EXTERIOR DISTANCE FUNCTION

ROMAN A. POLYAK

Abstract. We introduce and study exterior distance function (EDF) and correspondent exterior point method (EPM) for convex optimization.

The EDF is a classical Lagrangian for an equivalent problem obtained from the initial one by monotone transformation of both the objective function and the constraints.

The constraints transformation is scaled by a positive scaling parameter. Thus, the EDF is a particular realization of the Nonlinear Rescaling (NR) principle.

Along with the "center", the EDF has two extra tools: the barrier (scaling) parameter and the vector of Lagrange multipliers.

We show that EPM generates primal - dual sequence, which converges to the primal - dual solution in value under minimum assumption on the input data. Moreover, the convergence is taking place under any fixed interior point as a "center" and any fixed positive scaling parameter, just due to the Lagrange multipliers update.

If the second order sufficient optimality condition is satisfied, then the EPM converges with Q-linear rate under any fixed interior point as a "center" and any fixed, but large enough positive scaling parameter.

1. Introduction

The Interior Distance functions (IDFs) were introduced and the Interior Center Methods (ICMs) were developed by P. Huard in the mid - 60^{th} (see [4],[11],[12]).

Later IDFs and correspondent ICMs were incorporated into SUMT and studied by A. Fiacco and G. McCormick in [5] and other authors (see, for example, [10],[20] and references therein).

At each step ICM finds a central (in a sense) point of the Relaxation Feasible Set (RFS) and updates the level set using the new objective function value. The RFS is the intersection of the feasible set and the relaxation (level) set of the objective function at the attained level.

The "center" is sought as a minimizer of the IDF. It is a point in the RFS "most distant" from both the boundary of the objective function level set and the active constraints.

2010 *Mathematics Subject Classification.* Primary 65B9990C2090C25.

Key words and phrases. Lagrangian, Duality, Interior Distance Function, Non-Linear Rescaling, Multipliers Method.

Interest in IDFs and correspondent center methods has grown dramatically after N. Karmarkar published is 1984 his projective scaling method (see [14]). In fact, his potential function is an IDF and his method is a Center Method, which generates centers of spheres belonging to the interior of the polytope.

Mainly for this reason the concept of centers became extremely popular in the 80s. Centering and reducing the cost are two basic ideas behind the Interior Point Methods (IPMs), which was the main stream in Modern Optimization for a long time. Centering means to stay away from the boundary of the RFS. An answer to the basic question: how far from the boundary one should stay in case of LP was given by Sonnevend in [28] (see also [13]) through the definition of analytic center of a polytop. The central path is a curve formed by analytic centers. The curve plays an important role in the IPMs (see [8]).

Following the central path J. Renegar in [27] obtained the first path-following algorithm with $O(\sqrt{n}L)$ number of iterations, versus $O(nL)$ iterations for the N. Karmarkar's method.

Soon after C.Gonzaga [7] and P. Vaidya [29] developed algorithms for LP, based on the centering ideas, with overall complexity $O(n^3L)$ arithmetic operations, which is the best knows result so far.

After Yu. Nesterov and A. Nemirovsky developed their self-concondance theory it becomes evident that path-following methods with polynomial complexity for convex optimization problems is possible if the RFS can be equipped with self-concordant barrier (see [18],[19]).

If it is not the case, then one can use the classical IDF and correspondent ICM.

The classical IDF, however, has well known drawbacks: (1) the IDF, its gradient and Hessian does not exist at the primal solution; (2) the IDF, as well as, the condition number of IDF's Hessian unboundedly grows when the primal approximation approaches the solution. The singularity of the IDF at the solution leads to numerical instability, in particular, in the final phase. It means that from some point on, finding an accurate approximation for the IDF's minimizer is practically an impossible task.

In spite of a long history of IDF and correspondent ICM the fundamental question still is: *how the main idea of center methods: to stay away from the boundary consistent with the main purpose of constrained optimization: finding a solution on the boundary.*

The issue was partially addressed in [24], where the Modified Interior Distance Functions was introduced and correspondent theory and methods were developed. The results in [24], however, were obtained only under the second order sufficient optimality condition.

In this paper we address the issue by introducing the Exterior Distance Function (EDF) and correspondent Exterior Point Method (EPM). The

EDF is a classical Lagrangian for a convex optimization problem equivalent to the initial one and obtained from the latter by transforming both the objective function and the constraints.

So, EDF is a particular realization of the Nonlinear Rescaling (NR) principle, but the main EDF results do not follow from NR theory (see [22],[23],[25]).

We obtained the basic convergence results under minimum assumptions on the input data.

In contrast to the classical IDF, the EDF, its gradient and Hessian are defined on an extended feasible set. It eliminates the singularity of the EDF and its derivatives at the solution.

The EDF has two extra tools, which control the computational process: the positive barrier (scaling) parameter and the vector of Lagrange multipliers.

The EPM alternates finding the EDF primal minimizer with Lagrange multipliers update, while both the "center" and the barrier parameter can be fixed or updated from step to step.

Under a fixed "center" the EDF resemble the Modified barrier Function (MBF) (see [22]), but for a problem equivalent to the initial one. Due to the "center" it provides, on the top of the MBF qualities, an extra one. By changing the "center" from step to step it is possible strengthening convergence results typical for MBF without much extra computational work.

Convergence due to the Lagrange multipliers update allows keeping the condition number of the EDF's Hessian stable, which is critical for numerical stability. This is a fundamental departure from classical IDF theory and methods.

Under standard second order sufficient optimality condition EPM converges with Q-linear rate even when both the "center" and the scaling parameter are fixed, but the parameter is large enough. Therefore by changing the scaling parameter and/or the "center" from step to step one gets superlinear convergence rate versus sublinear, which is typical for the Classical IDF.

Also for a fixed, but large enough scaling parameter and any fixed interior point as a "center" the EDF is strongly convex in the neighborhood of the primal minimizer no matter the objective function and the active constraints are convex or not.

2. Problem formulation and basic assumptions

Let $f:\mathbb{R}^n\to\mathbb{R}$ be convex and all $c_i:\mathbb{R}^n\to\mathbb{R}$, $i=1,...,m$ are concave.

We consider the following convex optimization problem

$$f(x^*)=\min\{f(x)|x\in\Omega\}, \tag{2.1}$$

where

$$\Omega=\{x\in\mathbb{R}^n: c_i(x)\geq 0, i=1,...,m\} \tag{2.2}$$

is a feasible set.

We assume:

A. The primal solution set $X^* = \{x \in \Omega : f(x) = f(x^*)\}$ is not empty and bounded;

B. Slater condition

$$\exists x_0 \in \Omega : c_i(x_0) > 0,\ i = 1, ..., m$$

holds.

Let $y \in \text{int}\ \Omega$, then the relaxation feasible set (RFS)

$$\Omega(y) = \{x \in \Omega : f(x) < f(y)\}$$

is convex and bounded for any given $y \in \text{int}\ \Omega$. It follows from A, convexity f, concavity c_i, $i = 1, ..., m$ and Corollary 20 (see [5]).

Also, without losing generality, we can assume that $f(x) \geq 0$, because, otherwise, we can replace $f(x)$ by an equivalent objective function $f(x) := \ln(e^{f(x)} + 1) \geq 0$.

Throughout the paper we will use the following well known fact.

Lemma 2.1 (Debreu). *Let $A = A^T : \mathbb{R}^n \to \mathbb{R}^n$, $C : \mathbb{R}^n \to \mathbb{R}^r (n > r)$, rank $C = r$ and*

$$(Ax, x) \geq \mu(x, x), \quad \mu > 0, \quad \forall x : Cx = 0,$$

then there is $0 < \rho < \mu$ and large enough $k_0 > 0$ that for any $k \geq k_0$ the following inequality

$$((A + kC^T C)x, x) \geq \rho(x, x), \forall x \in \mathbb{R}^n$$

holds true.

In the following section we recall some IDF properties.

3. Classical Interior Distance Function

Let $y_0 \in \text{int}\ \Omega$ and $\tau_0 = f(y_0)$, then

$$\Omega(\tau_0) = \{x \in \Omega : f(x) \leq \tau_0\}$$

is the RFS at the level $\tau_0 = f(x_0)$.

Let $\tau \in T = \{\tau : \tau_0 > \tau > \tau^* = f(x^*)$, then Huard's IDF $H : \Omega(\tau) \times T \to \mathbb{R}^1$ is defined by

$$H(x, \tau) = -m \ln(\tau - f(x)) - \sum_{i=1}^{m} \ln c_i(x). \tag{3.1}$$

We assume $\ln t = -\infty$, for $t \leq 0$, then Interior Center Method (ICM) step finds the "center"

$$\hat{x} = \hat{x}(\tau) = \text{argmin}\{H(x, \tau)/x \in \mathbb{R}^n\} \tag{3.2}$$

and replaces $\Omega(\tau)$ by $\Omega(\hat{\tau})$, where $\hat{\tau} = f(\hat{x})$.

From the statement $x \to \partial\Omega(\tau) \Rightarrow H(x, \tau) \to \infty$ follows $\hat{x}(\tau) \in \text{int}\ \Omega(\tau)$ for any $\tau \in T$.

If the RFS can be equipped with self - concordant (SC) barrier, that is $H(x,\tau)$ is a SC-function in $x \in \operatorname{int} \Omega(\tau)$, then following the central trajectory $\{\hat{x}(\tau), \tau \in T\}$ one gets an ε-approximation of $f(x^*)$ in $O(\sqrt{n} \ln \varepsilon^{-1})$ steps by alternating Newton's step applied for minimization $H(x,\tau)$ with τ update (see [18], [19], [27] and references therein).

If the RFS can not be equipped with SC barrier, then Classical Huard's IDF and correspondent ICM (3.2) is used.

The IDF $F(x,\tau)$, as well as, its gradient and Hessian are singular at x^*.

For any given $\tau \in T$ we have $\lim_{x \to x^*} H(x,\tau) = \infty$ and the condition number of the Hessian $\nabla_{xx} H(x,\tau)$ unboundedly grows when $\hat{x}(\tau) \to x^*$, which makes finding a reasonable approximation for primal minimizer, from some point on, practically impossible.

Although approximations for the Lagrange multipliers can be found as a by-product of ICM, they cannot be effectively used in the computational process.

Let us consider the issues briefly. From the boundedness of RFS $\Omega(\tau)$ and $\lim_{x \to \partial\Omega(\tau)} H(x,\tau) = \infty$ the unconstrained minimizer always exists and $\hat{x} = \hat{x}(\tau) \in \operatorname{int} \Omega(\tau)$. So we have

$$\nabla_x H(\hat{x},\tau) = \frac{m}{\tau - f(\hat{x})} \, \nabla f(\hat{x}) - \sum_{i=1}^{m} \frac{\nabla c_i(\hat{x})}{c_i(\hat{x})} = 0 \tag{3.3}$$

or

$$\nabla f(\hat{x}) - \sum_{i=1}^{m} \frac{\tau - f(\hat{x})}{m \, c_i(\hat{x})} \, \nabla c_i(\hat{x}) = 0 \, . \tag{3.4}$$

Let

$$\hat{\lambda} = (\hat{\lambda}_i = \hat{\lambda}_i(\tau) = (\tau - f(\hat{x}))(m \, c_i(\hat{x}))^{-1}, \quad i = 1, \ldots, m) \tag{3.5}$$

be the vector of Lagrange multipliers.

Vector $\hat{\lambda}$ is positive because $f(\hat{x}) < \tau$ and all $c_i(\hat{x}) > 0$. The systems (3.3) and (3.4) can be rewritten as follows:

$$\nabla_x H(x,\tau) = \frac{m}{\tau - f(\hat{x})} \nabla_x L(\hat{x}, \hat{\lambda}) = \frac{m}{\tau - f(\hat{x})} (\nabla f(\hat{x}) - \nabla c(\hat{x})^T \hat{\lambda}) = 0, \tag{3.6}$$

where $\nabla c(x) = J(c(x))$ is the $m \times n$ Jacobian of $c(x) = (c_1(x), ..., c_m(x))^T$.

From (3.5) we have

$$\hat{\lambda}_i c_i(\hat{x}) = (\tau - f(\hat{x})) m^{-1}, \ i = 1, \ldots, m \, . \tag{3.7}$$

Summing up (3.7), we obtain

$$\hat{\lambda}^T c(\hat{x}) = \sum_{i=1}^{m} \hat{\lambda}_i c_i(\hat{x}) = \tau - f(\hat{x}) \, .$$

From (3.3) and $\tau > f(\hat{x})$ follows $\tau - f(\hat{x}) \to 0$, when $\tau \to \tau^*$, because f is bounded from below.

Vector $\hat{x} \in \text{ int } \Omega$ is primal feasible, vector $\hat{\lambda} \in R^m_{++}$, is dual feasible and from (3.7) follows asymptotic complementarity condition

$$\lim_{\tau \to \tau^*} \hat{\lambda}_i(\tau) c_i(\hat{x}(\tau)) \to 0,\ i = 1, ..., m.$$

To simplify considerations we assume at this point that the second order sufficient optimality condition for the problem (2.1) is satisfied. Then the primal-dual solution (x^*, λ^*) is unique. Therefore,

$$\lim_{\tau \to \tau^*} \hat{x}(\tau) = x^*, \quad \lim_{\tau \to \tau^*} \hat{\lambda}(\tau) = \lambda^*.$$

Let $I^* = \{i : c_i(x^*) = 0\} = \{1, ..., r\}$ be the active constraints set.

For Hessian $\nabla^2_{xx} H(x, \tau)$ at $x = \hat{x}$ we obtain

$$\begin{aligned}
\nabla^2_{xx} H(x, \tau)_{/x=\hat{x}} &= m(\tau - f(\hat{x}))^{-1} \Big[(\tau - f(\hat{x}))^{-1} \nabla f(\hat{x}) \nabla f^T(\hat{x}) + \nabla^2 f(\hat{x}) \\
&\quad - \sum_{i=1}^m \frac{(\tau - f(\hat{x}))}{m} \frac{\nabla^2 c_i(\hat{x})}{c_i(\hat{x})} + \sum_{i=1}^m \frac{(\tau - f(\hat{x}))}{m c_i^2(\hat{x})} \nabla c_i(\hat{x}) \nabla c_i(\hat{x})^T \Big] \\
&= m(\tau - f(\hat{x}))^{-1} \Big[\nabla^2_{xx} L(\hat{x}, \hat{\lambda}) + \nabla c(\hat{x})^T C^{-1}(\hat{x}) \hat{\Lambda}(\tau) \nabla c(\hat{x}) \\
&\quad + \ (\tau - f(\hat{x}))^{-1} \nabla f(\hat{x}) \nabla f(\hat{x})^T \Big],
\end{aligned}$$

where $C(x) = [\text{diag}\, c_i(x)]^m_{i=1}$ and $\Lambda(\tau) = [diag\, \lambda_i(\tau)]^m_{i=1}$ are diagonal matrices and $L(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i c_i(x)$ is the Lagrangian for problem (2.1).

In view of $\hat{x} = \hat{x}(\tau) \to x^*$ and $\hat{\lambda} = \hat{\lambda}(\tau) \to \lambda^*$ for τ close to τ^* we have

$$\begin{aligned}
\nabla^2_{xx} H(\hat{x}, \tau) &\approx m(\tau - f(x^*))^{-1} \left[\nabla^2_{xx} L(x^*, \lambda^*) + \nabla c(x^*)^T \Lambda^* C^{-1}(\hat{x}) \nabla c(x^*) \right. \\
&\quad \left. + \ (\tau - f(x^*))^{-1} \nabla f(x^*) \nabla f(x^*)^T \right].
\end{aligned}$$

From the K-K-T condition

$$\nabla f(x^*) = \sum_{i=1}^r \lambda_i^* \nabla c_i(x^*),$$

follows

$$\forall\, u : \nabla c_{(r)}(x^*)\, u = 0 \Rightarrow (\nabla f(x^*), u) = 0\,,$$

where $\nabla c_{(r)}(x^*) = J(c_{(r)}(x^*))$ is $r \times n$ Jacobian of the vector - function $c_{(r)}(x) = (c_1(x), ..., c_r(x))^T$, which corresponds to the active constraints.

Hence, for $\forall\, u : \nabla c_{(r)}(x^*) u = 0$, we obtain

$$\left(\nabla^2_{xx} H(\hat{x}, \tau) u, u \right) \approx m(\tau - f(x^*))^{-1} \times$$

$$\left(\left(\nabla^2_{xx} L(x^*, \lambda^*) + \nabla c^T_{(r)}(x^*) \Lambda^*_{(r)} C^{-1}_{(r)}(\hat{x}) \nabla c_{(r)}(x^*) \right) u, u \right), \tag{3.8}$$

where

$$\Lambda^*_{(r)} = \text{diag}\, (\lambda_i^*)^r_{i=1},\ C_{(r)}(x) = \text{diag}\, (c_i(\hat{x}))^r_{i=1}\,.$$

From the second order sufficient condition follows $\lambda_i^* > 0$, $i = 1, ..., r$, also from $\hat{x} = \hat{x}(\tau) \to x^*$ follows $c_i(\hat{x}) \to 0$, $\ i = 1, \ldots, r$.

Therefore

$$\lim_{\tau\to f(x^*)} M_i(\tau) = \lim_{\tau\to f(x^*)} \hat{\lambda}_i(\tau)c_i^{-1}(\hat{x}(\tau)) = \infty\,, \quad i = 1,\ldots,r\,. \tag{3.9}$$

For $\tau_0 > \tau > \tau^* = f(x^*)$ close to τ^* from Debreu's lemma with $A = \nabla^2_{xx}L(x^*,\lambda^*)$ and $C = \Lambda^{*\frac{1}{2}}_{(r)}(C_{(r)}(\hat{x}))^{-\frac{1}{2}}\nabla c_{(r)}(x^*)$ follows existence of $\rho > 0$, such that

$$\mu(\tau) = \text{mineigenval } \nabla^2_{xx}H(\hat{x},\tau) \to \rho$$

when $\tau \to \tau^*$.

On the other hand, from (3.9) follows

$$M(\tau) = \text{maxeigenval } \nabla^2_{xx}H(\hat{x},\tau) \to \infty,$$

when $\tau \to \tau^*$.

Therefore

$$\text{cond}\nabla^2_{xx}H(\hat{x},\tau) = M(\tau)\mu^{-1}(\tau) \to \infty$$

when $\tau \to \tau^*$.

The ill-conditioning of the Hessian $\nabla^2_{xx}H(\hat{x},\tau)$ is much more critical in nonlinear optimization than in LP. In case of LP, the term $\nabla^2_{xx}L(x,\lambda)$ in the expression of the Hessian $\nabla^2_{xx}H(\hat{x},\tau)$ disappears and by rescaling the input data properly, one can, to some extent, eliminate the ill-conditioning effect.

In nonlinear optimization, the situation is completely different and the ill-conditioning is an important issue, in particular, when solution with high accuracy is required.

In the following section we introduce and study the EDF, which eliminates the basic drawbacks of the Classical IDF.

4. Exterior Distance Function

For a given $y \in \text{int}\,\Omega$ let us consider the following problem

$$F(x^*,y) = \min\{F(x,y)|c_i(x) \geq 0, i = 1,...,m\}, \tag{4.1}$$

where

$$F(x,y) = -\ln\Delta(x,y) = -\ln(f(y) - f(x)).$$

For any $y \in \text{int}\,\Omega$ the function F is convex and monotone decreasing together with f for $x \in \Omega(y) = \{x \in \Omega : f(x) \leq f(y)\}$, therefore the solution $x^* \in \Omega(y)$ of the problem (4.1) belongs to X^* and vice versa any $x^* \in X^*$ solves (4.1), that is problems (2.1) and (4.1) are equivalent.

In what is following we consider the problem (4.1) instead (2.1) .

The correspondent to (4.1) Lagrangian $L_y : \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}$ is given by

$$L_y(x,\lambda) = F(x,y) - \sum_{i=1}^m \lambda_i c_i(x). \tag{4.2}$$

The correspondent to (4.1) dual function $d_y : \mathbb{R}^m_+ \to \mathbb{R}$ is

$$d_y(\lambda) = \inf_{x\in\mathbb{R}} L_y(x,\lambda)$$

and

$$d_y(\lambda^*) = \max\{d_y(\lambda)|\lambda \in \mathbb{R}^m_+\} \tag{4.3}$$

is the dual to (4.1) problem.

Let $\psi(t) = \ln(t+1)$ and $k > 0$, then the original set of constraints

$$c_i(x) \geq 0, \quad i = 1, ..., m$$

is equivalent to the following set

$$k^{-1}\psi(kc_i(x)) = k^{-1}\ln(kc_i(x)+1) \geq 0, \; i = 1, ..., m, \tag{4.4}$$

therefore for any given $k > 0$ we have

$$\Omega = \{x \in \mathbb{R}^n : k^{-1}\psi(kc_i(x)) = k^{-1}\ln(kc_i(x)+1) \geq 0, \; i = 1, ..., m\},$$

and for any given $y \in \text{int}\,\Omega$ and $k > 0$ the problem

$$F(x^*, y) = \min\{F(x, y)|k^{-1}\ln(kc_i(x)+1) \geq 0, \; i = 1, .., m\} \tag{4.5}$$

is equivalent to (4.1).

Let us fix $y \in \text{int}\,\Omega$, then the following extension

$$\Omega_{-k^{-1}}(y) = \{x \in \mathbb{R}^n : c_i(x) \geq -k^{-1}, i = 1, ..., m, \; f(y) > f(x)\} \tag{4.6}$$

of $\Omega(y)$ is convex and bounded due to convexity f, concavity c_i, $i = 1, ..., m$ boundedness $\Omega(y)$ and Corollary 20 ([5]) and so is the following contruction of $\Omega(y)$

$$\Omega_\gamma(y) = \{x \in \mathbb{R} : c_i(x) \geq \gamma, \; i = 1, .., m, \;\; f(y) > f(x)\}. \tag{4.7}$$

The set $\Omega_\gamma(y)$ for small $\gamma > 0$ is not empty due to the Slater condition.

Let us fix $y \in \text{int}\,\Omega$ and $k > 0$, then Lagrangian $\mathcal{L}_y : \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}_{++} \to \mathbb{R}$

$$\mathcal{L}_y(x, \lambda, k) = F(x, y) - k^{-1}\sum_{i=1}^m \lambda_i \ln(kc_i(x)+1), \tag{4.8}$$

for problem (4.5) we call the exterior distance function (EDF).

Thus, EDF is a particular realization of the NR principle (see [22], [23], [25] and references therein).

Let us consider the second order sufficient optimality conditions for problem (4.1).

There exists $\mu > 0$, such that

$$(\nabla^2_{xx}L_y(x^*, \lambda^*)u, u) \geq \mu(u, u), \forall u : \nabla c_{(r)}(x^*)u = 0 \tag{4.9}$$

and

$$\text{rank}\; \nabla c_{(r)}(x^*) = r. \tag{4.10}$$

We conclude the section by pointing out some EDF properties at the KKT's point (x^*, λ^*).

First of all, $\mathcal{L}$ is convex in $x \in \Omega_{-k^{-1}}(y)$ for any given int Ω, $k > 0$ and $\lambda \in \mathbb{R}^m_+$.

Proposition 4.1. *For a given $y \in \text{int}\,\Omega$ and $k > 0$ and any KKT's point (x^*, λ^*) we have:*

1^0 $\mathcal{L}_y(x^*,\lambda^*,k) = F(x^*,y) = -\ln(f(y)-f(x^*))$
or

$$f(x^*) = f(y) - e^{-F(x^*,y)};$$

2^0 $\nabla_x \mathcal{L}_y(x^*,\lambda^*,k) = \Delta^{-1}(x^*,y)\nabla f(x^*) - \sum_{i=1}^m (kc_i(x^*)+1)^{-1}\lambda_i^*\nabla c_i(x^*) =$

$$\Delta^{-1}(x^*,y)\nabla f(x^*) - \sum_{i=1}^m \lambda_i^*\nabla c_i(x^*) = \nabla_x L_y(x^*,\lambda^*) = 0;$$

3^0 $\nabla^2_{xx}\mathcal{L}_y(x^*,\lambda^*,k) = \Delta^{-2}(x^*,y)\nabla f(x^*)\nabla f^T(x^*) + \Delta^{-1}(x^*,y)\nabla^2 f(x^*)$

$$-\sum_{i=1}^m \lambda_i^*\nabla^2 c_i(x^*) + k\nabla c(x^*)^T\Lambda^*\nabla c(x^*) = \nabla^2_{xx}L_y(x^*,\lambda^*) + k\nabla c(x^*)^T\Lambda^*\nabla c(x^*) =$$

$$\nabla^2_{xx}L_y(x^*,\lambda^*) + k\nabla c_{(r)}(x^*)^T\Lambda^*_{(r)}\nabla c_{(r)}(x^*),$$

where $\Lambda^*_{(r)} = diag(\lambda_i^*)_{i=1}^r$, $\lambda_i^* = 0$, $i = r+1,...,m.$

Properties $1^0 - 3^0$ follow from the definition of EDF (4.8) and complementarity condition

$$\lambda_i^* c_i(x^*) = 0, i = 1,...,m. \tag{4.11}$$

The fundamental difference between EDF (4.8) and the Huard's IDF (3.1) follows from $1^0 - 3^0$.

First, the $\mathcal{L}_y(x,\lambda,k)$ is defined at the solution together with its gradient and Hessian.

Second, from 2^0 follows that for any given $y \in \text{int}\,\Omega$ and $k > 0$ the optimal solution of (4.1) can be found by solving one smooth unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} \mathcal{L}_y(x,\lambda^*,k) = \mathcal{L}_y(x^*,\lambda^*,k). \tag{4.12}$$

It means that $\mathcal{L}_y(x,\lambda^*,k)$ is an exact smooth approximation for the following non-smooth problem

$$\min_{x\in\mathbb{R}^n} \max\{F(x,y) - F(x^*,y), -c_i(x),\ i = 1,...,m\}, \tag{4.13}$$

which is for any given $y \in \text{int}\,\Omega$ is equivalent to (4.1).

Third, from 3^0 for any $u \in \mathbb{R}^n$ follows

$$(\nabla^2_{xx}\mathcal{L}_y(x^*,\lambda^*,k)u,u) = ((\nabla^2_{xx}L_y(x^*,\lambda^*) + k\nabla c_{(r)}(x^*)^T\Lambda^*_{(r)}\nabla c_{(r)}(x^*))u,u).$$

Proposition 4.2. *Under the second order sufficient optimality condition (4.9)-(4.10), for any given* $y \in \text{int}\,\Omega$, $k_0 > 0$ *large enough and any* $k \geq k_0$ *there exists* $0 < \rho < \mu$ *such that*

$$(\nabla^2_{xx}\mathcal{L}_y(x^*,\lambda^*,k)u,u) \geq \rho(u,u),\ \forall u \in \mathbb{R}^n.$$

Proposition 4.2 follows from the second order sufficient optimality condition (4.9)-(4.10) and Debreu's Lemma with

$$A = \nabla^2_{xx}L_y(x^*,\lambda^*)\ ,\ C = \Lambda^{*\frac{1}{2}}_{(r)}\nabla c_{(r)}(x^*).$$

In other words, for any fixed $y \in \text{int}\,\Omega$ as a "center" and any $k \geq k_0$ the EDF $\mathcal{L}_y(x, \lambda^*, k)$ is strongly convex in the neighborhood of x^* no matter if f and $-c_i$, $i = 1, ..., m$ are convex or not.

The EDF is related to the Classical Huard's interior distance function $H(x, \tau)$ as MBF (see [22]) is to the classical R. Frisch's barrier function $F(x, k) = f(x) - k^{-1} \sum_{i=1}^{m} \ln c_i(x)$ (see [6]).

It is worth mentioning that relatively to the MBF features the EDF has one extra tool-the "center", which we will use later to improve convergence properties.

The EDF properties lead to a new multipliers method, which converges under any fixed $y \in \text{int}\ \Omega$ as a "center" and any fixed scaling parameter $k > 0$, just due to the Lagrange multipliers update. This is a fundamental departure from the Classical IDF theory (see [4],[5],[10],[11],[12],[20]).

5. Exterior Point Method

The EPM at each step finds the primal minimizer of $\mathcal{L}_y$ following by Lagrange multipliers updates.

We start with $y \in \text{int}\ \Omega$ as a given fixed "center", fixed scaling parameter $k > 0$ and initial Lagrange multipliers vector $\lambda_0 = e = (1, ..., 1)^T \in \mathbb{R}^m_{++}$.

Let the primal-dual approximation (x_s, λ_s) has been found already.

The approximation (x_{s+1}, λ_{s+1}) we find by the following operations

$$x_{s+1} : \nabla_x \mathcal{L}_y(x_{s+1}, \lambda_s, k)$$

$$= \Delta^{-1}(x_{s+1}, y)\nabla f(x_{s+1}) - \sum_{i=1}^{m} \lambda_{i,s}\psi^{'}(kc_i(x_{s+1}))\nabla c_i(x_{s+1}) = 0 \tag{5.1}$$

$$\lambda_{s+1} : \lambda_{i,s+1} = \lambda_{i,s}\psi^{'}(kc_i(x_{s+1})) = \lambda_{i,s}(kc_i(x_{s+1})+1)^{-1},\ i = 1, ..., m. \tag{5.2}$$

The key ingredient of the EPM (5.1)-(5.2) convergence analysis is its equivalence to the proximal point method with φ-divergence distance function for the dual problem

Theorem 5.1. *If condition A and B hold, f, $c_i \in C^1$, $i = 1, ..., m$, f is convex and all c_i, $i = 1, ..., m$ are concave, then EPM (5.1)-(5.2) is:*

1) *well defined;*
2) *equivalent to the following proximal point method*

$$d_y(\hat{\lambda}) - k^{-1}D(\hat{\lambda}, \lambda) = \max\{d_y(u) - k^{-1}D(u, \lambda) | u \in \mathbb{R}^m_+\}, \tag{5.3}$$

where

$$D(u, \lambda) = \sum_{i=1}^{m} \lambda_i \varphi(u_i/\lambda_i)$$

is φ-divergence distance function based on the kernel $\varphi = -\psi^*$, where ψ^* is Legendre transform of ψ

Proof.

1) Due to convexity f, concavity c_i, $i = 1, ..., m$, Slater condition, boundedness $\Omega(y)$ and properties of log-barrier function the recession cone of $\Omega(y)$ is empty, that is we have

$$\lim_{t\to\infty} \mathcal{L}_y(x + td, \lambda, k) = \infty$$

for any $d \neq 0$ from $\mathbb{R}^n$, $y \in \operatorname{int}\Omega$, $k > 0$ and $\lambda \in \mathbb{R}^m_{++}$.

Hence, there exists $x_{s+1} \in \mathbb{R}^n$:

$$\mathcal{L}_y(x_{s+1}, \lambda_s, k) = \min\{\mathcal{L}_y(x, \lambda_s, k)|x \in \mathbb{R}^n\},$$

thus (5.1) holds.

From $\ln t = -\infty$ for $t \leq 0$ and (5.1) follows $kc_i(x_{s+1}) + 1 > 0$, $i = 1, ..., m$, therefore from (5.2) we have

$$\lambda_s \in \mathbb{R}^m_{++} \Rightarrow \lambda_{s+1} \in \mathbb{R}^m_{++}.$$

Hence, method (5.1)-(5.2) is well defined.

2) From (5.1) and (5.2) follows

$$\nabla_x\mathcal{L}_y(x_{s+1}, \lambda_s, k) = \Delta^{-1}(x_{s+1}, y)\nabla f(x_{s+1}) - \sum_{i=1}^{m} \lambda_{i,s+1}\nabla c_i(x_{s+1}) =$$

$$\nabla_x L_y(x_{s+1}, \lambda_{s+1}) = 0.$$

Therefore

$$\min_{x\in\mathbb{R}^n} L_y(x, \lambda_{s+1}) = L_y(x_{s+1}, \lambda_{s+1}) = d_y(\lambda_{s+1}).$$

The subdifferential $\partial d_y(\lambda_{s+1})$ contains $-c(x_{s+1})$, that is

$$0 \in c(x_{s+1}) + \partial d_y(\lambda_{s+1}). \tag{5.4}$$

From (5.2) we have

$$\psi'(kc_i(x_{s+1})) = \lambda_{i,s+1}/\lambda_{i,s},\ i = 1, ..., m.$$

Also $\psi''(kc_i(x_{s+1})) \neq 0$, therefore the inverse function ψ'^{-1} exists and

$$c_i(x_{s+1}) = k^{-1}\psi'^{-1}(\lambda_{i,s+1}/\lambda_{i,s}). \tag{5.5}$$

From (5.5) and Legendre identity $\psi'^{-1} \equiv \psi^{*'}$ follows

$$c_i(x_{s+1}) = k^{-1}\psi^{*'}(\lambda_{i,s+1}/\lambda_{i,s}),\ i = 1, ..., m. \tag{5.6}$$

From (5.4) and (5.6) we obtain

$$0 \in \partial d_y(\lambda_{s+1}) + k^{-1}\sum_{i=1}^{m} \psi^{*'}(\lambda_{i,s+1}/\lambda_{i,s})e_i, \tag{5.7}$$

where $e_i = (0, ..., 1, ..., 0)^T \in \mathbb{R}^m_+$.

Let $\varphi = -\psi^*$, then (5.7) we can rewrite as follows

$$0 \in \partial d_y(\lambda_{s+1}) - k^{-1} \sum_{i=1}^{m} \varphi^{'}(\lambda_{i,s+1}/\lambda_{i,s}) e_i,$$

which is the optimality condition for λ_{s+1} to be the solution in (5.3) with $\lambda = \lambda_s$, that is

$$d_y(\lambda_{s+1}) - k^{-1} \sum_{i=1}^{m} \lambda_{i,s} \varphi(\lambda_{i,s+1}/\lambda_{i,s}) =$$

$$\max\{d_y(u) - k^{-1} \sum_{i=1}^{m} \lambda_{i,s} \varphi(u_i/\lambda_{i,s}) | u \in \mathbb{R}^m_{++}\} \tag{5.8}$$

$$= \max\{d_y(u) - k^{-1} D(u, \lambda_s) | u \in \mathbb{R}^m_{++}\},$$

where $\psi^*(s) = \ln s - s + 1$.

Therefore for the kernel $\varphi(s) = -\psi^*(s)$ of the φ- divergence distance function

$$D(u, \lambda) = \sum_{i=1}^{m} \lambda_i \varphi(u_i/\lambda_i) = \sum_{i=1}^{m} [-\lambda_i \ln u_i/\lambda_i + u_i - \lambda_i]$$

we have:

$$\begin{aligned} &a) \quad \varphi(s) = -\ln s + s - 1 \geq 0, \quad \forall s > 0 \\ &b) \quad \min_{s>0} \varphi(s) = \varphi(1) = \varphi^{'}(1) = 0. \end{aligned} \tag{5.9}$$

In fact, $D(u, \lambda)$ is the Kullback-Leibler φ- divergence distance function (see, for example, [25]). The proof of Theorem 5.1 is completed.

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

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

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

For compact sets X and Y we have

$$d_H(X, Y) = 0 \Leftrightarrow X = Y. \tag{5.10}$$

Let $Q \subset \mathbb{R}^m_+$ be a compact set, $\hat{Q} = \mathbb{R}^m_+ \backslash Q$, $S(u, \varepsilon) = \{v \in \mathbb{R}^m_+ : \|u - v\| \leq \varepsilon\}$ and

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

be the boundary of Q.

For convex and compact sets $A \subset B \subset C$ the inequality

$$d_H(A, \partial B) < d_H(A, \partial C) \tag{5.11}$$

follows from the definition of Hausdorff distance.

For the dual sequence $\{\lambda_s\}_{s=0}^{\infty}$ we consider the dual level sets $\Lambda_s = \{\lambda \in \mathbb{R}^m_+ : d(\lambda) \geq d(\lambda_s)\}$, which are convex, due to concavity d, and bounded, due to the boundedness L^*,which is, in turn, a consequence of Slater's condition. Let $\partial\Lambda_s = \{\lambda \in \Lambda_s : d(\lambda) = d(\lambda_s)\}$ be the boundary of Λ_s.

6. Convergence of the EPM

The following Theorem establishes convergence of the EPM under minimum assumptions on the input data, just due to the Lagrange multipliers update.

Theorem 6.1. *Under assumptions of Theorem 5.1 for any fixed $y \in \text{int}\,\Omega$, as a "center", any scaling parameter $k > 0$ and any $\lambda_0 \in \mathbb{R}^m_{++}$ the EPM* (5.1)-(5.2) *generates primal-dual sequence $\{x_s, \lambda_s\}_{s+0}^{\infty}$ that:*

1) $d_y(\lambda_{s+1}) > d_y(\lambda_s),\ s \geq 0$
2) $\lim_{s\to\infty} d_y(\lambda_s) = d_y(\lambda^*),\ \lim_{s\to\infty} F(y, x_s) = F(y, x^*)$
3) $\lim_{s\to\infty} d_H(\partial\Lambda_s, L^*) = 0$
4) *there exists a subsequence $\{s_l\}_{l=1}^{\infty}$ such that for $\bar{x}_l = \sum_{s=s_l}^{s_{l+1}} (s_{l+1} - s_l)^{-1} x_s$ we have $\lim_{l\to\infty} \bar{x}_l = \bar{x} \in X^*$, that is the primal sequence converges to the primal solution in the ergodic sence.*

Proof.

1) From $\varphi(1) = 0$ and (5.8) with $u = \lambda_s$ follows

$$d_y(\lambda_{s+1}) \geq d_y(\lambda_s) + k^{-1} \sum_{i=1}^{m} \lambda_{i,s} \varphi(\lambda_{i,s+1}/\lambda_{i,s}) \tag{6.1}$$

$$= d_y(\lambda_s) + k^{-1} D(\lambda_{s+1}, \lambda_s).$$

From $\varphi(t) \geq 0$, $\forall t > 0$, (6.1) and $\lambda_s \in \mathbb{R}^m_{++}$ follows

$$d_y(\lambda_{s+1}) \geq d_y(\lambda_s). \tag{6.2}$$

Moreover, $d_y(\lambda_{s+1}) > d_y(\lambda_s)$ unless $\varphi(\lambda_{i,s+1}/\lambda_{i,s}) = 0$ for all $i = 1, ..., m$, which leads to $\lambda_{s+1} = \lambda_s = \lambda^*$.

2) The monotone increasing sequence $\{d_y(\lambda_s)\}_{s=0}^{\infty}$ is bounded from above by the optimal value of the primal objective function $F(y, x^*) = -\ln(f(y) - f(x^*))$, therefore there exists $\lim_{s\to\infty} d_y(\lambda_s) = \bar{d} \leq F(y, x^*)$.

Our next step is to show $\bar{d} = F(y, x^*)$.

From $-c(x_{s+1}) \in \partial d_y(\lambda_{s+1})$ and concavity of the dual function d_y follows

$$d_y(\lambda) - d_y(\lambda_{s+1}) \leq (-c(x_{s+1}), \lambda - \lambda_{s+1}),\ \forall \lambda \in \mathbb{R}^m_{++}.$$

For $\lambda = \lambda_s$ we obtain

$$d_y(\lambda_{s+1}) - d_y(\lambda_s) \geq (c(x_{s+1}), \lambda_s - \lambda_{s+1}). \tag{6.3}$$

From the update formula (5.2) we have

$$\lambda_{i,s} - \lambda_{i,s+1} = k c_i(x_{s+1}) \lambda_{i,s+1},\ i = 1, ..., m. \tag{6.4}$$

Therefore from (6.3) and (6.4) follows

$$d_y(\lambda_{s+1}) - d_y(\lambda_s) \geq k \sum_{i=1}^{m} c_i^2(x_{s+1})\lambda_{i,s+1}. \tag{6.5}$$

From boundedness of L^* and concavity d_y follows boundedness of the initial dual level set

$$\Lambda_0 = \{\lambda \in \mathbb{R}^m_+ : d_y(\lambda) \geq d_y(\lambda_0)\}.$$

From the dual monotonicity (6.2) and boundedness Λ_0 follows boundedness of the dual sequence $\{\lambda_s\}_{s=0}^{\infty} \subset \Lambda_0$.

Therefore there exists $L = \max_{i,s} \lambda_{i,s}$ and from (6.5) follows

$$d_y(\lambda_{s+1}) - d_y(\lambda_s) \geq kL^{-1}(c(x_{s+1}), \lambda_{s+1})^2. \tag{6.6}$$

Summing up (6.6) from $s = 1$ to $s = N$ we obtain

$$d_y(\lambda^*) - d_y(\lambda_0) \geq d_y(\lambda_{N+1}) - d_y(\lambda_0) \geq kL^{-1} \sum_{s+1}^{N} (\lambda_s, c(x_s))^2.$$

It leads to the asymptotic complementarity condition

$$\lim_{s \to \infty} (\lambda_s, c(x_s)) = 0. \tag{6.7}$$

Summing up (6.1) from $s = 0$ to $s = N$ we obtain

$$d_y(\lambda^*) - d_y(\lambda_0) \geq d_y(\lambda_N) - d_y(\lambda_0) \geq k^{-1} \sum_{s+1}^{N} D(\lambda_{s+1}, \lambda_s),$$

therefore $\lim_{s\to\infty} D(\lambda_{s+1}, \lambda_s) = 0$.

The diminishing divergence between two sequential Lagrange multipliers vectors leads us to believe that under any given $y \in \operatorname{int} \Omega$ as a "center" and any given scaling parameter $k > 0$ the map

$$\lambda \to \hat{\lambda}(\lambda, k) = (kC(\hat{x}) + I^m)^{-1}\lambda,$$

has a fixed point λ^*, where $C(\hat{x}) = \operatorname{diag}(c_i(\hat{x}))_{i=1}^m$, I^m- identical matrix in $\mathbb{R}^m$ and

$$\hat{x}(\lambda, k) \equiv \hat{x} : \nabla_x \mathcal{L}_y(\hat{x}, \lambda, k) = 0.$$

First, let us show that for any $\lambda^* \in \Lambda^*$ the sequence $\{D(\lambda_s, \lambda^*)\}_{s=0}^{\infty}$ is monotone decreasing.

We assume $x \ln x = 0$ for $x = 0$, then

$$D(\lambda_s, \lambda^*) - D(\lambda_{s+1}, \lambda^*) = \sum_{i=1}^{m} (\lambda_i^* \ln \lambda_{i,s+1}/\lambda_{i,s} + \lambda_{i,s} - \lambda_{i,s+1}).$$

Using the update formula (5.2) we obtain

$$D(\lambda_s, \lambda^*) - D(\lambda_{s+1}, \lambda^*) = \tag{6.8}$$

$$\sum_{i=1}^{m} \lambda_i^* \ln(kc_i(x_{s+1}) + 1)^{-1} + k \sum_{i=1}^{m} \lambda_{i,s+1} c_i(x_{s+1}).$$

From $-\ln(1+t) \geq -t,\ \forall t > -1$ and (6.8) follows

$$D(\lambda_s, \lambda^*) - D(\lambda_{s+1}, \lambda^*) \geq \tag{6.9}$$

$$k \sum_{i=1}^{m} (\lambda_{i,s+1} - \lambda_{i,s}^*) c_i(x_{s+1}) = k(-c(x_{s+1}), \lambda^* - \lambda_{s+1}).$$

From concavity d and $-c(x_{s+1}) \in \partial d_y(\lambda_{s+1})$ we obtain

$$0 \leq d_y(\lambda^*) - d_y(\lambda_{s+1}) \leq (-c(x_{s+1}), \lambda^* - \lambda_{s+1}). \tag{6.10}$$

From (6.9) and (6.10) follows

$$D(\lambda_s, \lambda^*) - D(\lambda_{s+1}, \lambda^*) \geq k(d_y(\lambda^*) - d_y(\lambda_{s+1})) \geq 0. \tag{6.11}$$

If $\lim_{s\to\infty} d_y(\lambda_s) = \bar{d} < d_y(\lambda^*) = F(y, x^*)$, then there is $\sigma > 0$ and s_0 that from (6.11) we have

$$D(\lambda_s, \lambda^*) - D(\lambda_{s+1}, \lambda^*) \geq k\sigma, \forall s \geq s_0.$$

Summing up the last inequalities from $s = s_0$ to $s = N$ we obtain

$$D(\lambda_0, \lambda^*) - D(\lambda_{N+1}, \lambda^*) =$$

$$\sum_{i=1}^{m} (\lambda_i^* \ln \lambda_{i,N+1}/\lambda_{i,s_0} + \lambda_{i,s_0} - \lambda_{i,N+1}) \geq k(N - s_0)\sigma,$$

which is impossible for large N due to the boundedness of $\{\lambda_s\}_{s=0}^{\infty} \in \Lambda_0$.

Therefore

$$d_y(\lambda^*) = \lim d_y(\lambda_s) = \lim_{s\to\infty} [F(y, x_s) - (\lambda_s, c(x_s))].$$

Keeping in mind asymptotic complementarity (6.7) we obtain

$$\lim_{s\to\infty} F(y, x_s) = F(y, x^*) = d_y(\lambda^*). \tag{6.12}$$

3) From boundedness of the dual sequence follows existence of a subsequence $\{\lambda_{s_i}\}_{i=1}^{\infty} \subset \{\lambda_s\}_{s=0}^{\infty}$:$\lim_{s_i\to\infty} \lambda_{s_i} = \bar{\lambda}$. From convergence of the dual sequence in value follows $\bar{\lambda} = \lambda^*$ and $L^* = \{\lambda \in \mathbb{R}_+^m : d_y(\lambda) = d_y(\bar{\lambda})\}$.

   From dual strong monotonicity: $d_y(\lambda_{s+1}) > d_y(\lambda_s)$ follows

$$L^* \subset ...\Lambda_{s+1} \subset \Lambda_s ... \subset \Lambda_0,$$

   therefore from (5.11) follows that $\{d_H(\partial\Lambda_s, L^*)\}_{s=0}^{\infty}$, is a monotone decreasing sequence of positive numbers. It has a limit, that is

$$\lim_{s\to\infty} d_H(\partial\Lambda_s, L^*) = \nu \geq 0,$$

   but $\nu > 0$ is impossible due to the dual convergence in value (6.12).

4) The ergodic convergence of the primal sequence one can prove by repeating the arguments used in the proof of item 4. Theorem 8 in [25]. The proof of Theorem 6.1 is completed.

So far, neither the fixed "center" $y \in \operatorname{int}\Omega$ nor the fixed scaling parameter $k > 0$ contributed to improvement of the EPM convergence.

In the following section we establish Q-linear convergence rate of the EPM under standard second order sufficient optimality condition (4.9)-(4.10), any $y \in \operatorname{int}\Omega$ as fixed "center" and a fixed, but large enough scaling parameter $k > 0$.

7. Convergence rate of the EPM

Let us first describe the dual domain, where the basic results are taking place.

We assume that $0 < \delta < \min_{1\le i\le r} \lambda_i^*$ is small enough and $k_0 > 0$ is large enough.

In the course of proving the Theorem it will be more clear quantitatively what "small" and "large" means.

We split the extended dual set into active and passive sub-sets, that is

$$\Lambda(\cdot) \equiv \Lambda(\lambda, k, \delta) = \Lambda_{(r)}(\cdot) \otimes \Lambda_{(m-r)}(\cdot),$$

where

$$\Lambda_{(r)}(\cdot) \equiv \Lambda_{(r)}(\lambda_{(r)}, k, \delta) = \{(\lambda_{(r)}, k, \delta) : \lambda_i \ge \delta,\ |\lambda_i - \lambda_i^*| \le \delta k, i = 1, ..., r, k \ge k_0\}$$

be the active dual sub-set and

$$\Lambda_{(m-r)}(\cdot) \equiv \Lambda_{(m-r)}(\lambda_{(m-r)}, k, \delta)$$
$$= \{(\lambda_{(m-r)}, k, \delta) : 0 \le \lambda_i \le \delta k,\ i = r+1, ..., m,\ k \ge k_0\}$$

be the passive dual sub-set. For a vector $a \in \mathbb{R}^n$ we use the following norm $\|a\| = \max_{1\le i\le n} |a_i|$. For a matrix $A \in \mathbb{R}^{m\times n}$ the correspondent norm is $\|A\| = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|$.

Theorem 7.1. *If f, $c_i \in C^2$, $i = 1, ..., m$ and the second order sufficient optimality condition (4.9)-(4.10) is satisfied, then exist a small enough $\delta > 0$ and large enough $k_0 > 0$, that for any $y \in \operatorname{int}\Omega$ as a fixed "center" and any $(\lambda, k) \in \Lambda(\cdot)$ the following statements hold true:*

1) *there exists*

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

and

$$\hat{\lambda} = (\hat{\lambda}_i = \lambda_i (k c_i(\hat{x}) + 1)^{-1},\ i = 1, ..., m).$$

2) *for $(\hat{x}, \hat{\lambda})$ the following bound holds*

$$\max\{\|\hat{x} - x^*\|, \|\hat{\lambda} - \lambda^*\|\} \le c k^{-1} \|\lambda - \lambda^*\|, \tag{7.1}$$

where $c > 0$ is independent on $k \ge k_0$. Also $x(\lambda^*, k) = x^*$ and $\hat{\lambda}(\lambda^*, k) = \lambda^*$, that is λ^* is a fixed point of the map $\lambda \to \hat{\lambda}(\lambda, k)$.

3) *The EDF $\mathcal{L}_y(x, \lambda, k)$ is strongly convex in the neighborhood of $\hat{x}$.*

Proof. Vector $\hat{x} = \hat{x}(\lambda, k)$ also depends on $y \in \text{int } \Omega$, therefore $\hat{\lambda}$ is a function of λ, k and $y \in \text{int } \Omega$. At this point y is fixed, so to simplify notation we omit y in the definition of $\hat{x}$ and $\hat{\lambda}$.

By introducing vector $t = (t_1, ..., t_r, t_{r+1}, ..., t_m)^T$ with $t_i = (\lambda_i - \lambda_i^*)k^{-1}$ we transform the dual set $\Lambda(\cdot)$ into the following neighborhood of the origin of the extended dual space

$$S(0, k, \delta) = S_{(r)}(0, k, \delta) \otimes S_{(m-r)}(0, k, \delta),$$

where

$$S_{(r)}(0, k, \delta) = \{(t_{(r)}, k) : |t_i| \leq \delta,\ t_i \geq (\delta - \lambda_i^*)k^{-1},\ i = 1, ..., r,\ k \geq k_0\}$$

and

$$S_{(m-r)}(0, k, \delta) = \{(t_{(m-r)}, k) : 0 \leq t_i \leq \delta, i = r+1, ..., m, k \geq k_0\}.$$

Let us consider vector-function $h : \mathbb{R}^{n+m-r+1} \to \mathbb{R}^n$ defined as follows

$$h(x, t_{(m-r)}, k) = k \sum_{i=r+1}^{m} t_i (kc_i(x) + 1)^{-1} \nabla c_i(x),$$

then

$$\nabla_t h(x, t_{(m-r)}, k) = [0^{n,r} \quad k \nabla c_{(m-r)}^T(x) \Psi'(kc_{(m-r)}(x)],$$

where $\Psi'(kc_{(m-r)}(x)) = \text{diag}[(kc_i(x) + 1)^{-1}]_{i=r+1}^m$.

$$\nabla_x h(x, t_{(m-r)}, k) = k^2 \sum_{i=r+1}^{m} t_i (kc_i(x) + 1)^{-2} \nabla c_i^T(x) \nabla c_i(x).$$

Therefore for any $k > 0$ we have

$$h(x^*, 0^{m-r}, k) = 0^n,\ \nabla_x h(x^*, 0^{m-r}, k) = 0^{n \times n}.$$

Our main tool is the map $\Phi_y : \mathbb{R}^{n+m+r+1} \to \mathbb{R}^{n+r}$ given by the following formula

$$\Phi_y(x, \hat{\lambda}_{(r)}, t, k) = \begin{pmatrix} \Delta^{-1}(y, x) \nabla f(x) - \sum_{i=1}^r \hat{\lambda}_i \nabla c_i(x) - h(x, t_{(m-r)}, k) \\ (t_i + k^{-1}\lambda_i^*)(kc_i(x) + 1)^{-1} - k^{-1}\hat{\lambda}_i,\ i = 1, ..., r \end{pmatrix}.$$

For a given fixed $y \in \text{int } \Omega$ we have

$$\Phi_y(x^*, \lambda_{(r)}^*, 0^m, k) = \begin{bmatrix} \Delta^{-1}(y, x^*) \nabla f(x^*) - \sum_{i=1}^r \lambda_i^* \nabla c_i(x^*) - h(x^*, 0^{m-r}, k) \\ k^{-1}(\lambda_i^* - \lambda_i^*),\ i = 1, ..., r \end{bmatrix}$$

$$= \begin{bmatrix} \nabla_x L_y(x^*, \lambda^*) \\ k^{-1}(\lambda_i^* - \lambda_i^*),\ i = 1, ..., r \end{bmatrix} = \begin{bmatrix} 0^n \\ 0^r \end{bmatrix}.$$

Let us consider the following Jacobian

$$\nabla_{x\hat{\lambda}_{(r)}} \Phi_y(x, \hat{\lambda}_{(r)}, t, k) = \nabla_{x\hat{\lambda}_{(r)}} \Phi_y(\cdot) =$$

$$= \begin{bmatrix} \nabla_{xx}^2 L_y(\cdot) & -\nabla c_{(r)}^T(\cdot) \\ -(T^r + k\Lambda_{(r)}^*)(kC_{(r)}(\cdot) + I^r)^{-2} \nabla c_{(r)}(\cdot) & -k^{-1} I^r \end{bmatrix},$$

where $\Lambda_{(r)} = \text{diag}(\lambda_i)_{i=1}^r$, $C_{(r)}(\cdot) = \text{diag}(c_i(\cdot))_{i=1}^r$, $T^r = \text{diag}(t_i)_{i=1}^r$ I^r - identical matrix in $\mathbb{R}^r$. For $x = x^*$, $\lambda_{(r)} = \lambda^*_{(r)}$ and $t = 0^m$ we have

$$\nabla_{x\hat{\lambda}_{(r)}} \Phi_y(x^*, \lambda^*_{(r)}, 0^m, k) = \begin{bmatrix} \nabla^2_{xx} L_y(x^*, \lambda^*) & -\nabla c^T_{(r)}(x^*) \\ -\Lambda^*_{(r)} \nabla c_{(r)}(x^*) & -k^{-1} I^r \end{bmatrix} \equiv$$

$$\equiv \begin{bmatrix} \nabla^2_{xx} L_y & -\nabla c^T_{(r)} \\ -\Lambda^*_{(r)} \nabla c_{(r)} & -k^{-1} I^r \end{bmatrix} \equiv \nabla \Phi_{(y,k)}.$$

The next step is to show that the matrix $\nabla \Phi_{(y,k)}$ is not singular for any given $y \in \text{int}\,\Omega$ and $k \geq k_0$, where $k_0 > 0$ is large enough.

Let $w = (u, v) \in \mathbb{R}^{n+r}$, then from

$$\nabla \Phi_{(y,k)} w = \begin{bmatrix} \nabla^2_{xx} L_y u & -\nabla c^T_{(r)} v \\ -\Lambda^*_{(r)} \nabla c_{(r)} u & -k^{-1} v \end{bmatrix} = \begin{bmatrix} 0^n \\ 0^r \end{bmatrix},$$

follows $v = -k \Lambda^*_{(r)} \nabla c_{(r)} u$ and

$$Nu = (\nabla^2_{xx} L_y + k \nabla c^T_{(r)} \Lambda^*_{(r)} \nabla c_{(r)}) u = 0,$$

therefore

$$(Nu, u) = ((\nabla^2_{xx} L_y + k \nabla c^T_{(r)} \Lambda^*_{(r)} \nabla c_{(r)}) u, u) = 0.$$

From sufficient optimality condition (4.9)-(4.10) and Debreu's lemma with $A = \nabla^2_{xx} L_y$ and $C = \Lambda^{*\frac{1}{2}} \nabla c_{(r)}$ follows the existence $0 < \rho < \mu$ that

$$0 = (Nu, u) \geq \rho(u, u),$$

hence $u = 0^n$, then from

$$\nabla^2_{xx} L_y u - \nabla c^T_{(r)} v = 0^n$$

and (4.10) follows $v = 0^r$. It means

$$\nabla \Phi_{(y,k)} w = 0^{n+r} \Rightarrow w = 0^{n+r},$$

therefore $\nabla \Phi^{-1}_{(y,k)}$ exists. Using argument similar to those in Theorem 1 from [22] one can prove existence of large enough $k_0 > 0$, that for any $k \geq k_0$ there exists $\rho_0 > 0$ independent on $k \geq k_0$ and $y \in \text{int}\,\Omega$ that

$$\|\Phi^{-1}_{(y,k)}\| \leq \rho_0. \tag{7.2}$$

Let $\infty > k_1 > k_0$, $k_0 > 0$ be large enough and $K = \{0^n\} \times [k_0, k_1]$. We consider the following neighborhood

$$S(K, \delta) = \{(t, k) : |t_i| \leq \delta,\ t_i \geq (\delta - \lambda^*_i) k^{-1}, i = 1, ..., r;$$
$$0 \leq t_i \leq \delta, i = r+1, ..., m\}$$

of K.

From the second implicit function Theorem (see, for example, [2] p.12) follows that for any $k \in [k_0, k_1]$ the system

$$\Phi_y(x, \hat{\lambda}_{(r)}, t, k) = 0^{m+r}$$

defines on $S(K,\delta)$ a unique pair of vectors

$$x(t,k) = (x_i(t,k),\ i = 1,...,n) \text{ and } \hat{\lambda}_{(r)}(t,k) = (\hat{\lambda}_i(t,k), i = 1,...,r)$$

that $x(0^m,k) = x^*$, $\hat{\lambda}_{(r)}(0^m,k) = \lambda^*_{(r)}$ and

$$\Phi_y(x(t,k), \hat{\lambda}_{(r)}(t,k), t, k) \equiv 0^{n+r}, \forall (t,k) \in S(K,\delta). \tag{7.3}$$

Identity (7.3) can be rewritten as follows

$$\Delta^{-1}(y, x(t,k))\nabla f(x(t,k)) - \sum_{i=1}^{r} \hat{\lambda}_i(t,k)\nabla c_i(x(t,k)) \\ -h(x(t,k), t_{(m-r)}, k) \equiv 0^n. \tag{7.4}$$

$$\hat{\lambda}_i(t,k) \equiv (kt_i + \lambda^*_i)\psi^{'}(kc_i(x(t,k))),\ i = 1,...,r, \tag{7.5}$$

also

$$\hat{\lambda}_i(t,k) \equiv kt_i\psi^{'}(kc_i(x(t,k))),\ i = r+1,...,m. \tag{7.6}$$

From (7.4)-(7.6) follows

$$\nabla_x \mathcal{L}_y(x(t,k), \lambda, k) = \nabla_x L_y(x(t,k), \hat{\lambda}(t,k)) \equiv 0^n,$$

where $\hat{\lambda}(t,k) = (\hat{\lambda}_{(r)}(t,k), \hat{\lambda}_{(m-r)}(t,k))^T$.

It completes the proof of item 1).

2) For a given small enough $\varepsilon > 0$ there is $\delta > 0$ such that

$$\|x(t,k) - x(0^m,k)\| = \|x(t,k) - x^*\| \leq \varepsilon \text{ for } \forall (t,k) \in S(K,\delta).$$

Hence, there is $\sigma = \min_{r+1\leq i\leq m}\{c_i(x^*)\}$ that for $r+1 \leq i \leq m$ we have

$$c_i(x(t,k)) \geq 0.5\sigma, \quad \forall (t,k) \in S(K,\delta).$$

From (5.2) follows

$$\hat{\lambda}_i = \lambda_i(kc(x(t,k)+1))^{-1} \leq 2(\sigma k)^{-1}\lambda_i,\ i = r+1,...,m,$$

where $\sigma > 0$ is independent on $k \geq k_0$.

To prove the bound (7.1) for $x(t,k)$ and $\hat{\lambda}_{(r)}(t,k)$ we will first estimate the norms $\|\nabla_t x(t,k)\|$, $\|\nabla_t \nabla \hat{\lambda}_{(r)}(t,k)\|$ at $t = 0^m$.

By differentiating identities (7.4) and (7.5) in t we obtain the following system for Jacobians $\nabla_t x(\cdot)$ and $\nabla_t \hat{\lambda}_{(r)}(\cdot)$

$$\nabla^2_{xx} L_y(x(\cdot), \hat{\lambda}_{(r)}(\cdot))\nabla_t x(\cdot) - \nabla c^T(\cdot)\nabla_t \hat{\lambda}_{(r)}(\cdot) \\ \equiv \nabla_t h(x(t,k), t_{(m-r)}, k) \\ = \nabla_x h(x(\cdot), \cdot)\nabla_t(x(\cdot)) + \nabla_t h(x(\cdot), \cdot) \tag{7.7}$$

$$k(kT^r + \Lambda^*_{(r)})\Psi^{''}(kc_{(r)}(x(\cdot)))\nabla c_{(r)}(x(\cdot))\nabla_t(x(\cdot)) - \nabla_t \hat{\lambda}_{(r)}(\cdot) \\ \equiv [k\Psi^{'}(kc_{(r)}(x(\cdot))); 0^{r,m-r}], \tag{7.8}$$

where $\Psi^{'}(kc_{(r)}(x(\cdot))) = \mathrm{diag}(\psi^{'}(kc_i(x(\cdot))))^r_{i=1}$, $\Psi^{''}(kc_{(r)}(x(\cdot))) = \mathrm{diag}(\psi^{''}(kc_i(x(\cdot))))^r_{i=1}$ are diagonal matrices.

In other words, for Jacobians $\nabla_t x(\cdot)$ and $\nabla_t \hat{\lambda}_{(r)}(\cdot)$ we have the following system

$$\nabla \Phi_{(y,k)}(\cdot) \begin{bmatrix} \nabla_t x(\cdot) \\ \nabla_t \hat{\lambda}(\cdot) \end{bmatrix} = R(\cdot), \tag{7.9}$$

where

$$\nabla \Phi_{(y,k)}(\cdot) = \begin{bmatrix} \nabla^2_{xx} L_y(x(\cdot), \hat{\lambda}_{(r)}(\cdot)) & -\nabla c^T_{(r)}(x(\cdot)) \\ (kT^r + \Lambda^*_{(r)})\Psi^{''}(kc_{(r)}(x(\cdot)))\nabla c_{(r)}(x(\cdot)) & -k^{-1}I^r \end{bmatrix}$$

and

$$R(x(t,k);t,k) = R(x(\cdot),\cdot) = \begin{bmatrix} \nabla_x h(x(\cdot),\cdot)\nabla_t x(\cdot) + \nabla_t h(x(\cdot),\cdot) \\ \Psi^{'}(kc_{(r)}(x(\cdot))); 0^{r,m-r} \end{bmatrix}.$$

Let us consider the system (7.9) for $t = 0^m$. We obtain

$$x(0^m,k) = x^*,\ \hat{\lambda}_{(r)}(0^m,k) = \lambda^*_{(r)},$$

$$\nabla^2_{xx} L_y(x(0^m,k), \hat{\lambda}_{(r)}(0^m,k)) = \nabla^2_{xx} L_y(x^*,\lambda^*) = \nabla^2_{xx} L_y,$$

$$\nabla c_r(x(0^m,k)) = \nabla c_{(r)}(x^*) = \nabla c_{(r)}$$

$$\Psi^{'}(kc_{(r)}(x(0^m,k))) = \Psi^{'}(kc_{(r)}(x^*)) = I^r,\ \Psi^{''}(kc_{(r)}(x^*)) = \psi^{''}(0)I^r.$$

We also have

$$\nabla_t h(x(t,k),t,k)_{|t=0^m} \equiv \nabla_t h(x(\cdot),\cdot)_{|t=0^m}$$

$$= [0^{n,r}\ \ k\nabla c^T_{(m-r)}(x^*)\Psi^{'}(kc_{(m-r)}(x^*))],$$

where $\Psi^{'}(kc_{(m-r)}(x^*)) = \operatorname{diag}[\psi^{'}(kc_i(x^*))]^m_{i=r+1}$.

From $c_i(x^*) \geq \sigma > 0$, $r+1 \leq i \leq m$ and the update formulas (5.2) follows

$$\|k(\nabla c_{(m-r)}(x^*))^T\Psi^{'}(kc_{(m-r)}(x^*))\| \leq 2\sigma^{-1}\|(\nabla c_{(m-r)}(x^*))^T\|.$$

The system (7.9) for $t = 0^m$ we can rewrite as follows

$$\begin{bmatrix} \nabla^2_{xx} L_y & -\nabla c^T_{(r)} \\ -\Lambda^*_{(r)}\nabla c_{(r)} & -k^{-1}I^r \end{bmatrix} \begin{bmatrix} \nabla_t x(0^m,k) \\ \nabla_t \hat{\lambda}_{(r)}(0^m,k) \end{bmatrix} = \begin{bmatrix} 0^{n,r} & k(\nabla c_{(m-r)}(x^*))^T\Psi^{'}(kc_{(m-r)}(x^*) \\ I^r & 0^{r,m-r} \end{bmatrix}.$$

or

$$\begin{bmatrix} \nabla_t x(0^m,k) \\ \nabla_t \hat{\lambda}_{(r)}(0^m,k) \end{bmatrix} \tag{7.10}$$

$$= \nabla \Phi^{-1}_{(y,k)} \begin{bmatrix} 0^{n,r} & k(\nabla_{(m-r)} c(x^*))^T\Psi^{'}(kc_{(m-r)}(x^*)) \\ I^r & 0^{r,m-r} \end{bmatrix}.$$

From (7.2), (7.10) and $k \geq k_0$ follows

$$\max\{\|\nabla_t \hat{x}(0^m,k)\|, \|\nabla_t \hat{\lambda}_{(r)}(0^m,k)\|\} \leq \rho_0 \max\{1, 2\sigma^{-1}\|(\nabla c_{(m-r)}(x^*))^T\|\} = c_0.$$

Thus, for $\delta > 0$ small enough and any $(t,k) \in S(K,\delta)$ from (7.9) follows

$$\|\Phi^{-1}_{(y,k)}(\cdot)R(\cdot)\| = \tag{7.11}$$

$$= \|\nabla \Phi^{-1}_{(y,k)}(x(\tau t,k), \hat{\lambda}_{(r)}(\tau t,k)) \cdot R(x(\tau t,k);\tau t,k)\| \leq 2c_0$$

for any $0 \le \tau \le 1$ and any $k \ge k_0$.

Using Newton-Leibniz formula

$$\begin{bmatrix} x(t,k) - x^* \\ \hat{\lambda}_{(r)}(t,k) - \lambda^*_{(r)} \end{bmatrix} = \begin{bmatrix} x(t,k) - x(0^m,k) \\ \hat{\lambda}_{(r)}(t,k) - \hat{\lambda}_{(r)}(0^m,k) \end{bmatrix} \tag{7.12}$$

$$= \int_0^1 \nabla\Phi^{-1}_{(y,k)} \Big((x(\tau t,k), \hat{\lambda}_{(r)}(\tau t,k)\Big) R(x(\tau t,k), \tau t, k)[t] d\tau$$

we obtain

$$\max\{\|x(t,k) - x^*\|, \|\hat{\lambda}_{(r)}(t,k) - \lambda^*\|\} \le 2c_0\|t\| = 2c_0 k^{-1}\|\lambda - \lambda^*\|.$$

Let $\hat{x}(\lambda,k) = x(\frac{\lambda-\lambda^*}{k}, k)$ and

$$\hat{\lambda}(\lambda,k) = (\hat{\lambda}_{(r)}(\frac{\lambda-\lambda^*}{k}), k), \hat{\lambda}_{(m-r)}(\frac{\lambda-\lambda^*}{k}, k)).$$

Then for $c = 2\max\{\sigma^{-1}, c_0\}$, which is independent on $k \ge k_0$, we obtain (7.1)

3) Let us consider the Hessian of the EDF $\mathcal{L}_y(x,\lambda,k)$ at $x = \hat{x}$ and $\lambda = \hat{\lambda}$. We have

$$\nabla^2_{xx}\mathcal{L}_y(\hat{x},\hat{\lambda},k) = \nabla^2_{xx}L_y(\hat{x},\hat{\lambda}) - k\nabla c(\hat{x})^T\Phi''(kc(\hat{x}))\hat{\Lambda}\nabla c(\hat{x}).$$

From (7.1) for $k \ge k_0$ large enough we have

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

$$\nabla^2_{xx}L_y(x^*,\lambda^*) + k\nabla c_{(r)}(x^*)^T\Lambda^*_{(r)}\nabla c_{(r)}(x^*).$$

The item 3) of Theorem 7.1 follows from the second order sufficient optimality condition (4.9)-(4.10) and Debreu's Lemma, with $A = \nabla^2_{xx}L_y(x^*,\lambda^*)$ and $C = \Lambda^{*\frac{1}{2}}_{(r)}\nabla c_{(r)}(x^*)$. The proof of Theorem 7.1 is completed.

8. Stopping Criteria

The EPM (5.1)-(5.2) is an infinite procedure, which require, at each step, solving an unconstrained optimization problem (5.1), which is, in turn, an infinite procedure as well.

The following result allows replacing x_{s+1} from (5.1) by an approximation $\bar{x}_{s+1}$, finding which requires finite procedure and does not compromise Q-linear convergence rate.

For a given $\alpha > 0$ let us consider the primal-dual approximation $(\bar{x}, \bar{\lambda})$:

$$\bar{x} = \bar{x}(\lambda,k) : \|\nabla_x\mathcal{L}_y(\bar{x},\lambda,k)\| \le \frac{\alpha}{k}\|\bar{\lambda} - \lambda\| \tag{8.1}$$

$$\bar{\lambda} = \bar{\lambda}(\lambda,k) = (\bar{\lambda}_i = \psi'(kc_i(\bar{x}))\lambda_i,\ i = 1,...,m.) \tag{8.2}$$

Obviously $\bar{x}$ depends not only on $\lambda \in \mathbb{R}^m_{++}$ and $k > 0$ but also on y and α as well.

At this point $y \in \operatorname{int}\Omega$ and $\alpha > 0$ are fixed, therefore to simplify notation we omitted y and α from the definition of $\bar{x}$ and $\bar{\lambda}$.

Theorem 8.1. *If $f,\ c_i \in C^2,\ i = 1, ..., m$ and the second order sufficient optimality condition (4.9)-(4.10) is satisfied, then for a given $\alpha > 0$, small enough $\delta > 0$, large enough k_0, any $k \geq k_0$ and any $(\lambda, k) \in \Lambda(\lambda, k, \delta)$ we have:*

1) *there exists $(\bar{x}, \bar{\lambda})$ defined by (8.1)-(8.2);*
2) *there is $c > 0$ independent on $k \geq k_0$ that the following bound*

$$\max\{\|\bar{x} - x^*\|, \|\bar{\lambda} - \lambda^*\|\} \leq \frac{c}{k}(1 + 2\alpha)\|\lambda - \lambda^*\| \tag{8.3}$$

holds;

3) *the Lagrangian $\mathcal{L}_y(x, \lambda, k)$ for the equivalent problem is strongly convex at the neighborhood of $\bar{x}$.*

Proof. For a small enough $\delta > 0$, large enough k_0 and any $k \geq k_0$ we define the following extended dual set

$$\Lambda(\lambda, k, \delta, \theta) = \Lambda(\lambda, k, \delta) \otimes \{\theta \in \mathbb{R}^n : \|\theta\| \leq \delta\}$$

$$= \{\lambda \in \mathbb{R}^m_+ : \lambda_i \geq \delta, |\lambda_i - \lambda_i^*| \leq \delta k,\ i = 1, ..., r\} \otimes \{0 < \lambda_i < \delta k,\ i = r+1, ..., m,\ k \geq k_0\}$$
$$\otimes \{\theta \in \mathbb{R}^n : \|\theta\| \leq \delta\} = \Lambda_{(r)}(\cdot) \otimes \Lambda_{(m-r)}(\cdot) \otimes \{\theta \in \mathbb{R}^n : \|\theta\| \leq \delta\}.$$

By introducing vector $t = (t_1, ..., t_r, t_{r+1}, ..., t_m)$ with $t_i = (\lambda_i - \lambda_i^*)k^{-1}$, $i = 1, ..., m$ we transform $\Lambda(\lambda, k, \delta, \theta)$ into the neighborhood of the origin of the extended dual space

$$S(0^m, k, \delta, 0^n) = S_{(r)}(0^r, k, \delta) \otimes S_{(m-r)}(0^{m-r}, k, \delta) \otimes \{\theta \in \mathbb{R}^n : \|\theta\| \leq \delta\}.$$

The following map $\Phi_y : \mathbb{R}^{2n+m+r+1} \to \mathbb{R}^{n+r}$:

$$\Phi_y(x, \bar{\lambda}_{(r)}, k, t, \theta) = \begin{pmatrix} \Delta^{-1}(y, x)\nabla f(x) - \sum_{i=1}^r \bar{\lambda}_i \nabla c_i(x) - h(x, t_{(m-r)}, k) - \theta \\ (t_i + k^{-1}\lambda_i^*)\psi^{'}(kc_i(x)) - k^{-1}\bar{\lambda}_i,\ \ i = 1, ..., r \end{pmatrix},$$

is the key ingredient of the proof.

Let us consider $\infty > k_1 > k_0$, then for any $k \in [k_0, k_1]$ and $y \in \text{int}\,\Omega$ we have

$$\Phi_y(x^*, \lambda^*_{(r)}, k, 0^m, 0^n) =$$
$$\begin{bmatrix} \Delta^{-1}(y, x^*)\nabla f(x^*) - \sum_{i=1}^r \lambda_i^* \nabla c_i(x^*) - h(x^*, 0^{m-r}, k) - 0^n \\ k^{-1}(\lambda_i^* - \lambda_i^*),\ \ i = 1, ..., r \end{bmatrix} = \begin{bmatrix} 0^n \\ 0^r \end{bmatrix}.$$

Further,

$$\nabla_{x\bar{\lambda}_{(r)}}\Phi_y \equiv \nabla_{x,\bar{\lambda}_{(r)}}\Phi_y(x^*, \lambda^*_{(r)}, k, 0^m, 0^n) =$$
$$\begin{bmatrix} \nabla^2_{xx}L_y(x^*, \lambda^*) & -\nabla c^T_{(r)}(x^*) \\ \Lambda^*_{(r)}\nabla c_{(r)}(x^*) & -k^{-1}I^r \end{bmatrix} = \nabla\Phi_{(y,k)}.$$

We saw already that the inverse $(\nabla\Phi_{(y,k)})^{-1}$ exists and there is $\rho_0 > 0$ that $\|(\nabla\Phi_{(y,k)})^{-1}\| \leq \rho_0$.

From the second Implicit Function Theorem follows existence of two vector-functions

$$x(\cdot) = x(t, k, \theta) = (x_1(t, k, \theta), ..., x_n(t, k, \theta))$$

and

$$\bar\lambda_{(r)}(\cdot)=\bar\lambda_{(r)}(t,k,\theta)=(\bar\lambda_1(t,k,\theta),...,\bar\lambda_r(t,k,\theta))$$

uniquely defined on $S(0^m,k,\delta,0^n)$ for small enough $\delta>0$ and $k\geq k_0$, that the following identities hold

$$\Delta^{-1}(y,x(\cdot))\nabla f(x(\cdot))-\sum_{i=1}^{r}\bar\lambda_i\nabla c_i(x(\cdot))-h(x(\cdot),t,k)-\theta\equiv 0 \tag{8.4}$$

$$\bar\lambda_i(\cdot)=\bar\lambda_i(t,k,\theta)\equiv(kt_i+\lambda_i^*)\psi^{'}(kc_i(x(\cdot))),\ \ i=1,..,r. \tag{8.5}$$

For a given small $\varepsilon>0$ there is $\delta>0$ that

$$\max\{\|x(t,k,\theta)-x^*\|,\|\bar\lambda_{(r)}(t,k,\theta)-\lambda^*_{(r)}\|\}\leq\varepsilon$$

for $\forall(t,k,\theta)\in S(0^m,k,\delta,0^n)$.

Therefore for the passive constraints we have

$$c_i(x(\cdot))=\bar c_i(x(t,k,\theta))\geq 0.5\sigma.$$

Hence

$$\bar\lambda_i=\lambda_i\psi^{'}(kc_i(x(\cdot)))\leq\lambda_i\psi^{'}(0.5k\sigma)\leq\frac{2}{\sigma k}\lambda_i,\ i=r+1,...,m$$

and $\sigma>0$ is independent on $k\in[k_0,k_1]$.

To prove the bound (8.3) we estimate the norms of Jacobians $\nabla_{t,\theta}x(t,k,\theta)$ and $\nabla_{t,\theta}\bar\lambda_{(r)}(t,k,\theta)$ for $t=0^m$, $\theta=0^n$.

By differentiating identities (8.4) and (8.5) in t and θ we obtain

$$\nabla^2_{xx}L_y(\cdot)\nabla_{t,\theta}x(\cdot)-\nabla c^T_{(r)}(x(\cdot))\nabla_{t,\theta}\bar\lambda_{(r)}(\cdot) \tag{8.6}$$
$$=[0^{n,r},\ \ \nabla_x h(x(\cdot),t,k)\nabla_{t,\theta}x(\cdot)+\nabla_t h(x(\cdot),t,k),\ \ I^n]$$

$$k(kT^r+\Lambda^*_{(r)})\Psi^{"}(kc_{(r)}(x(\cdot)))\nabla c_{(r)}(x(\cdot))\nabla_{t,\theta}x(\cdot)-\nabla_{t,\theta}\bar\lambda_{(r)}(\cdot) \tag{8.7}$$
$$=[k\Psi^{'}(kc_{(r)}(x(\cdot))),\ \ 0^{r,m-r},\ \ 0^{r,n}],$$

where $T^r=\mathrm{diag}(t_i)^r_{i=1}$, $\Lambda^*_{(r)}=\mathrm{diag}(\lambda^*_i)^r_{I=1}$, $\Psi^{'}(kc_{(r)}(x(\cdot)))=\mathrm{diag}(\psi^{'}(kc_i(x(\cdot))))^r_{i=1}$, $\Psi^{''}(kc_{(r)}(x(\cdot))=\mathrm{diag}(\psi^{''}(kc_i(x(\cdot))))^r_{i=1}$.

The system (8.6)-(8.7) can be rewritten as follows

$$\begin{bmatrix}\nabla_{t,\theta}x(\cdot)\\ \nabla_{t,\theta}\bar\lambda_{(r)}(\cdot)\end{bmatrix}= \tag{8.8}$$

$$\begin{bmatrix}\nabla^2 L_y(x(\cdot),\bar\lambda_{(r)}(\cdot)) & -\nabla c^T_{(r)}(\cdot)\\ (kT^r+\Lambda^*_{(r)})\Psi^{''}(kc_{(r)}(x(\cdot))\nabla c_{(r)}(x(\cdot))) & -k^{-1}I^r\end{bmatrix}^{-1}\times R(x(\cdot),t,k),$$

where

$$R(x(\cdot),t,k)=\begin{bmatrix}0^{n,r} & \nabla_x h(x(\cdot),t,k)\nabla_{t,\theta}x(\cdot)+\nabla_t h(x(\cdot),t,k) & I^n\\ \Psi^{'}(kc_{(r)}(x(\cdot))) & 0^{r,m-r} & 0^{r,n}\end{bmatrix}.$$

Let us consider system (8.8) for $t = 0^m$ and $\theta = 0^n$. We obtain

$$(8.9)\qquad \begin{bmatrix} \nabla_{t,\theta} x(0^m, k, 0^n) \\ \nabla_{t,\theta} \bar{\lambda}_{(r)}(0^m, k, 0^n) \end{bmatrix} = \begin{bmatrix} \nabla^2 L_y(x^*, \lambda^*_{(r)}) & -\nabla c^T_{(r)}(x^*) \\ -\Lambda^*_{(r)} \nabla c_{(r)}(x^*) & -k^{-1} I^r \end{bmatrix}^{-1} \times$$

$$\begin{bmatrix} 0^{n,r} & k\nabla c^T_{(m-r)}(x^*)\Psi'(kc_{(m-r)}(x^*) & I^n \\ I^r & 0^{r,m-r} & 0^{r,n} \end{bmatrix} = \Phi^{-1}_{(y,k)} R,$$

where

$$\Psi'(kc_{(m-r)}(x^*)) = \mathrm{diag}(\psi'(kc_i(x^*)))_{i=r+1}^m.$$

Keeping in mind (7.2) from (8.9) we obtain

$$\max\{\|\nabla_{t,\theta} x(0^m, k, 0^n)\|,\ \|\nabla_{t,\theta} \bar{\lambda}_{(r)}(0^m, k, 0^n)\|\} \leq$$

$$\rho_0 \max\{1, \sigma^{-1}\|(\nabla c_{(m-r)}(x^*))^T\|\} = c_0$$

and c_0 is independent on $k \geq k_0$. Thus,

$$\begin{bmatrix} x(t,k,\theta) - x^* \\ \bar{\lambda}_{(r)}(t,k,\theta) - \lambda^*_{(r)} \end{bmatrix} = \begin{bmatrix} x(t,k,\theta) - x(0^m,k,0^n) \\ \bar{\lambda}_{(r)}(t,k,\theta) - \bar{\lambda}_{(r)}(0^m,k,0^n) \end{bmatrix} =$$

$$= \int_0^1 \nabla \Phi^{-1}_{(y,k)}(x(\tau t, k, \tau\theta), \bar{\lambda}_{(r)}(\tau t, k, \tau\theta)) R(x(\tau t, k, \tau\theta); \tau t, k, \tau\theta) \begin{bmatrix} t \\ \theta \end{bmatrix} d\tau.$$

Hence,

$$\max\{\|x(t,k,\theta) - x^*\|, \|\bar{\lambda}_{(r)}(t,k,\theta) - \lambda^*_{(r)}\|\} \leq 2c_0 k^{-1}\|\lambda - \lambda^*\| + \|\theta\|.$$

Let

$$\bar{x} = \bar{x}(t,k,\theta) = x\left(\frac{\lambda - \lambda^*}{k}, k, \theta\right)$$

$$\bar{\lambda} = \bar{\lambda}(t,k,\theta) = \left(\bar{\lambda}_{(r)}\left(\frac{\lambda - \lambda^*}{k}, k, \theta\right); \bar{\lambda}_{(m-r)}\left(\frac{\lambda - \lambda^*}{k}, k, \theta\right)\right),$$

then for $c = 2\max\{\sigma^{-1}, c_0\}$ we have

$$(8.10)\qquad \|\bar{x} - x^*\| \leq \frac{c}{k}\|\lambda - \lambda^*\| + \|\theta\|$$

$$(8.11)\qquad \|\bar{\lambda} - \lambda^*\| \leq \frac{c}{k}\|\lambda - \lambda^*\| + \|\theta\|.$$

Keeping in mind the stopping criteria (8.1)-(8.2) we obtain

$$\|\nabla_x \mathcal{L}_y(\bar{x}, \lambda, k)\| = \|\theta\| \leq \frac{\alpha}{k}\|\bar{\lambda} - \lambda\|.$$

Therefore

$$(8.12)\qquad \|\bar{x} - x^*\| \leq \frac{c}{k}\|\lambda - \lambda^*\| + \frac{\alpha}{k}\|\bar{\lambda} - \lambda\|$$

$$(8.13)\qquad \|\bar{\lambda} - \lambda^*\| \leq \frac{c}{k}\|\lambda - \lambda^*\| + \frac{\alpha}{k}\|\bar{\lambda} - \lambda\|.$$

From (8.13) follows

$$\|\bar{\lambda} - \lambda^*\| \leq \frac{c}{k}\|\lambda - \lambda^*\| + \frac{\alpha}{k}\|\bar{\lambda} - \lambda^*\| + \frac{\alpha}{k}\|\lambda^* - \lambda\|$$

or

$$\left(1-\frac{\alpha}{k}\right)\|\bar{\lambda}-\lambda^*\| \le \frac{c+\alpha}{k}\|\lambda-\lambda^*\|. \tag{8.14}$$

For $k_0 > c+2\alpha$ and any $k \ge k_0$ from (8.14) follows

$$\|\bar{\lambda}-\lambda^*\| \le \frac{c+2\alpha}{k}\|\lambda-\lambda^*\| \tag{8.15}$$

From (8.12) and (8.15) we obtain

$$\|\bar{x}-x^*\| \le \frac{c}{k}\|\lambda-\lambda^*\| + \frac{\alpha}{k}\|\bar{\lambda}-\lambda^*\| + \frac{\alpha}{k}\|\lambda-\lambda^*\|$$

$$= \frac{c+\alpha}{k}\|\lambda-\lambda^*\| + \frac{\alpha}{k}\|\bar{\lambda}-\lambda^*\| \le \left[\frac{c+\alpha}{k} + \frac{\alpha}{k}\frac{(c+2\alpha)}{k}\right]\|\lambda-\lambda^*\|.$$

Again for $k_0 > c+2\alpha$ and any $k \ge k_0$ we have

$$\|\bar{x}-x^*\| \le \frac{c+2\alpha}{k}\|\lambda-\lambda^*\|.$$

The proof of Theorem 8.1 is completed.

We conclude the section by considering the numerical realization of the EPM.

The EPM scheme consists of inner and outer iteration. On the inner iteration we find an approximation $\bar{x}$ for the primal minimizer using the stopping criteria (8.1).

On the outer iteration we update the Lagrange multipliers by (8.2), using the approximation $\bar{x}$.

For finding $\bar{x}$ any unconstrained minimization technique can be used. Fast gradient method (see [19]) or regularized Newton method (see [21]) are two possible candidates.

Under usual convexity and smoothness assumptions both methods converges to the minimizer from any starting point and for both methods there exist complexity bounds, that is the upper bound for the number of step required for finding an ε- approximation for the minimizer.

To describe the numerical realization of EPM we need to introduce the relaxation operator $R : \Omega_{-k^{-1}} \times \mathbb{R}^m_{++} \to \Omega_{-k^{-1}} \times \mathbb{R}^m_{++}$, with is defined as follows

$$Ru = \bar{u} = (\bar{x}, \bar{\lambda}), \tag{8.16}$$

where $\bar{x}$ and $\bar{\lambda}$ are given by (8.1) and (8.2).

We also need the merit function $\nu_y : \Omega_{-k^{-1}} \times \mathbb{R}^m_+ \to R$, which is defined by the following formula

$$\nu_y(u) = \max\{\|\nabla_x L_y(x,\lambda)\|,\ \sum_{i=1}^m \lambda_i|c_i(x)|, -c_i(x),\ i=1,...,m\}. \tag{8.17}$$

From (8.17) follows $\nu_y(u) \ge 0$, $\forall u \in \Omega_{-k^{-1}} \times \mathbb{R}^m_+$, it is also easy to see that

$$\nu_y(u) = 0 \Leftrightarrow u = u^* = (x^*; \lambda^*) \tag{8.18}$$

holds.

Moreover, under the second order sufficient optimality condition and f, $c_i \in C^2$, $i = 1, ..., m$ the merit function ν_y in the neighborhood of u^* is similar to the norm of a gradient of a strong convex function with Lipschitz continuous gradient in the neighborhood of the minimizer (see [26]).

Let $\gamma > 0$ be small enough, $y \in \text{int } \Omega_\gamma$ be the initial "center", $u = (x; \lambda) \in \Omega_{-k^{-1}} \times \mathbb{R}^m_{++}$ be the initial primal-dual approximation, $\Delta > 0$ be the reduction parameter for the objective function, $k > 0$ be the scaling parameter and $\varepsilon > 0$ be the required accuracy.

The EPM consists of the following operations

1. find $\bar{u} = (\bar{x}; \bar{\lambda}) = Ru$
2. if $\nu_y(\bar{u}) \leq \varepsilon$, then $u^* = (x^*; \lambda^*) := (\bar{x}; \bar{\lambda})$ else;
3. find $\bar{\tau} = \max\{0 \leq \tau \leq 1 : x(\tau) = y + t(\bar{x} - y) \in \Omega_\gamma,\ (\nabla f(x(\tau)), \bar{x} - y) \leq 0\}$ and $x(\bar{\tau})$;
4. if

$$f(y) - f(x(\bar{\tau})) \geq \Delta, \tag{8.19}$$

then update the center $\bar{y} := 0.5(y + x(\bar{x}))$, set $x := \bar{x}$, $y := \bar{y}$ and go to 1;
else set $x := \bar{x}$; $\lambda := \bar{\lambda}$ and go to 1.

It follows from 3. and 4. that the sequence of centers is monotone decreasing in value, therefore from some point on the inequality (8.19) can't be satisfied, so the "center" is fixed.

Hence, from this point on the primal-dual sequence is generated only by the relaxation operator (8.16) and converge to the primal-dual solution with Q-linear rate due to Theorem 8.1

9. Concluding Remarks

It follows from 1.-4. that the efficiency of the EPM heavily depends on the efficiency of the unconstrained minimization algorithm used in operator R.

The absence of singularity of EDF at the solution combined with stability of its Hessian's condition number improves substantially the efficiency of the operator R.In particular, it allows to reduce the number of unconstrained minimization steps per Lagrange multipliers update.

On the other hand, under fixed Lagrange multipliers EDF possesses self-concordance properties for a wide classes of constrained optimization problems.

It provides an opportunity to combine the nice feature of the IPM at the beginning of the computational process with excellent EDF properties at the final phase.

The NR approach produced very strong numerical results for wide classes of large scale nonlinear optimization problems (see, for example, [1], [3], [9], [17]).

In particular, one of the most reliable NLP solver PENNON is based on NR theory (see [15], [16]).

It leads us to believe that the extra tool, which EDF possesses, can contribute to the numerical efficiency mainly because updating the center does not require much computational effort, but can substantial reduce the objective function value.

It means that updating the "center" will allow to reach the "hot start" faster (see [9], [22], [25]).

References

[1] A. Ben-Tal, M. Zibulevski, *Penalty-barrier methods for convex programming problems*, SIAM J. Optim. 7 (1997), 347-366.

[2] D. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, (1982).

[3] M. Breitfeld, D. Shanno, *Computational experience with modified log-barrier methods for nonlinear programming*, Annals of OR, 62 (1996), 439-464.

[4] Lieu Bui-Trong, P. Huard, *La methods des centres dans un espace topologique*, Num. Mat. 8, No. 1, (1966), 56-67.

[5] A. V. Fiacco, G. P. McCormick, *Nonlinear Programming:Sequential Unconstrained Minimization Techniques* Wiley, New York (1968).

[6] K.Frisch, *The logarithmic Potential Method for convex programming. Memorandum of may 13 1955*, University Institute of Economics, Oslo, (1955).

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

[8] C. Gonzaga, *Path-following methods for Linear Programming*, SIAM review, 34, N2, (1992), 167-224.

[9] I. Griva, R. Polyak, *Primal-dual nonlinear rescaling method with dynamic scaling parameter update.*, Math. Program. Ser. A 106 (2006), 237-259.

[10] K. Grossman, K. Kaplan , *Nonlinear Programming based on unconstrained minimization*, Novosibirsk Nauca (1981)(in Russian).

[11] P. Huard, *Resolution of mathematical programming with nonlinear constraints by the method of centres*, In: Nonlinear Programming, North Holland, Amsterdam (1967), 207-219.

[12] P. Huard, *A method of centres by upper-bounding function with applications*, In: Nonlinear Programming, North Holland, Amsterdam (1967), 1-30.

[13] F. Jarre, G. Sonnevend, G. Stoer, *An implementation of the method of analytic centers*, Benoussan, A., Lions, J. L., eds., Lecture Notes in Control and Information Sciences, No. 111, Springer, Berlin, (1988).

[14] N. A. Karmarkar, *A new polynomial - time algorithm for linear programming*, Combinatorica, 4, (1984), 373-395.

[15] M. Kocvara, M. Stingl, *PENNON. A code for convex nonlinear and semidefinite programming.*, Optimization methods and software, 18, 3 (2003), 317-333.

[16] M. Kocvara, M. Stingl, *Resent progress in the NLP-SDP code PENNON*, Workshop "Optimization and Applications", Oberwalfach (2005).

[17] S. Nash, R. Polyak and A. Sofer, *A numerical comparison of barrier and modified barrier method for large scale bound-constrained optimization*, Large Scale Optimization, State of the Art. W. Hager, D. Hearn, P. Pardalos (Eds.). Kluwer Academic Publishers, (1994), 319-338.

[18] Y. Nesterov, A. Nemirovsky , *Interior Point Polynomial Methods in Convex Programming: Theory and Algorithms*, SIAM, Philadelphia, (1994).

[19] Yu. Nesterov, *Inroductory Lectures on Convex Optimization: A basic course* Kluwer (2004).
[20] E. Polak, *Computational Method in Optimization: A Unified Approach*, Academic Press, New Yotk (1971).
[21] R.Polyak, *Regularized Newton method for unconstrained convex optimization*, Mathematical Programming, 1, Vol. 120, (2009), 125-145.
[22] R. Polyak, *Modified Barrier Functions.(Theory and Methods)*, Mathematical Programming, 54, Vol. 2, (1992), 177-222.
[23] R. Polyak, M. Teboulle *Nonlinear Rescaling and Proximal-like Methods in Convex Optimization*, Mathematical Programming, 76, (1997), 265-284.
[24] R. Polyak, *Modified Interior Distance functions*, Contemporary Mathematics, AMS Vol 209 (1997), 183-209.
[25] R. Polyak, *Legendre transform in modern optimization* , Optimization and Applications in Control and Data Sciences. Springer, vol. 115 (2016), 437-507.
[26] R. Polyak, *Local Quadratic Convergence of the Primal - Dual Augmented Lagrangian method Optimization method and Software*, v 24,(2009), 369-379.
[27] J. Renegar, *A polynomial-time algorithm, based on Newton's method for linear [programming]*, Mathematical Programming, 40 (1988), 59-93.
[28] G. Sonnevend, *An analytical centre for polyhedrons and new classes of global algorithms for linear (smooth convex) programming*, Lecture Notes, Control Inform. Sci 84, Springer-Verlag, New York, NY, (1985), 866-876.
[29] P. M. Vaidya, *An algorithm for linear programming which requires* $O(((m+n)n^2 + (m+n)^{1.5})L)$ *arithmetic operations*, Technical Report, AT & T Bell Laboratories, Murray Hill, NJ (1987).

(R. Polyak) Department of Mathematics, The Technion - Israel Institute of Technology, 32000 Haifa, Israel
E-mail address: rpolyak@techunix.technion.ac.il and rpolyak@gmu.edu