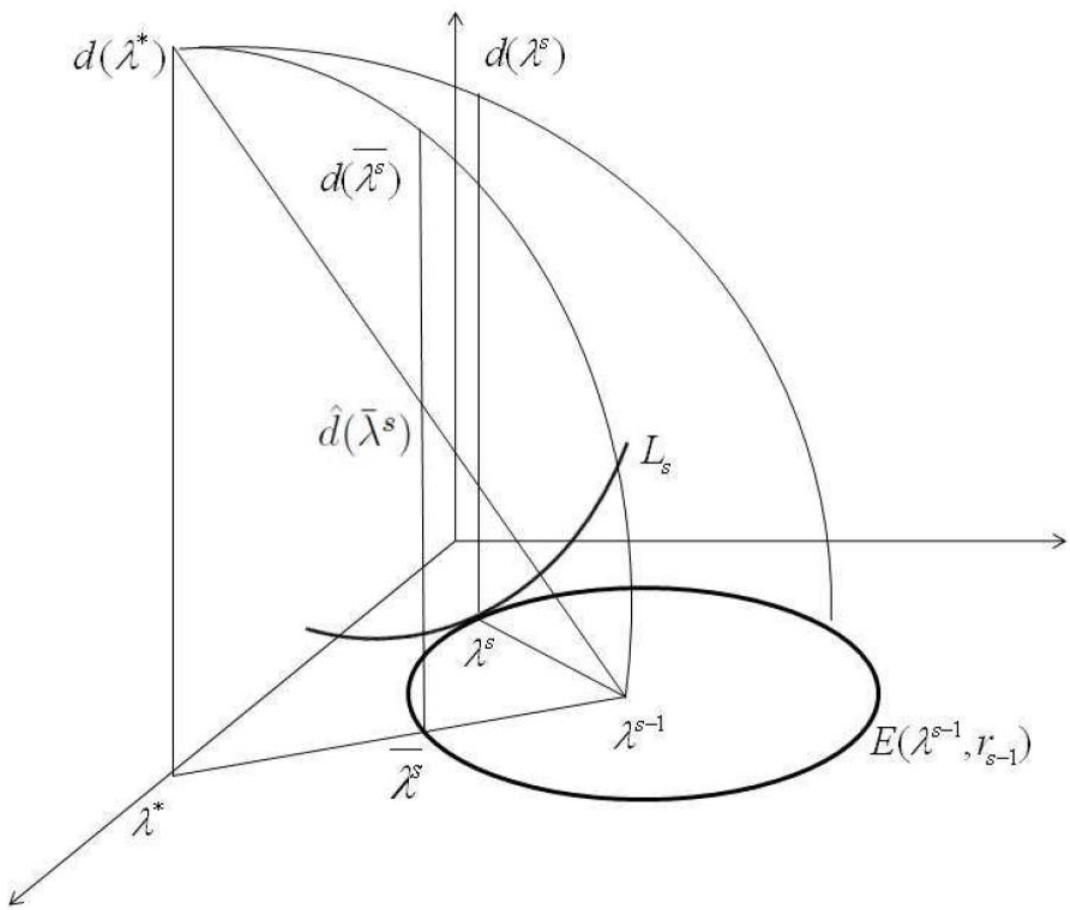

Figure 1: Interior Ellipsoid

Due to (33), (34) and (36) for a large enough $k_0 > 0$ and any $i \in I_+^* = \{i : \lambda_i^* > 0\}$ we have $\lambda_i^s \geq 0.5\lambda_i^*, s > 1$, therefore

$$\sum_{i=1}^{q} (\lambda_i^* - \lambda_i^{s-1})^2 (\lambda_i^{s-1})^{-2} \leq q.$$

Also assuming $\varphi''(1) = 1$ for $k_0 > 0$ large enough and a nondecreasing sequence $\{k_s\}_{s=0}^{\infty}$ due to (12) and (43) we have $\varphi''(1 + \theta_i^{s-1}(\lambda_i^s/\lambda_i^{s-1} - 1)) \leq 2\varphi''(1) = 2, s > 1$.

Then

$$\|\lambda^* - \lambda^{s-1}\|_{H_\varphi^{s-1}}^2 = \sum_{i=1}^{q} 2\varphi''(1 + \theta_i^{s-1}(\lambda_i^s/\lambda_i^{s-1} - 1)) \left(\frac{\lambda_i^* - \lambda_i^{s-1}}{\lambda_i^{s-1}} \right)^2 \leq 4q \tag{50}$$

From (49) and (50) we have

$$\frac{\hat{d}(\bar{\lambda}^s)}{d(\lambda^*)} \geq \frac{\|\lambda^s - \lambda^{s-1}\|_{H_\varphi^{s-1}}}{2\sqrt{q}}. \tag{51}$$

Keeping in mind the following inequalities (see Figure 1)

$$d(\lambda^*) - d(\lambda^s) \leq d(\lambda^*) - d(\bar{\lambda}^s) \leq d(\lambda^*) - \hat{d}(\bar{\lambda}^s)$$

and $d(\lambda^{s-1}) = 0$ from (51) follows the bound (47). $\square$

We are ready to estimate the number of steps N of the IQP method (28) needs for finding $x^N : \Delta_N = d(\lambda^*) - d(\lambda^N) \leq \epsilon$ where $1 >> \epsilon > 0$ is the required accuracy. Let $\sigma_s = \sum_{i=0}^{s} k_i, \Delta_s = d(\lambda^*) - d(\lambda^s)$. To simplify notation we will use H_s instead of H_φ^s.

Theorem 4. *If $\varphi \in \Phi$ is well defined, $k_0 > 0$ is large enough and $\{k_s\}_{s=0}^{\infty}$ is a nondecreasing sequence, then N is the smallest integer such that*

$$\sigma_N \geq O\left(\frac{q}{\epsilon} \ln \frac{\Delta_0}{\epsilon}\right) \tag{52}$$

Proof. Multiplying (38) by σ_{s-1} we obtain

$$(\sigma_s - k_s)d(\lambda^s) - \sigma_{s-1}d(\lambda^{s-1}) \geq \sigma_{s-1}\frac{1}{2k_s}\|\lambda^s - \lambda^{s-1}\|_{H_{s-1}}^2, s \geq 1.$$

Summing up the last inequality over $s = 1, \ldots, N$ we have

$$\sigma_N d(\lambda^N) - \sum_{s=1}^{N} k_s d(\lambda^s) \geq \frac{1}{2}\sum_{s=1}^{N} \frac{\sigma_{s-1}}{k_s}\|\lambda^s - \lambda^{s-1}\|_{H_{s-1}}^2. \tag{53}$$

Summing up (41) over $s = 1, \ldots, N$ we obtain

$$-\sigma_N d(\lambda) + \sum_{s=1}^{N} k_s d(\lambda^s) \geq$$

$$\frac{1}{2}\left(\sum_{s=1}^{N}(\|\lambda - \lambda^s\|^2_{H_{s-1}} - \|\lambda - \lambda^{s-1}\|^2_{H_{s-1}}) + \sum_{s=1}^{N}\|\lambda^s - \lambda^{s-1}\|^2_{H_{s-1}}\right). \tag{54}$$

Adding up (53) and (54) we obtain

$$\sigma_N(d(\lambda^N) - d(\lambda)) \geq \frac{1}{2}\sum_{s=1}^{N}(-\|\lambda - \lambda^{s-1}\|^2_{H_{s-1}} + \|\lambda - \lambda^s\|^2_{H_{s-1}}) + \sum_{s=1}^{N}(\frac{\sigma_{s-1}}{k_s} + 1)\|\lambda^s - \lambda^{s-1}\|^2_{Hs-1},$$

therefore using (50) for any $\lambda^* \in L^*$ we have

$$\begin{aligned}
\sigma_N(d(\lambda^*) - d(\lambda^N)) &\leq \frac{1}{2}\sum_{s=1}^{N}(\|\lambda^* - \lambda^{s-1}\|^2_{H_{s-1}} - \|\lambda^* - \lambda^s\|^2_{H_{s-1}}) \\
&= \frac{1}{2}\sum_{s=1}^{N}(\|\lambda^* - \lambda^{s-1}\|_{H_{s-1}} - \|\lambda^* - \lambda^s\|)(\|\lambda^* - \lambda^{s-1}\|_{H_{s-1}} + \|\lambda^* - \lambda^s\|_{H_{s-1}}) \\
&\leq 2\sqrt{q}\sum_{s=1}^{N}(\|\lambda^* - \lambda^{s-1}\|_{H_{s-1}} - \|\lambda^* - \lambda^s\|_{H_{s-1}}).
\end{aligned}$$

Using the triangle inequality $\|\lambda^s - \lambda^{s-1}\|_{H_{s-1}} \geq \|\lambda^* - \lambda^{s-1}\|_{H_{s-1}} - \|\lambda^* - \lambda^s\|_{H_{s-1}}$ we have

$$\sigma_N(d(\lambda^*) - d(\lambda^N)) \leq 2\sqrt{q}\sum_{s=1}^{N}\|\lambda^s - \lambda^{s-1}\|_{H_{s-1}}. \tag{55}$$

It follows from (47) that

$$d(\lambda^*) - d(\lambda^N) \leq \prod_{s=1}^{N}(1 - \frac{\|\lambda^s - \lambda^{s-1}\|_{H_{s-1}}}{2\sqrt{q}})(d(\lambda^*) - d(\lambda^0)),$$

therefore

$$\ln\frac{d(\lambda^*) - d(\lambda^N)}{d(\lambda^*) - d(\lambda^0)} \leq \sum_{s=1}^{N}\ln(1 - \frac{\|\lambda^s - \lambda^{s-1}\|_{H_{s-1}}}{2\sqrt{q}}).$$

Keeping in mind $\ln(1+x) \leq x, \forall x > -1$ we obtain,

$$\ln\frac{(d(\lambda^*) - d(\lambda^N))}{(d(\lambda^*) - d(\lambda^0))} \leq -\sum_{s=1}^{N}\frac{\|\lambda^s - \lambda^{s-1}\|_{H_{s-1}}}{2\sqrt{q}}.$$

or

$$\Delta_N = d(\lambda^*) - d(\lambda^N) \leq (d(\lambda^*) - d(\lambda^0))e^{-\sum_{s=1}^{N}\frac{\|\lambda^s - \lambda^{s-1}\|_{H_{s-1}}}{2\sqrt{q}}}, \tag{56}$$

i.e.

$$\Delta_N \le \Delta_0 e^{-\sum_{s=1}^{N} \frac{\|\lambda^s - \lambda^{s-1}\|_{H_{s-1}}}{2\sqrt{q}}}. \tag{57}$$

On the other hand from (55) we have

$$-2\sqrt{q}\sum_{s=1}^{N} \|\lambda^s - \lambda^{s-1}\|_{H_{s-1}} \le -\sigma_N \Delta_N$$

or

$$-\frac{1}{2\sqrt{q}}\sum_{s=1}^{N} \|\lambda^s - \lambda^{s-1}\|_{H_{s-1}} \le \frac{-\sigma_N \Delta_N}{4q}. \tag{58}$$

From (57) and (58) follows

$$\Delta_N \le \Delta_0 e^{\frac{-\sigma_N \Delta_N}{4q}}. \tag{59}$$

Let $1 >> \epsilon > 0$ be the required accuracy. Therefore if

$$\Delta_0 e^{-\frac{\sigma_N \Delta_N}{4q}} \le \epsilon$$

then

$$-\frac{\sigma_N \Delta_N}{4q} \le \ln \frac{\epsilon}{\Delta_0},$$

i.e.

$$\Delta_N \ge \frac{4q}{\sigma_N} \ln \frac{\Delta_0}{\epsilon}. \tag{60}$$

For $N : \Delta_N \le \epsilon$ from (60) we have

$$\sigma_N \ge O\left(\frac{q}{\epsilon} \ln \frac{\Delta_0}{\epsilon}\right), \tag{61}$$

i.e. N is the smallest integer for which (61) holds. □

Corollary By taking $\kappa_s = \kappa_0 + s$ we have $\sigma_N = Nk_0 + \frac{N(N+1)}{2}$, therefore it takes

$$N = O\left(\frac{q}{\epsilon} \ln \frac{\Delta_0}{\epsilon}\right)^{\frac{1}{2}} \tag{62}$$

steps to find an ϵ-approximation for $d(\lambda^*)$.

6 Lagrangian Transformation and Affine Scaling method for LP

Let $a \in \mathbb{R}^n, b \in \mathbb{R}^q$ and $A : \mathbb{R}^n \to \mathbb{R}^q$ are given. We consider the following linear programming problem LP

$$x^* \in X^* = \text{Argmin}\{(a, x) | c(x) = Ax - b \ge 0\} \tag{63}$$

and the dual LP

$$\lambda^* \in L^* = \text{Argmin}\{(b,\lambda)|r(\lambda) = A^T\lambda - a = 0, \lambda \in \mathbb{R}^q_+\}. \tag{64}$$

We assume that $X^* \neq \phi$ is bounded and so is the dual optimal set L^*.

The LT $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}_{++} \to \mathbb{R}$ for LP (63) is defined as follows

$$\mathcal{L}(x,\lambda,k) = (a,x) - k^{-1}\sum_{s=1}^{q} \psi(k\lambda_i c_i(x)) \tag{65}$$

where $c_i(x) = (Ax-b)_i = (a_i,x) - b_i, \quad i = 1,\dots,q$.

Let's consider the LT method with from step to step updated scaling parameter. We assume that $\{k_s\}_{s=0}^{\infty}$ is a positive nondecreasing sequence.

The LT step consists of finding the primal–dual pair (x^{s+1},λ^{s+1}) by formulas

$$x^{s+1} : \nabla_x \mathcal{L}(x^{s+1},\lambda^s,k_s) = 0 \tag{66}$$

$$\lambda^{s+1} : \lambda_i^{s+1} = \lambda_i^s \psi'(k_s\lambda_i^s c_i(x^{s+1})), i = 1,\dots,q. \tag{67}$$

Theorem 5 If the primal optimal X^* is bounded then the LT method (66)– (67) is well defined for any transformation $\psi \in \Psi$. For the dual sequence $\{\lambda^s\}_{s=0}^{\infty}$ generated by (67) the following statements hold true:

1) the LT method (66)– (67) is equivalent to the following Interior Prox

$$k_s(b,\lambda^{s+1}) - B_\varphi(\lambda^{s+1},\lambda^s) = \max\{k_s(b,\lambda) - B_\varphi(\lambda,\lambda^s)|A^T\lambda = 0\}$$

where $B_\varphi(u,v) = \sum_{i=1}^q \varphi(\frac{u_i}{v_i})$ is the Bregman type distance;

2) for $k > 0$ large enough the LT method (66)– (67) with truncated MBF transformation $\psi_2(t)$ is equivalent to the affine scaling type method for the dual LP.

Proof

1) We use the vector form for formula (67) assuming that the multiplication and division are componentwise, *i.e.* for vectors $a,b \in \mathbb{R}^n$, the vector $c = ab = (c_i = a_ib_i, \quad i = 1,\dots,n)$ and the vector $d = a/b = (d_i = a_i/b_i, \quad i = 1,\dots,n)$. We have

$$\frac{\lambda^{s+1}}{\lambda^s} = \psi'(k_s\lambda^s c(x^{s+1})). \tag{68}$$

Using again the inverse function formula we obtain

$$k_s\lambda^s c(x^{s+1}) = \psi'^{-1}(\lambda^{s+1}/\lambda^s). \tag{69}$$

It also follows from (66) and (67) that

$$\begin{aligned}\nabla_x \mathcal{L}(x^{s+1},\lambda^s,k_s) &= a - A^T\psi'(k_s\lambda^s c(x^{s+1}))\lambda^s = a - A^T\lambda^{s+1}\\ &= \nabla_x L(x^{s+1},\lambda^{s+1}) = 0,\end{aligned}$$

i.e.

$$d(\lambda^{s+1}) = L(x^{s+1}, \lambda^{s+1}) = (a, x^{s+1}) - (\lambda^{s+1}, Ax^{s+1} - b) \;\; = \\ (a - A^T\lambda^{s+1}, x^{s+1}) + (b, \lambda^{s+1}) = (b, \lambda^{s+1}).$$

Using the identity $\psi'^{-1} = \psi^{*\prime}$ and $\varphi = -\psi^*$ we can rewrite (69) as follows

$$-k_s c(x^{s+1}) - (\lambda^s)^{-1}\varphi'(\lambda^{s+1}/\lambda^s) = 0. \tag{70}$$

Keeping in mind $A^T\lambda^{s+1} = a, -c(x^{s+1}) \in \partial d(\lambda^{s+1})$ and $\lambda^{s+1} \in \mathbb{R}^q_{++}$ we can view (70) as the optimality criteria for the following problem

$$k_s(b, \lambda^{s+1}) - B_\varphi(\lambda^{s+1}, \lambda^s) = \max\{k_s d(\lambda) - B_\varphi(\lambda, \lambda^s)|A^T\lambda = a\}, \tag{71}$$

where $B_\varphi(\lambda, \lambda^s) = \sum_{i=1}^{q} \varphi(\lambda_i/\lambda_i^s)$ is Bregman type distance.

2) Let's consider the LT method with truncated MBF transformation $\psi_2(t)$. If follows from Corollory that for $k > 0$ large enough only MBF kernel $\varphi_2(t) = -\ln t + t - 1$, and correspondent Bregman distance

$$B_{\varphi_2}(\lambda, \lambda_s) = \sum_{i=1}^{q}(-ln\frac{\lambda_i}{\lambda_i^s} + \frac{\lambda_i}{\lambda_i^s} - 1)$$

will be used in (71) for any $s \geq 1$. Using considerations similar to those in item 4) Theorem 2 we can rewrite (71) as follows

$$k_s(b, \lambda^{s+1}) = \arg\max\{k_s(b, \lambda)|\lambda \in E(\lambda^s, r_s),\ A^T\lambda = a\}, \tag{72}$$

where $r_s^2 = Q_{\varphi_2}(\lambda^{s+1}, \lambda^s)$ and $E(\lambda^s, r_s) = \left\{\lambda : (\lambda - \lambda^s)^T H_s^{-2}(\lambda - \lambda^s) \leq r_s\right\}$ is Dikin's ellipsoid.

The problem type (72) one solves at each step of I. Dikin's type affine scaling method [7].

7 Concluding Remarks

We would like to point out that the convergence analysis of the LT method and its dual equivalent does not assume that the primal problem is well structured (see [21] Ch 4) Therefore the bound (52) is true for any convex optimization problems.

The numerical realization of the LT method (8)- (9) requires replacing at each step the minimizer x^s in (8) by its appropriate approximation.

The optimal gradient methods (2.2.6) and (2.2.8) seem to be very promising tools for finding an approximation for x^s in (8) (see Th. 2.2.2 and 2.2.3 in [21]).

The Primal-Dual approach, which happens to be efficient in the NR setting (see [11],[28]) can be used for LT if the second order optimality conditions are satisfied.

If (P) is a non-convex problem and the second order optimality conditions are satisfied then it follows from 3^0 that $\mathcal{L}(x, \lambda, k)$ is strongly convex in the neighborhood of x^* for any given $\lambda \in \mathbb{R}^q_{++}$ and any $k \geq k_\lambda$ where $k_\lambda > 0$ is large enough. Therefore after finding the primal approximation for the first minimizer in (8) the LT method (8)-(9) will require finding at each step an approximation for the minimizer of a strongly convex in x function $\mathcal{L}(x, \lambda^s, k_s)$.

Finally, the dual problem

$$d(\lambda^*) = \max\{d(\lambda) | \lambda \in \mathbb{R}^q_+\} \tag{73}$$

where $d : \mathbb{R}^q_+ \to R$ is a closed concave function is always a convex optimization problem no matter if the primal (P) is convex or not. Also, a number of real life application lead to the problem (73).

Therefore it is important to find an efficient algorithm for solving (73). There is evidence that in some instances IEM for the problem (73) generates a sequence which converges to the solution in value with linear rate and the ratio is dependent only on the dimension of the dual space.

We are planning to consider the convergence and complexity issues related to IEM in the up coming paper.

Acknowledgement. The research was supported by NSF Grant CCF–0836338.

References

[1] Auslender A, Teboulle M, Ben–Tiba S. “Interior primal and multipliers methods based on second–order homogeneous kernels,” Math. Oper. Res., vol.24, no. 3, pp. 645–668, 1999.

[2] Barnes E. “A variation on Karmarkar’s algorithm for solving linear programming problems,” Mathematical Programming, vol. 36, pp 174–182, 1986.

[3] Bertsekas D. “ Constrained Optimization and Lagrange Multipliers Methods. Academic Press: New York, 1982

[4] Bregman L. “ The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming,” USSR Computational Mathematics and Mathematical Physics, 7, 200–217, 1967.

[5] Censor Y, Zenios S. “The proximal minimization algorithm with d–functions,” Journal of Optimization Theory and Applications, 73, 451–464, 1992.

[6] Chen G. and Teboulle M. “ Convergence analysis of a proximal–like minimization algorithm using Bregman Functions,” SIAM J. Optimization, 3 (4), pp. 538–543.

[7] Dikin I. “ Iterative solutions of linear and quadratic programming problems,” Soviet Mathematic Doklady, **8** 674–675, 1967.

[8] Dikin I. “On the convergence of an iterative process”, Upravlyaemye systemi,12,54-60 (in Russian) 1974.

[9] Gonzaga C. An algorithm for solving Linear Programming problems in O(n^3L) operations, Progress in Mathematical Programming N. Megiddo. ed Springer-Verlag, Berlin 1-82(1988).

[10] Gill P, Murray W, Saunders M, Tomlin J, Wright M. “ On projected Newton barrier methods for linear programming and an equivalence to Karmarkar’s projected method,” Mathematical Programming **36**, 183—209, 1986.

[11] Griva I, Polyak R. “Primal-Dual nonlinear rescaling method with dynamic scaling parameter update,” Mathematical Programming Series A 106, 237-259 (2006).

[12] Guler O. “On the convergence of the proximal point algorithm for convex minimization,” SIAM J. Control Optim., vol 29, pp 403–419, 1991.

[13] Hestenes M R. “ Multipliers and gradient methods,” J. Optim. Theory Appl., vol. 4, pp. 303–320, 1969.

[14] Eckstein J.“ Nonlinear proximal point algorithms using Bregman functions with applications to convex programming,” Mathematics of Operations Research, 18 (1), 202–226, 1993.

[15] Karmarkar N.“ A new polynomial time algorithm for linear programing,” Combinatorica, **4**, 373–395, 1984.

[16] Kočvara M., Stingl M. “On the solution of large-scale SDP problems by the modified barrier method using iterative solvers.” Mathematical Programming, Series B, 109(2-3):413-444, 2007.

[17] Matioli L, Gonzaga C. “ A new family of penalties for Augmented Lagrangian methods,” Numerical Linear Algebra with Applications, 15, 925–944, 2008.

[18] Martinet B.“ Regularization d’ inequations variationelles par approximations successive,” Revue Francaise d’Automatique et Informatique Recherche Operationelle, 4, pp. 154–159, 1970.

[19] Moreau J.“ Proximite’ et dualite’ dans un espace Hilbertien,” Bull. Soc. Math. France, 93, pp. 273–299, 1965.

[20] Nesterov Yu., Nemirovsky A. “ Interior Point Polynomial Algorithms in Convex Programming,” SIAM, Philadelphia (1994).

[21] Nesterov Yu. “Introductory Lectures on Convex Optimization,” Kluwer Academic Publishers, Norwell, MA (2004).

[22] Polyak B. “ Introduction to Optimization,” Optimization Software, New York, NY, 1987.

[23] Polyak R.“ Modified barrier functions (theory and methods),” Math. Programming, 54, pp. 177–222, 1992.

[24] Polyak R, Teboulle M.“ Nonlinear rescaling and proximal–like methods in convex optimization. Mathematical Programming, 76, 265–284, 1997.

[25] Polyak R.“ Nonlinear rescaling vs. smoothing technique in convex optimization,” Math. Programming, 92, 197–235, 2002.

[26] Polyak R. Lagrangian Transformation in Convex Optimization, Research Report-072003, Department of SEOR and Mathematical Science Department, GMU, pg. 1-23 (2003).

[27] Polyak R.“ Nonlinear rescaling as Interior Quadratic Prox Method in convex optimization,” Computational Optimization and Applications, 35, 347–373, 2006.

[28] Polyak R.“ Primal–Dual Exterior Point Method for Convex Optimization,” Optimization methods and Software, 23(1), pp. 141–160, 2008.

[29] Potra F. A. and Wright S. J. "Interior-point methods." Journal of Computational and Applied Mathematics, 124:281,302, 2000.

[30] Powell M. J. D."A method for nonlinear constraints in minimization problems," in Fletcher (Ed.), Optimization, London Academic Press, pp. 283–298, 1969.

[31] Renegar J. "A polynomial time algorithm, based on Newton's method for linear programming", Mathematical Programming 40 (1988) 59-93.

[32] Rockafellar R. T."A dual approach to solving nonlinear programming problems by unconstrained minimization," Math. Programming, 5, pp. 354–373, 1973.

[33] Rockafellar R. T." Monotone operators and the proximal point algorithm," SIAM J. Control Optim, 14, pp. 877–898, 1976.

[34] Rockafellar R. T." Augmented Lagrangians and applications of the proximal points algorithms in convex programming," Math. Oper. Res., 1, 97–116, 1976.

[35] Tseng P. Bertsekas D."On the convergence of exponential multiplier method for convex programming," Mathematical Programming, 60, 1–19, 1993.

[36] Teboulle M, "Entropic proximal mappings with applications to nonlinear programming," Mathematics of Operations, 17, 97–116, 1992.

[37] Vandarbei R, Meketon M, and Freedman B." A modification of Karmarkar's linear programming algorithm," Algorithmica **1**, 4, 395–407, 1986.