

MINIMAX EXACTNESS AND GLOBAL SADDLE POINTS OF NONLINEAR AUGMENTED LAGRANGIANS

MAKSIM V. DOLGOPOLIK

Institute for Problems in Mechanical Engineering, Russian Academy of Sciences, St. Petersburg, Russia

Abstract. We study global minimax exactness of merit functions for constrained optimization problems. This concept arises as a natural generalization of the definition of global saddle points in the unified theory of exactness of penalty and augmented Lagrangian functions. We obtain necessary and sufficient conditions for the global minimax exactness of nonlinear augmented Lagrangians in the form of the localization principle, which allow one to reduce the study of the existence of global saddle points (or the existence of solutions of the augmented dual problem) to a local analysis of sufficient optimality conditions. With the use of the localization principle, we obtain simple necessary and sufficient conditions for the existence of global saddle points of He-Wu-Meng's augmented Lagrangian for inequality constrained problems and nonlinear rescaling Lagrangians for nonconvex semidefinite programs. We also introduce and analyze a new nonlinear smooth augmented Lagrangian for constrained minimax problems and provide necessary and sufficient conditions for the existence of a global saddle point of this augmented Lagrangian, which, in particular, expose some limitations of exponential penalty function methods.

Keywords. Augmented Lagrangian; Global saddle point; Augmented duality; Semidefinite programming; Minimax problem.

1. INTRODUCTION

The existence of global saddle points is one of the central questions of the general theory of augmented Lagrangian functions for constrained optimization problems. The existence of global saddle points was studied for general cone constrained optimization problems [1, 2], mathematical programming problems [3, 4, 5, 6, 7, 8, 9, 10], nonlinear second order cone programming problems [11], nonlinear semidefinite programming problems [12, 13], semi-infinite programming problems [14, 15], etc. A general theory of the existence of global saddle points of augmented Lagrangian functions for nonlinear cone constrained optimization problems that unified and subsumed many existing results in this area was developed in [16].

Another closely related problem is the problem of the existence of the so-called *augmented Lagrange multipliers* of Rockafellar-Wets' augmented Lagrangian. This augmented Lagrangian was introduced in [17] and thoroughly analyzed by many researchers [18, 19, 20, 21, 22, 23]. The existence of augmented Lagrange multipliers of Rockafellar-Wets' augmented Lagrangian for various types of optimization problems was studied in [1, 2, 14, 15, 24, 25, 26].

E-mail address: maxim.dolgopolik@gmail.com.

Received October 19, 2020; Accepted November 24, 2020.

An analysis of the aforementioned papers on the existence of global saddle points and augmented Lagrange multipliers indicates that the main ideas of these papers largely overlap and the existence theorems are very similar for all kinds of augmented Lagrangians and optimization problems. As was shown in [16, 24], one can formulate a general principle, called *the localization principle*, that allows one to obtain all known results on the existence of global saddle points/augmented Lagrange multipliers in the finite dimensional case in a unified manner. However, the results of [16, 24] cannot be applied to nonlinear (in the objective function) augmented Lagrangians that were studied in, e.g., [9, 23, 27]

In [28, 29], a unified theory of various concepts of exactness of penalty and augmented Lagrangian functions was developed. This theory provides a unified approach to an analysis of linear [30, 31, 32, 33] and nonlinear [34, 35, 36] exact penalty functions, Fletcher's continuously differentiable exact penalty functions [37, 38, 39, 40], exact augmented Lagrangians [16, 41, 42, 43], Huyer and Neumaier's exact penalty functions [44, 45], and the existence of augmented Lagrange multipliers in the finite dimensional case.

The main goal of this paper is to demonstrate that the theory of exact merit functions developed in [28, 29] admits a natural extension that allows one to include various existing results on global saddle points of nonlinear augmented Lagrangian functions into the unified theory and apply the ideas developed in [16, 24] to all types of *nonlinear* augmented Lagrangians. To this end, we introduce a new concept of minimax exactness of a merit function and analyze its connections with the existence of global saddle points and some properties of the augmented dual problem. Then we extend the localization principle for global saddle points/augmented Lagrange multipliers from [16, 24] to the case of general nonlinear merit functions and apply this principle to an analysis of three particular kinds of augmented Lagrangians. Namely, we study He-Wu-Meng's augmented Lagrangian [46], which has not attracted much attention of researchers despite its interesting properties pointed out in [16]. We also study nonlinear rescaling Lagrangian functions for nonlinear semidefinite programming problems [12, 16, 47, 48, 49, 50, 51], which can be viewed as a natural extension of the modified barrier functions and nonlinear rescaling methods developed by professor R. Polyak [52, 53, 54] to the case of semidefinite programming problems. Finally, we analyze a nonlinear augmented Lagrangian function for constrained minimax problems inspired by the paper [27] in order to demonstrate that the general theory developed in this paper can be applied to *nonlinear* (in the objective function) augmented Lagrangians and to show how this theory can be utilized in order to understand theoretical limitations of various augmented Lagrangian methods.

The paper is organized as follows. Some preliminary ideas and definitions from the general theory of exact merit functions are collected in Section 2. Minimax exactness and global saddle points of nonlinear merit functions are analyzed in Section 3. Section 4 contains an extension of the localization principle to the case of minimax exactness. Finally, applications of the general theory developed in this paper to three particular classes of augmented Lagrangians are presented in Section 5.

2. PRELIMINARIES

Let X be a finite dimensional real normed space and $M, A \subset X$ be nonempty sets. Throughout this article we study the following optimization problem:

$$\min f(x) \quad \text{subject to} \quad x \in M, \quad x \in A. \quad (\mathcal{P})$$

Here $f: X \rightarrow \mathbb{R} \cup \{+\infty\}$ is a given function, while the inclusions $x \in M$ and $x \in A$ represent two different types of constraints (e.g. nonlinear and linear constraints or equality and inequality constraints, etc.) written in a nonfunctional form. Below we suppose that there exists a globally optimal solution of (\mathcal{P}) and the optimal value of this problem is finite.

Our aim is to understand how one can remove the constraint $x \in M$ of the problem (\mathcal{P}) with the use of a merit function $F(\cdot)$ without losing any information about globally optimal solutions of the problem (\mathcal{P}) .

Let Λ be a nonempty set of parameters that are denoted by λ , and $c > 0$ be the *penalty parameter*. We suppose that a merit function $F: X \times \Lambda \times (0, +\infty) \rightarrow \mathbb{R} \cup \{+\infty\}$, $F = F(x, \lambda, c)$, is given. A connection between this function and the problem (\mathcal{P}) is specified below.

The function F can be, for instance, a penalty function with Λ being the empty set or an augmented Lagrangian with λ being a Lagrange multiplier. However, in order not to restrict ourselves to any specific case, we simply call $F(x, \lambda, c)$ a *merit function* for the problem (\mathcal{P}) .

Our main goal is to obtain general conditions under which the problem (\mathcal{P}) is equivalent (in some sense) to a problem of minimizing the function $F(x, \lambda, c)$ over a certain set. In other words, we are interested in conditions ensuring global exactness (in some sense) of the merit function $F(x, \lambda, c)$.

There are essentially three different ways to define the concept of exactness of merit functions. The first one is called *parametric exactness*. According to this concept, one fixes a parameter $\lambda_* \in \Lambda$ and considers the following auxiliary problem:

$$\min_x F(x, \lambda_*, c) \quad \text{subject to} \quad x \in A.$$

If there exists $c^* > 0$ such that for all $c \geq c^*$ points of global minimum of this problem coincide with points of global minimum of the problem (\mathcal{P}) , then the merit function $F(x, \lambda, c)$ is called *globally parametrically exact* with the *exact tuning parameter* λ_* . Thus, if $F(x, \lambda, c)$ is globally parametrically exact and an exact tuning parameter λ_* is known, then one can choose sufficiently large $c > 0$ and minimize the function $F(\cdot, \lambda_*, c)$ over the set A in order to find globally optimal solutions of the problem (\mathcal{P}) . This approach to exactness naturally suits the theory of linear [30, 31, 32, 33] and nonlinear [34, 35, 36] penalty functions, Fletcher's continuously differentiable exact penalty functions [37, 38, 39, 40], and the theory of existence of augmented Lagrange multipliers [1, 2, 15, 24], in which augmented Lagrange multipliers plays the role of exact tuning parameters. Parametric exactness of merit functions was introduced and studied in detail in [28].

The second concept of exactness is called *extended exactness*. Since the problem of finding an exact tuning parameter is very complicated (unless, of course, $F(x, \lambda, c)$ is a *penalty function*, i.e. unless $F(x, \lambda, c)$ does not depend on λ), a different approach to the definition of global exactness is needed. Namely, instead of fixing λ_* one considers the problem of minimizing $F(x, \lambda, c)$ with respect to both x and λ , i.e. one considers the extended problem:

$$\min_{(x, \lambda)} F(x, \lambda, c) \quad \text{subject to} \quad (x, \lambda) \in A \times \Lambda.$$

If there exists $c^* \geq 0$ such that for all $c \geq c^*$ a pair (x_*, λ_*) is a point of global minimum of this problem iff x_* is a globally optimal solution of the problem (\mathcal{P}) , then the merit function $F(x, \lambda, c)$ is called *extendedly exact*. Extended exactness is used within the theory of exact

augmented Lagrangians [16, 41, 42, 43] and Huyer and Neumaier's exact penalty functions [44, 45]. A general theory of extendedly exact merit functions was developed in [29].

Finally, the third concept of exactness naturally arises within the theory of global saddle points of augmented Lagrangians and we call it *minimax exactness*. The main goal of this article is to study minimax exactness of merit functions and its connections with global saddle points and augmented dual problems.

3. MINIMAX EXACTNESS AND GLOBAL SADDLE POINTS

We start with an extension of the definition of global saddle point to the case of merit functions.

Definition 3.1. A pair $(x_*, \lambda_*) \in A \times \Lambda$ is called a *global saddle point* of the merit function $F(x, \lambda, c)$ iff there exists $c^* > 0$ such that

$$\max_{\lambda \in \Lambda} F(x_*, \lambda, c) \leq F(x_*, \lambda_*, c) \leq \min_{x \in A} F(x, \lambda_*, c) \quad \forall c \geq c^*$$

and $F(x_*, \lambda_*, c) < +\infty$ for all $c \geq c^*$. The greatest lower bound of all such c^* is denoted by $c_{gsp}^*(x_*, \lambda_*)$ and is called *the least exact penalty parameter* at (x_*, λ_*) .

Definition 3.2. The merit function $F(x, \lambda, c)$ is called *globally minimax exact* iff there exists $\lambda_* \in \Lambda$ and $c^* > 0$ such that

- (1) for any globally optimal solution x_* of the problem (\mathcal{P}) the pair (x_*, λ_*) is a global saddle point of $F(x, \lambda, c)$ and $c_{gsp}^*(x_*, \lambda_*) \leq c^*$;
- (2) if (x_*, λ_*) is a global saddle point of $F(x, \lambda, c)$, then x_* is a globally optimal solution of the problem (\mathcal{P}) .

Any such λ_* is called a *minimax exact tuning parameter* of the merit function $F(x, \lambda, c)$, while the greatest lower bound of all such c^* is called *the least minimax exact penalty parameter* and is denoted by $c_{gsp}^*(\lambda_*)$.

It is easy to see that if $F(x, \lambda, c)$ is globally minimax exact with a minimax exact tuning parameter λ_* , then $F(x, \lambda, c)$ is globally parametrically exact with λ_* being an exact tuning parameter. On the other hand, minimax exactness provides one with a criterion for distinguishing λ_* from other tuning parameters $\lambda \in \Lambda$, which allows one to construct primal-dual methods for solving the problem (\mathcal{P}) based on the use of the merit function $F(x, \lambda, c)$.

Let us point out an obvious connection between minimax exactness and the existence of global saddle points. Denote by Ω_* the set of globally optimal solutions of the problem (\mathcal{P}) .

Proposition 3.1. Suppose that the following two assumptions are valid:

- (1) if (x_*, λ_*) is a global saddle point of $F(x, \lambda, c)$, then $x_* \in \Omega_*$,
- (2) if (x_*, λ_*) is a global saddle point of $F(x, \lambda, c)$, then there exists c^* such that for any $y_* \in \Omega_*$ the pair (y_*, λ_*) is a global saddle point of $F(x, \lambda, c)$ and $c_{gsp}^*(y_*, \lambda_*) \leq c^*$.

Then the merit function $F(x, \lambda, c)$ is globally minimax exact iff there exists a global saddle point of this function. Furthermore, λ_* is a minimax exact tuning parameter iff there exists $x_* \in A$ such that (x_*, λ_*) is a global saddle point of $F(x, \lambda, c)$.

Thus, under the assumptions of the proposition above the problem of whether $F(x, \lambda, c)$ is globally minimax exact is equivalent to the problem of whether a global saddle point of

$F(x, \lambda, c)$ exists. Furthermore, in this case there is a one to one correspondence between minimax exact tuning parameters and global saddle points of the merit function $F(x, \lambda, c)$.

Let us note that the first assumption of Proposition 3.1 is satisfied for most existing augmented Lagrangian functions, while the second assumptions of this proposition can usually be verified with the use of the augmented dual problem. Recall that *the augmented dual problem* associated with the merit function $F(x, \lambda, c)$ has the form

$$\max \Theta(\lambda, c) \quad \text{subject to} \quad \lambda \in \Lambda, \quad c > 0,$$

where $\Theta(\lambda, c) = \inf_{x \in A} F(x, \lambda, c)$ is *the augmented dual function* associated with $F(x, \lambda, c)$. One says that there is *no duality gap* between the primal problem (\mathcal{P}) and the augmented dual problem iff

$$\inf_{x \in \Omega} f(x) = \sup_{\lambda \in \Lambda, c > 0} \Theta(\lambda, c),$$

where $\Omega = M \cap A$ is the feasible set of the problem (\mathcal{P}). The propositions below describe interconnections between global minimax exactness, the existence of global saddle points, duality gap, and the existence of solutions of the augmented dual problem.

Proposition 3.2. *Let the function $F(x, \lambda, c)$ be nondecreasing in c and let (x_*, λ_*) with $x_* \in \Omega_*$ be a global saddle point of $F(x, \lambda, c)$ such that $F(x_*, \lambda_*, c) = f(x_*)$ for all $c > 0$. Then there is no duality gap between the problem (\mathcal{P}) and the augmented dual problem, and for any $c > c_{gsp}^*(x_*, \lambda_*)$ the pair (λ_*, c) is a globally optimal solution of the dual problem.*

Proof. By the definition of global saddle point for any $c > c_{gsp}^*(x_*, \lambda_*)$ one has

$$\max_{\lambda \in \Lambda} F(x_*, \lambda, c) \leq F(x_*, \lambda_*, c) \leq \min_{x \in A} F(x, \lambda_*, c) =: \Theta(\lambda_*, c).$$

Therefore for any $c > c_{gsp}^*(x_*, \lambda_*)$ one has

$$\Theta(\lambda_*, c) = F(x_*, \lambda_*, c) = f(x_*) \geq F(x_*, \lambda, c) \geq \Theta(\lambda, c) \quad \forall \lambda \in \Lambda.$$

Hence taking into account the fact that $F(x, \lambda, c)$ is nondecreasing in c one obtains that for any $c > c_{gsp}^*(x_*, \lambda_*)$ the following inequalities hold true

$$\Theta(\lambda_*, c) = f(x_*) \geq \Theta(\lambda, r) \quad \forall \lambda \in \Lambda, r > 0.$$

Thus, (λ_*, c) is a globally optimal solution of the augmented dual problem, and there is no duality gap between the problem (\mathcal{P}) and the augmented dual problem. \square

Proposition 3.3. *Suppose that there is no duality gap between the problem (\mathcal{P}) and the augmented dual problem and the function $F(x, \lambda, c)$ is nondecreasing in c . Suppose also that*

$$F(x_*, \lambda, c) \leq f(x_*) \quad \forall x_* \in \Omega_*, \lambda \in \Lambda, c > 0. \quad (3.1)$$

Then for any $x_ \in \Omega_*$ and for any globally optimal solution (λ_*, c_*) of the augmented dual problem the pair (x_*, λ_*) is a global saddle point of $F(x, \lambda, c)$ with $c_{gsp}^*(x_*, \lambda_*) \leq c_*$.*

Proof. Fix $x_* \in \Omega_*$ and let (λ_*, c_*) be a globally optimal solution of the augmented dual problem. Applying inequality (3.1) and the facts that there is no duality gap and $F(x, \lambda, c)$ is nondecreasing in c one obtains that

$$\Theta(\lambda_*, c) = f(x_*) \geq F(x_*, \lambda_*, c) \quad \forall c \geq c_*.$$

By definition one has $\Theta(\lambda_*, c) = \inf_{x \in A} F(x, \lambda_*, c)$. Therefore

$$f(x_*) = F(x_*, \lambda_*, c) = \min_{x \in A} F(x, \lambda_*, c) \quad \forall c \geq c_*.$$

Applying inequality (3.1) once again one gets that

$$\sup_{\lambda \in \Lambda} F(x_*, \lambda, c) \leq f(x_*) = F(x_*, \lambda_*, c) \leq \min_{x \in A} F(x, \lambda_*, c) \quad \forall c \geq c_*,$$

which implies the desired result. \square

Combining Propositions 3.1–3.3 we obtain two equivalent characterizations of global minimax exactness.

Theorem 3.1. *Let the following assumptions be valid:*

- (1) $F(x, \lambda, c)$ is nondecreasing in c ;
- (2) if (x_*, λ_*) is a global saddle point of $F(x, \lambda, c)$, then $x_* \in \Omega_*$;
- (3) $F(x_*, \lambda_*, c) = f(x_*)$ for any global saddle point (x_*, λ_*) and for all $c > 0$;
- (4) $F(x_*, \lambda, c) \leq f(x_*)$ for all $x_* \in \Omega_*$, $\lambda \in \Lambda$ and $c > 0$.

Then the three following statements are equivalent:

- (1) $F(x, \lambda, c)$ is globally minimax exact;
- (2) there exists a global saddle point of $F(x, \lambda, c)$;
- (3) there is no duality gap between the primal problem (\mathcal{P}) and the augmented dual problem and there exists a globally optimal solution of the augmented dual problem.

Proof. If $F(x, \lambda, c)$ is globally minimax exact, then, clearly, there exists a global saddle point of $F(x, \lambda, c)$.

If there exists a global saddle point of $F(x, \lambda, c)$, then by Proposition 3.2 there is no duality gap and the augmented dual problem has optimal solutions.

Finally, if there is no duality gap and there exists a globally optimal solution (λ_*, c_*) of the augmented dual problem, then by Proposition 3.3 for any $x_* \in \Omega_*$ the pair (x_*, λ_*) is a global saddle point of $F(x, \lambda, c)$ and $c_{gsp}^*(x_*, \lambda_*) \leq c_*$. Consequently, applying Proposition 3.1 one obtains that $F(x, \lambda, c)$ is globally minimax exact. \square

Corollary 3.1. *Let all assumptions of the previous theorem be valid and let (x_*, λ_*) be a global saddle point of the merit function $F(x, \lambda, c)$. Then for any $y_* \in \Omega_*$ the pair (y_*, λ_*) is a global saddle point of $F(x, \lambda, c)$ as well and $c_{gsp}^*(y_*, \lambda_*) = c_{gsp}^*(x_*, \lambda_*)$.*

Remark 3.1. (i) Let us note that the assumptions of the previous theorem are satisfied for most of the existing augmented Lagrangian functions (see [16] and the examples below).

(ii) It should be mentioned that in most particular cases if (x_*, λ_*) is a global saddle point of the merit function $F(x, \lambda, c)$, then (x_*, λ_*) is a KKT-point of the problem (\mathcal{P}) (cf. [16, Prp. 9]). Therefore, according to the corollary above, for the existence of a global saddle point in these cases it is necessary that there exists $\lambda_* \in \Lambda$ such that for any $x_* \in \Omega_*$ the pair (x_*, λ_*) is a KKT-point of the problem (\mathcal{P}) . Thus, in particular, if there exist two globally optimal solutions of the problem (\mathcal{P}) with disjoint sets of Lagrange multipliers, then a global saddle point of $F(x, \lambda, c)$ does not exist and $F(x, \lambda, c)$ cannot be globally minimax exact.

The theorem above demonstrates that the question of whether a merit function $F(x, \lambda, c)$ is globally minimax exact is directly related to the existence of global saddle points and duality.

In particular, Theorem 3.1 indicates the importance of the notion of minimax exactness for the study of primal-dual methods.

4. LOCALIZATION PRINCIPLE

Our aim now is to obtain general and easily verifiable necessary and sufficient conditions for the global minimax exactness of the merit function $F(x, \lambda, c)$. These conditions allow one to reduce the study of global minimax exactness to a local analysis of the behaviour of the function $F(x, \lambda, c)$ near globally optimal solutions of the problem (\mathcal{P}) . That is why the theorem containing these conditions is called *the localization principle*.

At first, let us describe a desired local behaviour of the function $F(x, \lambda, c)$ near globally optimal solutions that is necessary to ensure its global minimax exactness. This desired behaviour is described in terms of the existence of *local saddle points*.

Definition 4.1. A pair $(x_*, \lambda_*) \in A \times \Lambda$ is called a *local saddle point* of the merit function $F(x, \lambda, c)$ iff there exists a neighbourhood U of x_* and $c^* > 0$ such that

$$\max_{\lambda \in \Lambda} F(x_*, \lambda, c) \leq F(x_*, \lambda_*, c) \leq \min_{x \in U \cap A} F(x, \lambda_*, c) \quad \forall c \geq c^*$$

and $F(x_*, \lambda_*, c) < +\infty$ for all $c \geq c^*$. The greatest lower bound of all such c^* is denoted by $c_{lsp}^*(x_*, \lambda_*)$ and is called *the least local exact penalty parameter* at (x_*, λ_*) .

Note that in the definitions of local and global saddle points it is assumed that a pair (x_*, λ_*) is a local or global saddle point of the function $(x, \lambda) \mapsto F(x, \lambda, c)$ for any sufficiently large value of the parameter $c > 0$, which alludes to its role as a *penalty* parameter. However, a connection between the parameter $c > 0$ and penalization is not specified in the definition of the merit function $F(x, \lambda, c)$. We need the following definition in order to clarify this connection.

Definition 4.2. Let $\lambda_* \in \Lambda$ be fixed. One says that $F(x, \lambda, c)$ is a *penalty-like* merit function for $\lambda = \lambda_*$ if there exist an increasing unbounded sequence $\{c_n\} \subset (c_0, +\infty)$ and $x_n \in \arg \min_{x \in A} F(x, \lambda_*, c_n)$, $n \in \mathbb{N}$, such that the sequence $\{x_n\}$ is bounded and all its cluster points are globally optimal solution of the problem (\mathcal{P}) .

Roughly speaking, $F(x, \lambda, c)$ is a penalty-like merit function for $\lambda = \lambda_*$ iff the function $F(\cdot, \lambda_*, c)$ attains a global minimum on the set A for any sufficiently large $c > 0$ and global minimizers of this function on the set A tend to globally optimal solutions of the problem (\mathcal{P}) , which is the case for a standard penalty function, provided this function has bounded sublevel sets. However, to be more precise we must point out that in the definition above it is *not* assumed that *all* minimizers of $F(\cdot, \lambda_*, c)$ tend to globally optimal solutions of the problem (\mathcal{P}) . Instead, we only suppose that one can find an increasing unbounded sequence $\{c_n\}$ and a sequence of global minimizers of $F(\cdot, \lambda_*, c_n)$ that converges to the set of globally optimal solutions.

Now, we can formulate and prove the main result of this section.

Theorem 4.1 (Localization principle in the minimax form I). *Suppose that the following assumptions are valid:*

- (1) $F(x, \lambda, c)$ is nondecreasing in c ;
- (2) if (x_*, λ_*) is a global saddle point of $F(x, \lambda, c)$, then there exists $c^* > 0$ such that for any $y_* \in \Omega_*$ the pair (y_*, λ_*) is a global saddle point of $F(x, \lambda, c)$ and $c_{gsp}^*(y_*, \lambda_*) \leq c^*$;
- (3) if (x_*, λ_*) is a global saddle point of $F(x, \lambda, c)$, then $x_* \in \Omega_*$;

- (4) if $\max_{\lambda \in \Lambda} F(x_*, \lambda, c) \leq F(x_*, \lambda_*, c) < +\infty$ for some $x_* \in A$, $\lambda_* \in \Lambda$ and $c > 0$, then the function $F(x_*, \lambda_*, \cdot)$ is constant.

Then the merit function $F(x, \lambda, c)$ is globally minimax exact if and only if there exists $\lambda_* \in \Lambda$ such that

- (1) $F(x, \lambda, c)$ is a penalty-like merit function for $\lambda = \lambda_*$;
 (2) for any $x_* \in \Omega_*$ the pair (x_*, λ_*) is a local saddle point of $F(x, \lambda, c)$.

Proof. Let $F(x, \lambda, c)$ be globally minimax exact with a minimax exact tuning parameter λ_* . Then for any $x_* \in \Omega_*$ the pair (x_*, λ_*) is a global (and hence local) saddle point of $F(x, \lambda, c)$. From the definition of global saddle point it follows that for any $c > c_{gsp}(x_*, \lambda_*)$ the point x_* is a global minimizer of the function $F(\cdot, \lambda_*, c)$ on the set A , which implies that $F(x, \lambda, c)$ is a penalty-like merit function for $\lambda = \lambda_*$.

Let us prove the “if” part of the theorem. Suppose that there exists $\lambda_* \in \Lambda$ such that $F(x, \lambda, c)$ is a penalty-like merit function for $\lambda = \lambda_*$ and for any $x_* \in \Omega_*$ the pair (x_*, λ_*) is a local saddle point of this function. Since $F(x, \lambda, c)$ is a penalty-like merit function for $\lambda = \lambda_*$, without loss of generality one can suppose that there exist an increasing unbounded sequence $\{c_n\} \subset (0, +\infty)$ and $x_n \in \arg \min_{x \in A} F(x, \lambda_*, c_n)$, $n \in \mathbb{N}$, such that the sequence $\{x_n\}$ converges to some $x_* \in \Omega_*$. By our assumption, the pair (x_*, λ_*) is a local saddle point of $F(x, \lambda, c)$. Hence there exist $c^* > 0$ and a neighbourhood U of x_* such that

$$\max_{\lambda \in \Lambda} F(x_*, \lambda, c) \leq F(x_*, \lambda_*, c) \leq \min_{x \in U \cap A} F(x, \lambda_*, c) \quad \forall c \geq c^*. \quad (4.1)$$

Taking into account the fact that $\{x_n\}$ converges to x_* one obtains that $x_n \in U \cap A$ for any sufficiently large n , which implies that

$$F(x_*, \lambda_*, c) \leq F(x_n, \lambda_*, c) \quad \forall c \geq c^*$$

for any n large enough. Hence taking into account the definition of x_n one obtains that there exists $n_0 \in \mathbb{N}$ such that for any $n \geq n_0$ the point x_* is a global minimizer of $F(\cdot, \lambda_*, c_n)$ on the set A .

Note that from (4.1) and assumption 4 of the theorem it follows that $F(x_*, \lambda_*, \cdot)$ is a constant function. Consequently, with the use of (4.1) and the fact the function $F(x, \lambda, c)$ is nondecreasing in c one gets that

$$\max_{\lambda \in \Lambda} F(x_*, \lambda, c) \leq F(x_*, \lambda_*, c) \leq \min_{x \in A} F(x, \lambda_*, c_{n_0}) \leq \min_{x \in A} F(x, \lambda_*, c)$$

for all $c \geq c_{n_0}$. Thus, (x_*, λ_*) is a global saddle point of $F(x, \lambda, c)$. Hence taking into account Proposition 3.1 one gets the required result. \square

Remark 4.1. Let us note that assumption 4 is satisfied for most of the existing augmented Lagrangian functions (see examples below and [16, Remark 6]).

Let us also give a slightly different formulation of the localization principle in the minimax form that is more convenient for applications. For this purpose, let us introduce a stronger version of global minimax exactness in which one specifies the optimal value of the augmented dual problem.

Definition 4.3. Let $\Theta_* \in \mathbb{R}$ be fixed. The merit function $F(x, \lambda, c)$ is called *globally minimax exact with optimal value Θ_** iff $F(x, \lambda, c)$ is globally minimax exact and there exists $c_0 > 0$ such that for any $c \geq c_0$ and any global saddle point (x_*, λ_*) of $F(x, \lambda, c)$ one has $F(x_*, \lambda_*, c) = \Theta_*$.

It is easy to check that if $F(x, \lambda, c)$ is nondecreasing in c and for any global saddle point (x_*, λ_*) the function $c \mapsto F(x_*, \lambda_*, c)$ is constant for all $c > c_{gsp}^*(x_*, \lambda_*)$, then the merit function $F(x, \lambda, c)$ is globally minimax exact with optimal value Θ_* iff $F(x, \lambda, c)$ is globally minimax exact and the optimal value of the augmented dual problem is equal to Θ_* (cf. the proof of Prp. 3.2). In particular, the case $\Theta_* = \inf_{x \in \Omega} f(x)$ corresponds to the case when there is no duality gap between the primal and the augmented dual problems.

We need to recall an auxiliary definition from [28, 29].

Definition 4.4. Let $\lambda_* \in \Lambda$ be fixed. One says that $F(x, \lambda, c)$ is a *penalty-type* merit function for $\lambda = \lambda_*$ iff there exists $c_0 > 0$ such that if

- (1) $\{c_n\} \subset [c_0, +\infty)$ is an increasing unbounded sequence,
- (2) $x_n \in \arg \min_{x \in A} F(x, \lambda_*, c_n)$, $n \in \mathbb{N}$,
- (3) x_* is a cluster point of the sequence $\{x_n\}$,

then x_* is a globally optimal solution of the problem (\mathcal{P}) .

Roughly speaking, $F(x, \lambda, c)$ is a penalty-type merit function for $\lambda = \lambda_*$ iff global minimizers of the function $F(\cdot, \lambda_*, c)$ on the set A either tend to globally optimal solutions of the problem (\mathcal{P}) or escape to infinity as $c \rightarrow +\infty$. Note that unlike Definition 4.2, here we do not suppose that global minimizers of $F(\cdot, \lambda_*, c)$ exist and, furthermore, bounded. Instead, the above definition describes a general property of a penalty-type merit function: if this function attains a global minimum in x on the set A and a sequence of points of global minimum converges as the penalty parameter c increases unboundedly, then the limit point of this sequence must be a globally optimal solution of the problem (\mathcal{P}) . This property is satisfied for most existing penalty and augmented Lagrangian functions under very mild assumptions.

For any $x_* \in \Omega_*$ denote by $\Lambda_{loc}(x_*)$ the set of all $\lambda_* \in \Lambda$ for which the pair (x_*, λ_*) is a local saddle point of $F(x, \lambda, c)$. Define also $\Lambda_{loc}(\mathcal{P}) = \bigcap_{x_* \in \Omega_*} \Lambda_{loc}(x_*)$. Note that under the assumptions of Theorem 4.1 for the global minimax exactness of the merit function $F(x, \lambda, c)$ it is necessary that $\Lambda_{loc}(\mathcal{P}) \neq \emptyset$. Moreover, Theorem 4.1, in essence, states that λ_* is a minimax exact tuning parameter of $F(x, \lambda, c)$ iff $\lambda_* \in \Lambda_{loc}(\mathcal{P})$ and $F(x, \lambda, c)$ is a penalty-like merit function for $\lambda = \lambda_*$.

Now, we can give another formulation of the localization principle in the minimax form.

Theorem 4.2 (Localization principle in the minimax form II). *Let A be closed and the function $F(\cdot, \lambda, c)$ be l.s.c. on A for any $c > 0$ and $\lambda \in \Lambda$. Suppose also that the following assumptions are valid:*

- (1) $F(x, \lambda, c)$ is nondecreasing in c ;
- (2) if (x_*, λ_*) is a global saddle point of $F(x, \lambda, c)$, then there exists $c^* > 0$ such that for any $y_* \in \Omega_*$ the pair (y_*, λ_*) is a global saddle point of $F(x, \lambda, c)$ and $c_{gsp}^*(y_*, \lambda_*) \leq c^*$;
- (3) if (x_*, λ_*) is a global saddle point of $F(x, \lambda, c)$, then $x_* \in \Omega_*$;
- (4) if $\max_{\lambda \in \Lambda} F(x_*, \lambda, c) \leq F(x_*, \lambda_*, c) < +\infty$ for some $x_* \in A$, $\lambda_* \in \Lambda$ and $c > 0$, then the function $F(x_*, \lambda_*, \cdot)$ is constant;
- (5) $F(x, \lambda, c)$ is a penalty-type merit function for any $\lambda \in \Lambda_{loc}(\mathcal{P})$.

Then $F(x, \lambda, c)$ is globally minimax exact with optimal value $\Theta_* \in \mathbb{R}$ if and only if there exists $\lambda_* \in \Lambda$ and $c_0 > 0$ such that

- (1) (x_*, λ_*) is a local saddle point of $F(x, \lambda, c)$ for any $x_* \in \Omega_*$;

- (2) for any $x_* \in \Omega_*$ and $c \geq c_0$ one has $F(x_*, \lambda_*, c) = \Theta_*$;
 (3) there exists a bounded set $K \subset \mathbb{R}^d$ such that

$$S(\lambda_*, \Theta_*, c) := \left\{ x \in A \mid F(x, \lambda_*, c) < \Theta_* \right\} \subseteq K \quad \forall c \geq c_0.$$

Proof. If $F(x, \lambda, c)$ is globally minimax exact with optimal value Θ_* and exact tuning parameter λ_* , then by definition for any $x_* \in \Omega_*$ the pair (x_*, λ_*) is a global (and hence local) saddle point of $F(x, \lambda, c)$ and there exists $c_0 > 0$ such that $F(x_*, \lambda_*, c) = \Theta_*$ for any $c \geq c_0$ and $x_* \in \Omega_*$. Consequently, with the use of the definition of global saddle point one obtains that

$$\min_{x \in A} F(x, \lambda_*, c) \geq F(x_*, \lambda_*, c) = \Theta_* \quad \forall x_* \in \Omega_* \quad \forall c > \max\{c_0, c_{gsp}^*(\lambda_*)\}.$$

Therefore the set $S(\lambda_*, \Theta_*, c)$ is empty for all $c > \max\{c_0, c_{gsp}^*(\lambda_*)\}$ and, thus, $S(\lambda_*, \Theta_*, c)$ is contained in any bounded set $K \subset X$ for all sufficiently large $c > 0$.

Let us prove the converse statement. More precisely, let us verify that under the assumptions of the theorem $F(x, \lambda, c)$ is a penalty-like merit function for $\lambda = \lambda_*$. Then applying Theorem 4.1 one obtains that $F(x, \lambda, c)$ is globally minimax exact. Hence taking into account the fact that for any $x_* \in \Omega_*$ and $c \geq c_0$ one has $F(x_*, \lambda_*, c) = \Theta_*$ one obtains that $F(x, \lambda, c)$ is globally minimax exact with optimal value Θ_* .

At first, suppose that there exists $\bar{c} \geq c_0$ such that for any $c \geq \bar{c}$ the set $S(\lambda_*, \Theta_*, c)$ is empty. Then for any $c \geq \bar{c}$ any point $x_* \in \Omega_*$ is a global minimizer of the function $F(\cdot, \lambda_*, c)$ on the set A , which obviously implies that $F(x, \lambda, c)$ is a penalty-like merit function for $\lambda = \lambda_*$.

Thus, one can suppose that there exists an increasing unbounded sequence $\{c_n\} \subset [c_0, +\infty)$ such that for any $n \in \mathbb{N}$ the set $S_n := S(\lambda_*, \Theta_*, c_n)$ is not empty. By our assumptions the set S_n is contained in a bounded set K and $F(\cdot, \lambda_*, c_n)$ is l.s.c. on A . Therefore for any $n \in \mathbb{N}$ the function $F(\cdot, \lambda_*, c_n)$ attains a global minimum on the set A at a point $x_n \in K$, which implies that the sequence $\{x_n\}$ is a bounded. Hence without loss of generality one can suppose that this sequence converges to a point x_* , which is a globally optimal solution of the problem (\mathcal{P}) due to the fact that $F(x, \lambda, c)$ is a penalty-type merit function for any $\lambda \in \Lambda_{loc}(\mathcal{P})$. Thus, $F(x, \lambda, c)$ is a penalty-like merit function for $\lambda = \lambda_*$. \square

5. APPLICATIONS OF THE LOCALIZATION PRINCIPLE

Below, we utilize the main results of the previous sections in order to obtain simple necessary and sufficient conditions for the global minimax exactness (or, equivalently, for the existence of global saddle points) of several particular augmented Lagrangian functions.

5.1. Example I: He-Wu-Meng's augmented Lagrangian. In this subsection we study the augmented Lagrangian function for inequality constrained optimization problem introduced by He, Wu, and Meng in [46]. Let $X = \mathbb{R}^d$ and suppose that the set M has the form

$$M = \left\{ x \in \mathbb{R}^d \mid g_i(x) \leq 0, \quad i \in I \right\},$$

where $g_i: \mathbb{R}^d \rightarrow \mathbb{R}$, $i \in I = \{1, \dots, m\}$, are given functions.

Following the ideas of He, Wu, and Meng [46], introduce the function

$$\Phi(y, \lambda, c) = \frac{1}{c} \int_0^{cy} \left(\sqrt{t^2 + \lambda^2} + t \right) dt.$$

It is easy to see that $\Phi(y, 0, c) = cy(|y| + y)/2$, while, for any $\lambda \neq 0$,

$$\Phi(y, \lambda, c) = \frac{y}{2} \sqrt{c^2 y^2 + \lambda^2} + \frac{cy^2}{2} + \frac{\lambda^2}{2c} \ln \left(\sqrt{c^2 y^2 + \lambda^2} + cy \right) - \frac{\lambda^2}{2c} \ln |\lambda|.$$

Let us note some simple properties of the function $\Phi(y, \lambda, c)$.

Lemma 5.1. *The following statements hold true:*

- (1) $\Phi(y, \lambda, c) \geq 0$ for any $y \geq 0$ and $\Phi(y, \lambda, c) \leq 0$ for any $y \leq 0$;
- (2) $\Phi(y, 0, c) > 0$ for any $y > 0$ and $\Phi(y, 0, c) = 0$ for any $y \leq 0$;
- (3) the function $\Phi(y, \lambda, c)$ is nondecreasing in c and in y ;
- (4) $\lim_{c \rightarrow +\infty} \Phi(y, \lambda, c) = +\infty$ for any $y > 0$ and $\lambda \in \mathbb{R}$;
- (5) $\lim_{c \rightarrow +\infty} \Phi(y, \lambda, c) = 0$ for any $y \leq 0$ and $\lambda \in \mathbb{R}$.

With the use of the function $\Phi(y, \lambda, c)$, we can define an augmented Lagrangian function for the problem (\mathcal{P}) . Namely, set $\Lambda = \mathbb{R}^m$. Following [46], define

$$\mathcal{L}(x, \lambda, c) = f(x) + \sum_{i=1}^m \Phi(g_i(x), \lambda^{(i)}, c),$$

where $\lambda = (\lambda^{(1)}, \dots, \lambda^{(m)})$. We call this function He-Wu-Meng's augmented Lagrangian. Our aim is to obtain simple necessary and sufficient conditions for the global minimax exactness (or, equivalently, for the existence of global saddle points) of this augmented Lagrangian with the use of the localization principle. To this end, let us prove a simple lemma, which is very helpful for verifying the validity of assumptions of Theorem 4.2.

Lemma 5.2. *Let $(x_*, \lambda_*) \in A \times \Lambda$ be such that*

$$\max_{\lambda \in \Lambda} \mathcal{L}(x_*, \lambda, c_0) \leq \mathcal{L}(x_*, \lambda_*, c_0) < +\infty \quad (5.1)$$

for some $c_0 > 0$. Then x_ is a feasible point of the problem (\mathcal{P}) , the complementarity condition $\lambda_*^{(i)} g_i(x_*) = 0$, $i \in I$, holds true, and $\mathcal{L}(x_*, \lambda_*, c) = f(x_*)$ for all $c > 0$.*

Proof. Arguing by reductio ad absurdum, suppose that x_* is infeasible. Then $g_i(x_*) > 0$ for some $i \in I$. For all $s \in \mathbb{R}$ define $\lambda^{(i)}(s) = s$ and $\lambda^{(j)}(s) = 0$ for any $j \neq i$. From (5.1) it follows that for any $s \in \mathbb{R}$ one has

$$\begin{aligned} \mathcal{L}(x_*, \lambda_*, c_0) &\geq \mathcal{L}(x_*, \lambda(s), c_0) \geq f(x_*) + \Phi(g_i(x_*), s, c_0) \\ &\geq f(x_*) + \frac{1}{c_0} \int_0^{c_0 g_i(x_*)} \sqrt{s^2} dt = f(x_*) + g_i(x_*) |s|. \end{aligned}$$

(here we used the second statement from Lemma 5.1). Passing to the limit as $s \rightarrow +\infty$ one gets that $\mathcal{L}(x_*, \lambda_*, c_0) = +\infty$, which is impossible. Thus, x_* is a feasible point of the problem (\mathcal{P}) .

Let us now check that the complementarity condition holds true. Then, as is easy to see, $\mathcal{L}(x_*, \lambda_*, c) = f(x_*)$ for all $c > 0$. Arguing by reductio ad absurdum suppose that there exists $i \in I$ such that $g_i(x_*) < 0$, but $\lambda_*^{(i)} \neq 0$. Then $\Phi(g_i(x_*), \lambda_*^{(i)}, c) < 0$ for any $c > 0$, which implies that $\mathcal{L}(x_*, \lambda_*, c_0) < f(x_*)$ (recall that x_* is feasible). On the other hand, from (5.1) it follows that

$$f(x_*) > \mathcal{L}(x_*, \lambda_*, c_0) \geq \mathcal{L}(x_*, 0, c_0) = f(x_*),$$

which is impossible. Thus, $\lambda_*^{(i)} g_i(x_*) = 0$ for all $i \in I$. □

Now, we can formulate and prove the main result of this subsection. Denote by f_* the optimal value of the problem (\mathcal{P}) .

Theorem 5.1 (Localization principle for He-Wu-Meng's augmented Lagrangian). *Let A be closed, f be l.s.c. on A and g_i , $i \in I$, be continuous on A . Then a global saddle point of He-Wu-Meng's augmented Lagrangian $\mathcal{L}(x, \lambda, c)$ exists (or, equivalently, $\mathcal{L}(x, \lambda, c)$ is globally minimax exact with optimal value f_*) if and only if there exists $\lambda_* \in \mathbb{R}^m$ and $c_0 > 0$ such that*

- (1) *for any $x_* \in \Omega_*$ the pair (x_*, λ_*) is a local saddle point of $\mathcal{L}(x, \lambda, c)$;*
- (2) *the set $\{x \in A \mid \mathcal{L}(x, \lambda_*, c_0) < f_*\}$ is either bounded or empty.*

Proof. Applying Lemma 5.1 one obtains that the function $\mathcal{L}(x, \lambda, c)$ is nondecreasing in c and $\mathcal{L}(x, \lambda, c) \leq f(x)$ for any feasible point x .

Let (x_*, λ_*) be a global saddle point of $\mathcal{L}(x, \lambda, c)$. Then x_* is feasible and $\mathcal{L}(x_*, \lambda_*, c) = f(x_*)$ for all $c > 0$ by Lemma 5.2. Hence with the use of the definition of global saddle point one obtains that

$$f(x_*) = \mathcal{L}(x_*, \lambda_*, c) \leq \mathcal{L}(x, \lambda_*, c) \leq f(x) \quad \forall c > c_{gsp}^*(x_*, \lambda_*)$$

for any feasible point x , which implies that $x_* \in \Omega_*$. Therefore for any $y_* \in \Omega_*$ the pair (y_*, λ_*) is a global saddle point of $\mathcal{L}(x, \lambda, c)$ and $c_{gsp}^*(y_*, \lambda_*) = c_{gsp}^*(x_*, \lambda_*)$ by virtue of Corollary 3.1.

Thus, it remains to check that $\mathcal{L}(x, \lambda, c)$ is a penalty-type merit function for any $\lambda_* \in \Lambda_{loc}(\mathcal{P})$. Then applying the localization principle (Theorem 4.2) one obtains the desired result.

Fix arbitrary $\lambda_* \in \Lambda_{loc}(\mathcal{P})$. Let $\{c_n\} \subset (0, +\infty)$ be an increasing unbounded sequence, $x_n \in \arg \min_{x \in A} \mathcal{L}(x, \lambda_*, c_n)$ for all $n \in \mathbb{N}$, and x_* be a cluster point of the sequence $\{x_n\}$. Replacing, if necessary, the sequence $\{c_n\}$ with its subsequence one can suppose that x_n converges to x_* . Note that $x_* \in A$, since A is closed.

Let x_0 be a globally optimal solution of the problem (\mathcal{P}) . Taking into account the fact that $\lambda_* \in \Lambda_{loc}(\mathcal{P})$ and applying Lemma 5.2 one obtains that $\mathcal{L}(x_0, \lambda_*, c) = f(x_0) = f_*$ for any $c > 0$. Hence with the use of fact that $\mathcal{L}(x, \lambda, c)$ is nondecreasing in c one obtains that

$$\mathcal{L}(x_n, \lambda_*, c) \leq \mathcal{L}(x_n, \lambda_*, c_n) \leq \mathcal{L}(x_0, \lambda_*, c_n) = f_*$$

for any $c > 0$ and for any sufficiently large $n \in \mathbb{N}$. Passing to the limit as $n \rightarrow +\infty$ and taking into account the facts that f is l.s.c. on A and g_i , $i \in I$, are continuous on A one obtains that $\mathcal{L}(x_*, \lambda_*, c) \leq f_*$ for any $c > 0$. Let us check that x_* is feasible. Then passing to the limit as $c \rightarrow +\infty$ with the use of the last statement of Lemma 5.1 one gets that $f(x_*) \leq f_*$, which implies that x_* is a globally optimal solution of the problem (\mathcal{P}) and $\mathcal{L}(x, \lambda, c)$ is a penalty-type merit function for $\lambda = \lambda_*$.

Arguing by reductio ad absurdum suppose that x_* is infeasible. Then $g_i(x_*) > 0$ for some $i \in I$. Since x_n converges to x_* , there exists $n_0 \in \mathbb{N}$ such that

$$g_i(x_n) \geq \frac{g_i(x_*)}{2} > 0, \quad g_j(x_n) \geq y_0 := \min_{j \neq i} \min\{g_j(x_*) - 1, 0\} \quad \forall j \neq i$$

for any $n \geq n_0$. Hence taking into account the fact that the function $\Phi(y, \lambda, c)$ is nondecreasing in y by Lemma 5.1 one obtains that

$$f_* \geq \mathcal{L}(x_n, \lambda_*, c_n) \geq f(x_n) + \Phi\left(\frac{g_i(x_*)}{2}, \lambda_*^{(i)}, c_n\right) + \sum_{j \neq i} \Phi(y_0, \lambda_*^{(j)}, c_n)$$

for any $n \geq n_0$. Therefore passing to the limit as $n \rightarrow +\infty$ with the use of the last two statements of Lemma 5.1 and the fact that $y_0 \leq 0$ one gets that $\mathcal{L}(x_n, \lambda_*, c_n) \rightarrow +\infty$ as $n \rightarrow +\infty$, which is impossible. Thus, x_* is a feasible point of the problem (\mathcal{P}) . \square

Remark 5.1. (i) One can easily verify that if (x_*, λ_*) is a local saddle point of He-Wu-Meng's augmented Lagrangian function $\mathcal{L}(x, \lambda, c)$ and the functions f and g_i , $i \in I$ are differentiable at x_* , then the pair $(x_*, \bar{\lambda})$ with $\bar{\lambda} = (|\lambda_*^{(1)}|, \dots, |\lambda_*^{(m)}|)$ is a KKT-point of the problem (\mathcal{P}) satisfying the standard second order necessary optimality conditions. On the other hand, one can check (see [16, Thrm. 2]) that if (x_*, λ_*) is a KKT-point satisfying second order sufficient optimality conditions and the set A is convex, then (x_*, λ_*) is a local saddle point of $\mathcal{L}(x, \lambda, c)$. With the use of these results and the localization principle one can easily formulate sufficient conditions for the existence of a global saddle point of He-Wu-Meng's augmented Lagrangian function in terms of sufficient optimality conditions.

(ii) Observe that for any $y < 0$ one has

$$\Phi(y, \lambda, c) = \frac{1}{c} \int_{-c|y|}^0 (-\sqrt{t^2 + \lambda^2} - t) dt \geq \frac{1}{c} \int_{-c|y|}^0 (-\sqrt{\lambda^2}) dt = -|\lambda y|.$$

With the use of the inequality above one can easily verify that the set $\{x \in A \mid \mathcal{L}(x, \lambda_*, c_0) < f_*\}$ is bounded for some $c_0 > 0$, provided the set $\{x \in A \mid f(x) + \sum_{i \in I} |\lambda_*^{(i)}| \min\{g_i(x), 0\} < f_*\}$ is bounded. Alternatively, it is sufficient to suppose that the functions g_i , $i \in I$, are bounded below on the set A and the set $\{x \in A \mid f(x) < f_* + \alpha\}$ is bounded for some $\alpha > 0$. Let us note that similar assumptions were utilized in various papers on augmented Lagrangian functions (cf. [3, 9, 55]).

5.2. Example II: nonlinear rescaling Lagrangians for nonlinear semidefinite optimization problems. Let us apply the general theory of globally minimax exact merit function to the study of a class of augmented Lagrangian functions, called nonlinear rescaling Lagrangians, for nonlinear semidefinite programming problems [12, 16, 47, 48, 49, 50, 51]. Suppose that $X = \mathbb{R}^d$ and the set M has the form

$$M = \left\{ x \in \mathbb{R}^d \mid G(x) \preceq 0, h(x) = 0 \right\},$$

where $G: X \rightarrow \mathbb{S}^\ell$ and $h: X \rightarrow \mathbb{R}^s$ are given functions, \mathbb{S}^ℓ is the set of all $\ell \times \ell$ real symmetric matrices, and the relation $G(x) \preceq 0$ means that the matrix $G(x)$ is negative semidefinite. In this case the problem (\mathcal{P}) is a nonlinear semidefinite programming problem.

We suppose that the space \mathbb{S}^ℓ is equipped with the inner product $\langle A, B \rangle = \text{Tr}(AB)$, and the corresponding norm $\|A\|_F = \sqrt{\text{Tr}(A^2)}$, which is called the Frobenius norm of a matrix $A \in \mathbb{S}^\ell$. Denote by \mathbb{S}_+^ℓ the cone of $\ell \times \ell$ positive semidefinite matrices and by \mathbb{S}_-^ℓ the cone of $\ell \times \ell$ negative semidefinite matrices.

Recall that for any function $\psi: \mathbb{R} \rightarrow \mathbb{R}$ and for any $A \in \mathbb{S}^\ell$ the matrix function (or Löwner's operator) associated with ψ [56, 57, 58] is defined as

$$\Psi(A) = P \text{diag} \left(\psi(\lambda_1(A)), \dots, \psi(\lambda_\ell(A)) \right) P^T,$$

where $A = P \text{diag}(\lambda_1(A), \dots, \lambda_\ell(A)) P^T$ is a spectral decomposition of the matrix A and $\lambda_i(A)$ are the eigenvalues of A listed in the decreasing order.

Let us define an augmented Lagrangian function for nonlinear semidefinite programming problems. Let $\psi: \mathbb{R} \rightarrow \mathbb{R}$ be a nondecreasing twice continuously differentiable convex function such that $\psi(0) = 0$ and $\psi'(0) = 1$. Let also $\xi: \mathbb{R} \rightarrow \mathbb{R}$ be a twice continuously differentiable convex function such that $\xi(t) = 0$ for any $t \leq 0$ and $\xi(t) > 0$ for all $t > 0$. In particular, one can define $\psi(t) = e^t - 1$ and $\xi(t) = \max\{0, t\}^3$.

Denote $\Lambda = \mathbb{S}_+^\ell \times \mathbb{R}^s$. Following the ideas of [12], for any $\lambda = (\mu, \nu) \in \Lambda$, define the following augmented Lagrangian:

$$\mathcal{L}(x, \lambda, c) = f(x) + \frac{1}{c} \langle \mu, \Psi(cG(x)) \rangle + \frac{1}{c} \text{Tr}(\Xi(cG(x))) + \langle \nu, h(x) \rangle + \frac{c}{2} \|h(x)\|^2.$$

Here Ξ is the matrix function associated with ξ and $\|\cdot\|$ is the Euclidean norm. One can easily verify that the function $\mathcal{L}(x, \mu, \nu, c)$ is nondecreasing in c .

Our goal is to apply the localization principle in the minimax form to the augmented Lagrangian function $\mathcal{L}(x, \mu, \nu, c)$. To this end, let us prove a simple auxiliary result.

Lemma 5.3. *Let $(x_*, \mu_*, \nu_*) \in A \times \Lambda$ be such that*

$$\max_{(\mu, \nu) \in \Lambda} \mathcal{L}(x_*, \mu, \nu, c_0) \leq \mathcal{L}(x_*, \mu_*, \nu_*, c_0) < +\infty \quad (5.2)$$

for some $c_0 > 0$. Then x_ is a feasible point of the problem (\mathcal{P}) and for all $c > 0$ one has $\mathcal{L}(x_*, \mu_*, \nu_*, c) = f(x_*)$.*

Proof. Arguing by reductio ad absurdum, suppose that x_* is infeasible. Then either $h(x_*) \neq 0$ or $\lambda_1(G(x_*)) > 0$, where, as above, $\lambda_1(G(x_*))$ is the largest eigenvalue of the matrix $G(x_*)$.

Suppose, at first, that $\lambda_1(G(x_*)) > 0$. For any $s > 0$ define $\mu(s) = P \text{diag}(s, 0, \dots, 0) P^T$, where P is a matrix from a spectral decomposition $G(x_*) = P \text{diag}(\lambda_1(G(x_*)), \dots, \lambda_\ell(G(x_*))) P^T$ composed from the eigenvectors of the matrix $G(x_*)$. Then, as is easy to see, for any $c > 0$ one has $\langle \mu(s), \Psi(cG(x_*)) \rangle = s\psi(c\lambda_1(G(x_*))) > 0$. Therefore $\mathcal{L}(x_*, \mu(s), 0, c_0) \rightarrow +\infty$ as $s \rightarrow +\infty$ and $\mathcal{L}(x_*, \mu_*, \nu_*, c_0) = +\infty$ due to (5.2), which is impossible. Thus, $G(x_*) \leq 0$.

Suppose, now, that $h(x_*) \neq 0$. Then, for $\nu(s) = sh(x_*)$, one has $\mathcal{L}(x_*, 0, \nu(s), c_0) \rightarrow +\infty$ as $s \rightarrow +\infty$, which is impossible by virtue of (5.2). Thus, x_* is a feasible point of the problem (\mathcal{P}) .

Let us verify that $\langle \mu_*, \Psi(cG(x_*)) \rangle = 0$ for all $c > 0$. Then one can conclude that for all $c > 0$ one has $\mathcal{L}(x_*, \mu_*, \nu_*, c) = f(x_*)$. Suppose that $\langle \mu_*, \Psi(cG(x_*)) \rangle \neq 0$. From the facts that $G(x_*) \in \mathbb{S}_-^\ell$, ψ is nondecreasing and $\psi(0) = 0$ it follows that $\Psi(cG(x_*)) \in \mathbb{S}_-^\ell$ as well. Therefore there exists $A_c \in \mathbb{S}^\ell$ such that $\Psi(cG(x_*)) = -A_c^2$. By definition one has $\mu_* \in \mathbb{S}_+^\ell$, which implies that there exists $B \in \mathbb{S}^\ell$ such that $\mu_* = B^2$. Hence taking into account the fact that the trace is invariant under cyclic permutations one obtains that

$$\langle \mu_*, \Psi(cG(x_*)) \rangle = \text{Tr}(\mu_* \Psi(cG(x_*))) = -\text{Tr}(BA_c^2 B) \leq 0,$$

since the matrix $BA_c^2 B$ is obviously positive semidefinite. Hence $\langle \mu_*, \Psi(cG(x_*)) \rangle < 0$ and $\mathcal{L}(x_*, \mu_*, \nu_*, c_0) < f(x_*)$. On the other hand, from (5.2) and the feasibility of x_* it follows that

$$f(x_*) = \mathcal{L}(x_*, 0, 0, c_0) \leq \mathcal{L}(x_*, \mu_*, \nu_*, c_0) < f(x_*),$$

which is impossible. Thus, the proof is complete. \square

Remark 5.2. Note that from the proof of the second part of the lemma it follows that for any feasible point x one has $\mathcal{L}(x, \mu, \nu, c) \leq f(x)$.

Now we can obtain simple necessary and sufficient conditions for the existence of a global saddle point of the augmented Lagrangian $\mathcal{L}(x, \lambda, c)$.

Theorem 5.2 (Localization principle for nonlinear rescaling Lagrangians). *Let A be closed, f be l.s.c. on A , while the functions G and h be continuous on A . Suppose also that $\psi'(t) \rightarrow 0$ as $t \rightarrow -\infty$ and $\xi(t)/t \rightarrow +\infty$ as $t \rightarrow +\infty$. Then a global saddle point of the augmented Lagrangian $\mathcal{L}(x, \lambda, \mu, c)$ exists (or, equivalently, $\mathcal{L}(x, \lambda, c)$ is globally minimax exact with the optimal value f_*) if and only if there exist $\mu_* \in \mathbb{S}_+^\ell$, $v_* \in \mathbb{R}^s$ and $c_0 > 0$ such that*

- (1) *for any $x_* \in \Omega_*$ the triplet (x_*, μ_*, v_*) is a local saddle point of $\mathcal{L}(x, \mu, v, c)$;*
- (2) *the set $\{x \in A \mid \mathcal{L}(x, \mu_*, v_*, c_0) < f_*\}$ is either bounded or empty.*

Proof. One can readily verify that the function $\mathcal{L}(\cdot, \mu, v, c)$ is l.s.c. on A . Furthermore, as was noted above, the function $\mathcal{L}(x, \mu, v, c)$ is nondecreasing in c and $\mathcal{L}(x, \lambda, \mu, c) \leq f(x)$ for any feasible point x .

Let (x_*, μ_*, v_*) be a global saddle point of $\mathcal{L}(x, \lambda, \mu, c)$. Then applying Lemma 5.3 one obtains that $\mathcal{L}(x_*, \mu_*, v_*, c) = f(x_*)$ for any $c > 0$ and

$$f(x_*) = \mathcal{L}(x_*, \mu_*, v_*, c) \leq \mathcal{L}(x, \mu_*, v_*, c) \leq f(x) \quad \forall c > c_{gsp}^*(x_*, \mu_*, v_*)$$

for any feasible point x . Therefore x_* is a globally optimal solution of the problem (\mathcal{P}) . Consequently, for any $y_* \in \Omega_*$ the triplet (y_*, μ_*, v_*) is a global saddle point of $\mathcal{L}(x, \mu, v, c)$ and $c_{gsp}^*(y_*, \mu_*, v_*) = c_{gsp}^*(x_*, \mu_*, v_*)$ by Corollary 3.1.

Thus, it remains to check that $\mathcal{L}(x, \mu, v, c)$ is a penalty-type merit function for any $\lambda_* = (\mu_*, v_*) \in \Lambda_{loc}(\mathcal{P})$. Then applying the localization principle in the minimax form (Theorem 4.2) one obtains the desired result.

Fix arbitrary $(\mu_*, v_*) \in \Lambda_{loc}(\mathcal{P})$. Let $\{c_n\} \subset (0, +\infty)$ be an increasing unbounded sequence, $x_n \in \arg \min_{x \in A} \mathcal{L}(x, \mu_*, v_*, c_n)$ for all $n \in \mathbb{N}$ and x_* be a cluster point of the sequence $\{x_n\}$. Without loss of generality one can suppose that x_n converges to x_* . Note that $x_* \in A$, since A is closed.

Let x_0 be a globally optimal solution of the problem (\mathcal{P}) . Taking into account the fact that $(\mu_*, v_*) \in \Lambda_{loc}(\mathcal{P})$ and applying Lemma 5.2 one obtains that $\mathcal{L}(x_0, \mu_*, v_*, c) = f(x_0) = f_*$ for any $c > 0$. Hence with the use of the fact that $\mathcal{L}(x, \mu, v, c)$ is nondecreasing in c one gets that

$$\mathcal{L}(x_n, \mu_*, v_*, c) \leq \mathcal{L}(x_n, \mu_*, v_*, c_n) \leq \mathcal{L}(x_0, \mu_*, v_*, c_n) = f_* \quad (5.3)$$

for any $c > 0$ and for any sufficiently large $n \in \mathbb{N}$. Passing to the limit as $n \rightarrow +\infty$ one obtains that $\mathcal{L}(x_*, \mu_*, v_*, c) \leq f_*$ for any $c > 0$ due to the fact that the function $\mathcal{L}(\cdot, \mu, v, c)$ is l.s.c. on A . Let us verify that x_* is feasible. Then passing to the limit as $c \rightarrow +\infty$ with the use of the facts that $\psi(0) = 0$ and $\psi'(t) \rightarrow 0$ as $t \rightarrow -\infty$ (i.e. $\psi(ct)/c \rightarrow 0$ as $c \rightarrow +\infty$ for all $t \leq 0$) one gets that $f(x_*) \leq f_*$, which implies that x_* is a globally optimal solution of the problem (\mathcal{P}) and $\mathcal{L}(x, \mu, v, c)$ is a penalty-type merit function for $(\mu, v) = (\mu_*, v_*)$.

Arguing by reductio ad absurdum, suppose that the point x_* is infeasible. Let us check that $\liminf_{n \rightarrow \infty} c_n^{-1} \langle \mu_*, \Psi(c_n G(x_n)) \rangle \geq 0$. Then taking into account the facts that the eigenvalues $\lambda_i(G(x))$ continuously depend on x due to the continuity of $G(\cdot)$, the function ξ is nonnegative, $\xi(t)/t \rightarrow +\infty$ as $t \rightarrow \infty$ and

$$\frac{1}{c} \text{Tr} \left(\Xi(cG(x)) \right) = \frac{1}{c} \sum_{i=1}^{\ell} \xi(c\lambda_i(G(x))), \quad \langle v_*, h(x) \rangle + \frac{c}{2} \|h(x)\|^2 \geq -\frac{\|v_*\|^2}{2c}$$

one can easily verify that $\mathcal{L}(x_n, \mu_*, \nu_*, c_n) \rightarrow +\infty$ as $n \rightarrow +\infty$, which contradicts (5.3). Thus, x_* is a feasible point.

For any $n \in \mathbb{N}$ define

$$A_n = \frac{1}{c_n} P_n \text{diag} \left(\psi(c_n \lambda_1(G(x_n))), \dots, \psi(c_n \lambda_{k(n)}(G(x_n))), 0, \dots, 0 \right) P_n^T,$$

$$B_n = \frac{1}{c_n} P_n \text{diag} \left(0, \dots, 0, \psi(c_n \lambda_{k(n)+1}(G(x_n))), \dots, \psi(c_n \lambda_\ell(G(x_n))) \right) P_n^T,$$

where $G(x_n) = P_n \text{diag}(\lambda_1(G(x_n)), \dots, \lambda_\ell(G(x_n))) P_n^T$ is a spectral decomposition of the matrix $G(x_n)$ and $k(n) \in \{1, \dots, \ell\}$ is such that $\lambda_{k(n)}(G(x_n)) \geq 0$ and $\lambda_{k(n)+1}(G(x_n)) < 0$. In the case when $G(x_n)$ is negative definite, set $A_n = 0$ and $B_n = c_n^{-1} \Psi(c_n G(x_n))$, while in the case $k(n) = \ell$ put $B_n = 0$.

Observe that $c_n^{-1} \Psi(c_n G(x_n)) = A_n + B_n$. Furthermore, $A_n \in \mathbb{S}_+^\ell$, since ψ is a nondecreasing function and $\psi(0) = 0$. Hence taking into account the fact that by definition $\mu_* \in \mathbb{S}_+^\ell$ one obtains that $\langle \mu_*, A_n \rangle \geq 0$. Therefore it is sufficient to show that $\|B_n\|_F \rightarrow 0$ as $n \rightarrow \infty$.

By definition one has $\lambda_i(G(x_n)) < 0$ for any $i \in \{k(n)+1, \dots, \ell\}$ and

$$\|B_n\|_F = \sqrt{\frac{1}{c_n^2} \sum_{i=k(n)+1}^{\ell} \psi(c_n \lambda_i(G(x_n)))^2} \leq \frac{1}{c_n} \sum_{i=k(n)+1}^{\ell} \psi(c_n \lambda_i(G(x_n))).$$

Consequently, if we check that for any bounded sequence $\{b_n\} \subset (-\infty, 0)$ one has $\psi(c_n b_n)/c_n \rightarrow 0$ as $n \rightarrow \infty$, then $\|B_n\|_F \rightarrow 0$ as $n \rightarrow \infty$ and the proof is complete.

Define $M = \sup_n |b_n|$ and fix arbitrary $\varepsilon > 0$. Since $\psi'(t) \rightarrow 0$ as $t \rightarrow -\infty$, there exists $t_0 > 0$ such that $|\psi(-Mt)/t| < \varepsilon$ for any $t \geq t_0$. Hence taking into account the fact that the function ψ is nondecreasing and $\psi(0) = 0$ one obtains that

$$-\varepsilon < \frac{\psi(-Mt)}{t} \leq \frac{\psi(-st)}{t} \leq 0 \quad \forall s \in [0, M] \quad \forall t \geq t_0.$$

Therefore $|\psi(c_n b_n)/c_n| < \varepsilon$ for any $n \in \mathbb{N}$ large enough, which implies the desired result. \square

Remark 5.3. (i) Suppose that the function ψ is bounded below. Then one can easily verify that the set $\{x \in A \mid \mathcal{L}(x, \mu_*, \nu_*, c_0) < f_*\}$ is bounded for some $c_0 > 0$, provided there exists $\alpha > 0$ such that the set $\{x \in A \mid f(x) < f_* + \alpha, \text{dist}(G(x), \mathbb{S}_-^\ell) < \alpha, \|h(x)\| < \alpha\}$ is bounded and one can find $c > 0$ such that the function $\mathcal{L}(\cdot, \mu_*, \nu_*, c)$ is bounded below on A (see [16, Prp. 6 and Remark 11]).

(ii) Let us note that there is a direct connection between local saddle points of the augmented Lagrangian function $\mathcal{L}(x, \mu, \nu, c)$ and KKT-point of the problem (\mathcal{P}) . Namely, if (x_*, μ_*, ν_*) is a local saddle point of $\mathcal{L}(x, \mu, \nu, c)$, the set A is convex, and the functions f , G and h are twice differentiable at x_* , then (x_*, μ_*, ν_*) is a KKT-point of the problem (\mathcal{P}) satisfying standard second order necessary optimality conditions for semidefinite programs. Conversely, if (x_*, μ_*, ν_*) is a KKT-point of the problem (\mathcal{P}) satisfying second order sufficient optimality conditions, then (x_*, μ_*, ν_*) is a local saddle of the augmented Lagrangian $\mathcal{L}(x, \mu, \nu, c)$ (see [12, 16] for more details).

5.3. Example III: A nonlinear augmented Lagrangian for constrained minimax problems.

In this section we introduce a new nonlinear augmented Lagrangian function for constrained minimax problems and obtain necessary and sufficient conditions for the existence of a global

saddle point of this function with the use of the general theory of globally minimax exact merit functions. We also demonstrate how the theoretical results developed in this paper allow one to better understand limitations of some augmented Lagrangian methods. Unlike augmented Lagrangian functions from the previous examples, the augmented Lagrangian from this section nonlinearly depends on the objective function. The results of this subsection were inspired by the paper [27].

Let $X = \mathbb{R}^d$ and $f(x) = \max_{1 \leq j \leq m} f_j(x)$, where $f_j: \mathbb{R}^d \rightarrow \mathbb{R}$ are given functions. Suppose also that $M = \{x \in \mathbb{R}^d \mid h(x) = 0\}$, where $h: \mathbb{R}^d \rightarrow \mathbb{R}^s$. For the sake of simplicity we suppose that there are no inequality constraints, but the results below can be easily extended to the case of minimax problems with both equality and inequality constraints.

Thus, the problem (\mathcal{P}) is an equality constrained minimax problem, which, as is well-known and easy to check, is equivalent to the following mathematical programming problem:

$$\min_{(x,z)} z \quad \text{s.t.} \quad f_j(x) - z \leq 0, \quad j \in J, \quad h(x) = 0, \quad x \in A. \quad (5.4)$$

Here $J = \{1, \dots, m\}$. Let us introduce an augmented Lagrangian function for this problem first. We incorporate the constraints $f_j(x) - z \leq 0$ into this augmented Lagrangian with the use of the exponential penalty function [3, 8, 55, 59, 60], while the equality constraints are handled via the Hestenes-Powell augmented Lagrangian [61, 63, 62]. Let us note that the choice of the exponential penalty function for the constraints $f_j(x) - z \leq 0$ was inspired by [27] and it will allow us to obtain a simple analytic expression for the augmented Lagrangian function for the problem (\mathcal{P}) .

For any $\alpha \in \mathbb{R}_+^m$ and $\mu \in \mathbb{R}^s$, define

$$L(x, z, \alpha, \mu, c) = z + \frac{1}{c} \sum_{j=1}^m \alpha^{(j)} (e^{c(f_j(x) - z)} - 1) + \langle \mu, h(x) \rangle + \frac{c}{2} \|h(x)\|^2,$$

where $\mathbb{R}_+ = [0, +\infty)$. Let us impose a natural additional assumption on the multipliers $\alpha^{(j)}$. Namely, hereinafter we suppose that α belongs to the standard simplex, which we denote by $S = \{\alpha \in \mathbb{R}_+^m \mid \alpha^{(1)} + \dots + \alpha^{(m)} = 1\}$. Define $\Lambda = S \times \mathbb{R}^s$.

Minimizing the function $L(x, z, \alpha, \mu, c)$ in z , we arrive at the following augmented Lagrangian function for the original constrained minimax problem:

$$\mathcal{L}(x, \alpha, \mu, c) = \frac{1}{c} \ln \left(\sum_{j=1}^m \alpha^{(j)} e^{c f_j(x)} \right) + \langle \mu, h(x) \rangle + \frac{c}{2} \|h(x)\|^2.$$

Let us note that the first term in the expression above is very similar to the well-known smoothing regularization of the max-function (see, e.g., [64]). The introduction of the multipliers $\alpha^{(j)}$ allows one to minimize the augmented Lagrangian $\mathcal{L}(x, \alpha, \mu, v, c)$ with a finite fixed value of the penalty parameter $c > 0$, while the direct usage of smoothing regularization methods might result in ill-conditioning due to the need to unboundedly increase the parameter c (see the discussion in [27]). However, note that adaptive smoothing techniques might help to avoid ill-conditioning [65]). On the other hand, the introduction of the multipliers $\alpha^{(j)}$ imposes some additional restrictions on the problem (\mathcal{P}) , which can be avoided with the use of smoothing regularization methods (see Lemma 5.6 below).

Remark 5.4. It is possible to directly apply the theory of globally minimax exact merit function to the augmented Lagrangian $L(x, z, \alpha, \mu, c)$ for problem (5.4). However, one can verify that the

quadruplet $(x_*, z_*, \alpha_*, \mu_*)$ is a global saddle point of $L(x, z, \alpha, \mu, c)$ iff the triplet (x_*, α_*, μ_*) is a global saddle point of $\mathcal{L}(x, \alpha, \mu, c)$ and z_* coincides with the optimal value of the problem (\mathcal{P}) . Therefore the augmented Lagrangian $L(x, z, \alpha, \mu, c)$ is globally minimax exact iff the augmented Lagrangian $\mathcal{L}(x, \alpha, \mu, c)$ is globally minimax exact.

Note that for any $x \in \mathbb{R}^d$, $\alpha \in S$, and $c > 0$ one has

$$f(x, \alpha, c) := \frac{1}{c} \ln \left(\sum_{j=1}^m \alpha^{(j)} e^{cf_j(x)} \right) \leq f(x), \quad (5.5)$$

and this inequality turns into an equality if and only if the complementarity condition, that is, $\alpha^{(j)}(f_j(x) - f(x)) = 0$, $j \in J$, holds true. Hence, in particular, for any feasible point x and all $\lambda = (\alpha, \mu) \in \Lambda$ and $c > 0$ one has $\mathcal{L}(x, \alpha, \mu, c) \leq f(x)$. Let us point out some other useful properties of the augmented Lagrangian function $\mathcal{L}(x, \alpha, \mu, c)$.

Lemma 5.4. *The function $\mathcal{L}(x, \alpha, \mu, c)$ is nondecreasing in c .*

Proof. Clearly, it is sufficient to prove that the function $f(x, \alpha, c)$ is nondecreasing in c . To this end, observe that for any continuously differentiable function $q: \mathbb{R}_+ \rightarrow \mathbb{R}$, the function $c \mapsto q(c)/c$ is nondecreasing on $(0, +\infty)$, in particular, if the function q is convex. In turn, the convexity of the function $q(c) = cf(x, \alpha, c)$ follows directly from the convexity of the function $p(x) = \ln(\alpha_1 e^{x_1} + \dots + \alpha_m e^{x_m})$, which can be easily verified by computing the Hessian of this function. \square

Lemma 5.5. *Let $(x_*, \alpha_*, \mu_*) \in A \times \Lambda$ be such that*

$$\max_{(\alpha, \mu) \in \Lambda} \mathcal{L}(x_*, \alpha, \mu, c_0) \leq \mathcal{L}(x_*, \alpha_*, \mu_*, c_0) < +\infty \quad (5.6)$$

for some $c_0 > 0$. Then x_ is a feasible point of the problem (\mathcal{P}) , the complementarity condition $\alpha_*^{(j)}(f_j(x_*) - f(x_*)) = 0$, $j \in J$, holds true, and $\mathcal{L}(x_*, \alpha_*, \mu_*, c) = f(x_*)$ for all $c > 0$.*

Proof. The feasibility of x_* is proved in the same way as in Lemmas 5.2 and 5.3. Let us check that the complementarity condition holds true. Then one obtains the required result.

Suppose that $\alpha_*^{(j)}(f_j(x_*) - f(x_*)) \neq 0$ for some $j \in J$. Then $\alpha_*^{(j)} e^{c_0 f_j(x_*)} < \alpha_*^{(j)} e^{c_0 f(x_*)}$, which implies that $f(x_*, \alpha_*, c_0) < f(x_*)$. Therefore taking into account the feasibility of x_* one obtains that $\mathcal{L}(x_*, \alpha_*, \mu_*, c_0) < f(x_*)$. On the other hand, setting $\alpha^{(k)} = 1$ for some $k \in J$ such that $f_k(x) = f(x)$ and $\alpha^{(s)} = 0$ for any $s \neq k$ and applying (5.6) one gets

$$f(x_*) = \mathcal{L}(x_*, \alpha, 0, c_0) \leq \mathcal{L}(x_*, \alpha_*, \mu_*, c_0) < f(x_*),$$

which is impossible. \square

Applying Lemmas 5.4 and 5.5 and arguing in the same way as in the proofs of Theorems 5.1 and 5.2 one can try to extend the localization principle to the case of the augmented Lagrangian $\mathcal{L}(x, \alpha, \mu, c)$. However, one can check that the proofs of these theorems cannot be directly extended to the case of the augmented Lagrangian $\mathcal{L}(x, \alpha, \mu, c)$ due to the fact that this augmented Lagrangian is not a penalty-type merit function for all $\lambda_* \in \Lambda_{loc}(\mathcal{P})$ in the general case. The following simple result helps us to understand an additional assumption that one has to impose on the problem (\mathcal{P}) in order to extend the localization principle to the case of the function $\mathcal{L}(x, \alpha, \mu, c)$.

Let $J(x) = \{j \in J \mid f_j(x) = f(x)\}$ and for all $\alpha \in S$ denote $\text{supp}(\alpha) = \{j \in J \mid \alpha^{(j)} > 0\}$. Finally, define $g(x; y, \alpha) = \max_{j \in J(y) \cap \text{supp}(\alpha)} f_j(x)$ for all $y \in \mathbb{R}^d$ and $\alpha \in S$ satisfying the condition $J(y) \cap \text{supp}(\alpha) \neq \emptyset$. Note that if the functions f_j are continuous and $J(y) \subseteq \text{supp}(\alpha)$, then $g(x; y, \alpha) = f(x)$ for any x in a neighbourhood of y . Furthermore, $g(x; y, \alpha) \leq f(x)$ for all $x, y \in \mathbb{R}^d$. Consider the following relaxation of the problem (\mathcal{P}) :

$$\min_x g(x; y, \alpha) \quad \text{s.t.} \quad h(x) = 0, \quad x \in A. \quad (\mathcal{P}(y, \alpha))$$

Denote by $\Omega_*(y, \alpha)$ the set of globally optimal solutions of this problem.

Lemma 5.6. *Let (x_*, α_*, μ_*) be a global saddle point of $\mathcal{L}(x, \alpha, \mu, c)$. Then $\Omega_* \subseteq \Omega_*(y_*, \alpha_*)$ for any $y_* \in \Omega_*$.*

Proof. Taking into account Lemma 5.5 and inequality (5.5) one obtains that

$$f(x_*) = \mathcal{L}(x_*, \alpha_*, \mu_*, c) \leq \mathcal{L}(x, \alpha_*, \mu_*, c) \leq f(x) \quad \forall c > c_{gsp}^*(x_*, \alpha_*, \mu_*)$$

for any feasible point x , which yields that $x_* \in \Omega_*$. Therefore for any $y_* \in \Omega_*$ the triplet (y_*, α_*, μ_*) is a global saddle point of $\mathcal{L}(x, \alpha, \mu, c)$ by Corollary 3.1.

Let $y_*, w_* \in \Omega_*$ be arbitrary. By Lemma 5.5 for any $j \in J$ one has $\alpha_*^{(j)}(f_j(y_*) - f(y_*)) = 0$. Therefore

$$f(x, \alpha_*, c) = \frac{1}{c} \ln \left(\sum_{j \in J(y_*)} \alpha_*^{(j)} e^{cf_j(x)} \right) \leq g(x; y_*, \alpha_*) \quad \forall x \in \Omega,$$

which implies that $\mathcal{L}(x, \alpha_*, \mu_*, c) \leq g(x; y_*, \alpha_*)$ for any feasible x . Hence applying Lemma 5.5 once again one gets that

$$g(w_*; y_*, \alpha_*) \leq f(w_*) = f(y_*) = \mathcal{L}(y_*, \alpha_*, \mu_*, c) \leq \mathcal{L}(x, \alpha_*, \mu_*, c) \leq g(x; y_*, \alpha_*)$$

for all $c > c_{gsp}^*(y_*, \alpha_*, \mu_*)$ and for any $x \in \Omega$. Thus, w_* is a globally optimal solution of the problem $(\mathcal{P}(y_*, \alpha_*))$, that is, $\Omega_* \subseteq \Omega_*(y_*, \alpha_*)$. \square

Let us explain the meaning of the previous lemma. Suppose that (x_*, α_*, μ_*) is a global saddle point of $\mathcal{L}(x, \alpha, \mu, c)$. Then, as the lemma above states, the removal of those functions $f_j(x)$, $j \in J$, that are inactive at a globally optimal solution y_* of the problem (\mathcal{P}) (i.e. $j \notin J(y_*)$) and those functions $f_j(x)$, $j \in J(y_*)$, that correspond to zero multipliers $\alpha_*^{(j)}$, roughly speaking, must not change the set of globally optimal solutions of the minimax problem under consideration. More precisely, every globally optimal solution of the problem (\mathcal{P}) must be a globally optimal solution of the relaxed problem $(\mathcal{P}(y_*, \alpha_*))$. Let us note that this assumption is very restrictive for nonconvex problems, but one can verify that it always holds true in the convex case.

Remark 5.5. Let us note that Lemma 5.6 exposes the main drawback of the exponential penalty function. Namely, for the existence of a global saddle point of the exponential penalty function it is necessary that the removal of inactive inequality constraints does not change the set of globally optimal solutions of the problem under consideration (cf. [16], Example 18, Proposition 8 and Remark 14).

With the use of Lemma 5.6 we can prove the localization principle for the augmented Lagrangian function $\mathcal{L}(x, \alpha, \mu, c)$.

Theorem 5.3 (Localization Principle). *Let A be closed and the functions f_j , $j \in J$, and h be continuous on A . Then for the existence of a global saddle point of the augmented Lagrangian $\mathcal{L}(x, \alpha, \mu, c)$ (or, equivalently, for the global minimax exactness of $\mathcal{L}(x, \alpha, \mu, c)$ with the optimal value f_*) it is necessary that there exists $(\alpha_*, \mu_*) \in \Lambda$ and $c_0 > 0$ such that*

- (1) (x_*, α_*, μ_*) is a local saddle point of $\mathcal{L}(x, \alpha, \mu, c)$ for all $x_* \in \Omega_*$;
- (2) $\Omega_* \subseteq \Omega_*(x_*, \alpha_*)$ for any $x_* \in \Omega_*$;
- (3) the set $S(c_0, \alpha_*, \mu_*) := \{x \in A \mid \mathcal{L}(x, \alpha_*, \mu_*, c_0) < f_*\}$ is bounded.

Moreover, these conditions become sufficient, if $\Omega_* = \Omega_*(x_*, \alpha_*)$ for any $x_* \in \Omega_*$.

Proof. Let (x_*, α_*, μ_*) be a global saddle point of $\mathcal{L}(x, \alpha, \mu, c)$. Then, as was shown in the proof of Lemma 5.6, one has $x_* \in \Omega_*$ and for any $y_* \in \Omega_*$ the triplet (y_*, α_*, μ_*) is a global saddle point of $\mathcal{L}(x, \alpha, \mu, c)$ and $c_{gsp}^*(x_*, \alpha_*, \mu_*) = c_{gsp}^*(y_*, \alpha_*, \mu_*)$ by Corollary 3.1. Furthermore, $\Omega_* \subseteq \Omega_*(x_*, \alpha_*)$ for any $x_* \in \Omega_*$ due to Lemma 5.6 and the set $S(c, \alpha_*, \mu_*)$ is empty for all $c > c_{gsp}^*(x_*, \alpha_*, \mu_*)$ by virtue of Lemma 5.5. Thus, the “necessity” part of the theorem is proved.

Let us prove the converse statement. To this end, let us check that $\mathcal{L}(x, \alpha, \mu, c)$ is a *penalty-like* merit function for $\lambda = (\alpha_*, \mu_*)$. Then applying the first version of the localization principle in the minimax form (Theorem 4.1) one obtains the required result.

Let $\{c_n\} \subset (c_0, +\infty)$ be an increasing unbounded sequence. From the boundedness of the set $S(c_0, \alpha_*, \mu_*)$ and the continuity of the functions f_j and h it follows that for any $n \in \mathbb{N}$ there exists $x_n \in \arg \min_{x \in A} \mathcal{L}(x, \alpha_*, \mu_*, c_n)$. If $S(c_n, \alpha_*, \mu_*) = \emptyset$ for some $n \in \mathbb{N}$, then one can set $x_n = x_*$ for some fixed $x_* \in \Omega_*$ (note that in this case $\mathcal{L}(x_*, \alpha_*, \mu_*, c) = f(x_*) = f_*$, since $(\alpha_*, \mu_*) \in \Lambda_{loc}(\mathcal{P})$ by our assumption). The sequence $\{x_n\}$ is bounded, since $S(\tau, \alpha_*, \mu_*) \subseteq S(c, \alpha_*, \mu_*)$ for any $\tau > c$ by Lemma 5.4. Therefore, without loss of generality one can suppose that the sequence x_n converges to a point x_* , which belongs to the set A due to the fact that this set is closed.

From Lemma 5.4 and the definition of x_n , it follows that

$$\mathcal{L}(x_n, \alpha_*, \mu_*, c) \leq \mathcal{L}(x_n, \alpha_*, \mu_*, c_n) \leq f_* \quad (5.7)$$

for any $c > 0$ and for any sufficiently large $n \in \mathbb{N}$. Passing to the limit as $n \rightarrow +\infty$, one gets that

$$\mathcal{L}(x_*, \alpha_*, \mu_*, c) \leq f_* \quad \forall c > 0 \quad (5.8)$$

due to the continuity of the function f_j and h on the set A .

Note that x_* is feasible, since otherwise $\mathcal{L}(x_*, \alpha_*, \mu_*, c) \rightarrow +\infty$ as $c \rightarrow +\infty$, which contradicts (5.8). Therefore, passing to the limit superior as $c \rightarrow +\infty$ in inequality (5.8), one obtains that $\limsup_{c \rightarrow +\infty} f(x_*, \alpha_*, c) \leq f_*$.

Let y_* be a globally optimal solution of the problem (\mathcal{P}) . Applying Lemma 5.5 (recall that $(\alpha_*, \mu_*) \in \Lambda_{loc}(\mathcal{P})$), one obtains that $\alpha_*^{(j)} = 0$ for any $j \notin J(y_*)$. Let $j_0 \in J(y_*) \cap \text{supp}(\alpha_*)$ be such that $f_{j_0}(x_*) = g(x_*, y_*, \alpha_*)$, i.e. $f_{j_0}(x_*) \geq f_j(x_*)$ for all $j \in J(y_*) \cap \text{supp}(\alpha_*)$. Then

$$\begin{aligned} f(x_*, \alpha_*, c) &= \frac{1}{c} \ln \left(\sum_{j \in J(y_*) \cap \text{supp}(\alpha_*)} \alpha_*^{(j)} e^{cf_j(x_*)} \right) \\ &= g(x_*, y_*, \alpha_*) + \frac{1}{c} \ln \left(\sum_{j \in J(y_*) \cap \text{supp}(\alpha_*)} \alpha_*^{(j)} e^{c(f_j(x_*) - f_{j_0}(x_*))} \right) \geq g(x_*, y_*, \alpha_*). \end{aligned}$$

Hence the inequality $g(x_*; y_*, \alpha_*) \leq f_*$ holds true, which yields $x_* \in \Omega_*(y_*, \alpha_*) = \Omega_*$, since $g(y_*; y_*, \alpha_*) = f(y_*) = f_*$ and $y_* \in \Omega_* = \Omega_*(y_*, \alpha_*)$ by our assumption. Thus, $\mathcal{L}(x, \alpha, \mu, c)$ is a penalty-like merit function for $\lambda = (\alpha_*, \mu_*)$. \square

Remark 5.6. Let us note that the existence of a local saddle point of the augmented Lagrangian $\mathcal{L}(x, \alpha, \mu, v, c)$ can be easily proved with the use of the same second order sufficient optimality conditions as in [27]. For the sake of shortness, we leave a proof of this result to the interested reader.

6. CONCLUSIONS

In this paper we presented a general theory of globally minimax exact merit function for finite dimensional constrained optimization problems. Under some natural assumptions the equivalence between global minimax exactness, the existence of global saddle points and the existence of solutions of the augmented dual problem was established. We obtained simple necessary and sufficient conditions for the global minimax exactness of merit functions in the form of the localization principle and applied this principle in order to obtain simple necessary and sufficient conditions for the existence of global saddle points of He-Wu-Meng's augmented Lagrangian for inequality constrained problems, nonlinear rescaling Lagrangians for nonconvex semidefinite problems, and a new class of nonlinear augmented Lagrangian functions for constrained minimax problems.

REFERENCES

- [1] A. Shapiro, J. Sun, Some properties of the augmented Lagrangian in cone constrained optimization, *Math. Oper. Res.* 29 (2004), 479–491.
- [2] Y. Y. Zhou, J. C. Zhou, X. Q. Yang, Existence of augmented Lagrange multipliers for cone constrained optimization problems, *J. Glob. Optim.* 58 (2014), 243–260.
- [3] Q. Liu, X. Yang, Zero duality and saddle points of a class of augmented Lagrangian functions in constrained non-convex optimization. *Optim.* 57 (2008), 655–667.
- [4] Q. Liu, W. M. Tang, X. M. Yang, Properties of saddle points for generalized augmented Lagrangian, *Math. Meth. Oper. Res.* 69 (2009), 111–124.
- [5] H. Z. Luo, G. Mastroeni, H. X. Wu, Separation approach for augmented Lagrangians in constrained nonconvex optimization, *J. Optim. Theory Appl.* 144 (2010), 275–290.
- [6] J. Zhou, N. Xiu, C. Wang, Saddle point and exact penalty representation for generalized proximal Lagrangians, *J. Glob. Optim.* 56 (2012), 669–687.
- [7] H. X. Wu, H. Z. Luo, Saddle points of general augmented Lagrangians for constrained nonconvex optimization, *J. Glob. Optim.* 53 (2012), 683–697.
- [8] X. L. Sun, D. Li, K. I. M. McKinnon, On saddle points of augmented Lagrangians for constrained nonconvex optimization, *SIAM J. Optim.* 15 (2005), 1128–1146.
- [9] C. Wang, Q. Liu, B. Qu, Global saddle points of nonlinear augmented Lagrangian functions, *J. Glob. Optim.* 68 (2017), 125–146.
- [10] C. Wang, J. Zhou, X. Xu, Saddle points theory of two classes of augmented Lagrangians and its applications to generalized semi-infinite programming, *Appl. Math. Optim.* 59 (2009), 413–434.
- [11] J. Zhou, J. S. Chen, On the existence of saddle points for nonlinear second-order cone programming problems, *J. Glob. Optim.* 62 (2015), 459–480.
- [12] H. Luo, H. Wu, J. Liu, On saddle points in semidefinite optimization via separation scheme, *J. Optim. Theory Appl.* 165 (2015), 113–150.
- [13] H. X. Wu, H. Z. Luo, J. F. Yang, Nonlinear separation approach for the augmented Lagrangian in nonlinear semidefinite programming, *J. Glob. Optim.* 59 (2014), 695–727.

- [14] J.-J. Rückmann, A. Shapiro, Augmented Lagrangians in semi-infinite programming, *Math. Program., Ser. B.*, 116 (2009), 499–512.
- [15] R. S. Burachik, X. Q. Yang, Y. Y. Zhou, Existence of augmented Lagrange multipliers for semi-infinite programming problems, *J. Optim. Theory Appl.* 173 (2017), 471–503.
- [16] M. V. Dolgopolik, Augmented Lagrangian functions for cone constrained optimization: the existence of global saddle points and exact penalty property, *J. Glob. Optim.* 71 (2018), 237–296.
- [17] R. T. Rockafellar, R. J.-B. Wets. *Variational Analysis*, Springer-Verlag, Berlin, 1998.
- [18] X. X. Huang, X. Q. Yang, A unified augmented Lagrangian approach to duality and exact penalization, *Math. Oper. Res.* 28 (2003), 533–552.
- [19] X. X. Huang, X. Q. Yang. Further study on augmented Lagrangian duality theory, *J. Glob. Optim.*, 31 (2005), 193–210.
- [20] Y. Y. Zhou, X. Q. Yang, Duality and penalization in optimization via an augmented Lagrangian function with applications, *J. Optim. Theory Appl.* 140 (2009), 171–188.
- [21] R. S. Burachik, A. N. Iusem, J. G. Melo, Duality and exact penalization for general augmented Lagrangians, *J. Optim. Theory Appl.* 147 (2010), 125–140.
- [22] Y. Y. Zhou, X. Q. Yang. Augmented Lagrangian functions for constrained optimization problems, *J. Glob. Optim.* 52 (2012), 95–108.
- [23] C. Y. Wang, X. Q. Yang, X. M. Yang, Nonlinear augmented Lagrangian and duality theory, *Math. Oper. Res.* 38 (2012), 740–760.
- [24] M. V. Dolgopolik, Existence of augmented Lagrange multipliers: reduction to exact penalty functions and localization principle, *Math. Program.* 166 (2017), 297–326.
- [25] C. Kan, W. Song, Augmented Lagrangian duality for composite optimization problems, *J. Optim. Theory Appl.* 165 (2015), 763–784.
- [26] C. Kan, W. Song, Second-order conditions for existence of augmented Lagrange multipliers for eigenvalue composite optimization problems, *J. Glob. Optim.* 63 (2015), 77–97.
- [27] S. He, S. Zhou, A nonlinear augmented Lagrangian for constrained minimax problems, *Appl. Math. Comput.* 218 (2011), 4567–4579.
- [28] M. V. Dolgopolik, A unified approach to the global exactness of penalty and augmented Lagrangian functions I: parametric exactness, *J. Optim. Theory Appl.* 176 (2018), 728–744.
- [29] M. V. Dolgopolik, A unified approach to the global exactness of penalty and augmented Lagrangian functions II: extended exactness, *J. Optim. Theory Appl.* 176 (2018), 745–762.
- [30] G. Di Pillo, L. Grippo. On the exactness of a class of nondifferentiable penalty functions, *J. Optim. Theory Appl.* 57 (1998), 399–410.
- [31] V. F. Demyanov, Nonsmooth optimization. In G. Di Pillo and F. Schoen, editors, *Nonlinear Optimization. Lecture Notes in Mathematics*, vol. 1989, pp. 55–164. Springer-Verlag, Berlin-Heidelberg, 2010.
- [32] A. J. Zaslavski, *Optimization on Metric and Normed Spaces*, Springer, New York, 2010.
- [33] Z. Y. Wu, F. S. Bai, X. Q. Yang, L. S. Zhang, An exact lower order penalty function and its smoothing in nonlinear programming, *Optim.* 53 (2004), 51–68.
- [34] A. M. Rubinov, B. M. Glover, X. Q. Yang, Decreasing functions with applications to penalization, *SIAM J. Optim.*, 10 (1999), 289–313.
- [35] A. M. Rubinov, X. Q. Yang, A. M. Bagirov, Penalty functions with a small penalty parameter, *Optim. Methods Softw.*, 17 (2002), 931–964.
- [36] A. M. Rubinov, X. Q. Yang, *Lagrange-Type Functions in Constrained Non-Convex Optimization*, Kluwer Academic Publishers, Dordrecht, 2003.
- [37] R. Fletcher, A class of methods for nonlinear programming with termination and convergence properties. In J. Abadie, (ed), *Integer and nonlinear programming*, pp. 157–173, North-Holland, Amsterdam, 1970.
- [38] R. Fletcher, An exact penalty function for nonlinear programming with inequalities, *Math. Program.* 5 (1973), 129–150.
- [39] S. Lucidi, New results on a continuously differentiable exact penalty function, *SIAM J. Optim.* 2 (1992), 558–574.
- [40] E. H. Fukuda, P. J. S. Silva, M. Fukushima, Differentiable exact penalty functions for nonlinear secondorder cone programs, *SIAM J. Optim.* 22 (2012), 1607–1633.

- [41] G. Di Pillo, L. Grippo, A new class of augmented Lagrangians in nonlinear programming, *SIAM J. Control Optim.* 17 (1979), 618–628.
- [42] G. Di Pillo, S. Lucidi, An augmented Lagrangian function with improved exactness properties, *SIAM J. Optim.* 12 (2001), 376–406.
- [43] H. Luo, H. Wu, J. Liu, Some results on augmented Lagrangians in constrained global optimization via image space analysis, *J. Optim. Theory Appl.* 159 (2013), 360–385.
- [44] W. Huyer, A. Neumaier, A new exact penalty function, *SIAM J. Optim.* 13 (2003), 1141–1158.
- [45] C. Wang, C. Ma, J. Zhou, A new class of exact penalty functions and penalty algorithms, *J. Glob. Optim.* 58 (2014), 51–73.
- [46] S. He, L. Wu, H. Meng, A nonlinear Lagrangian for constrained optimization problems, *J. Appl. Math. Comput.* 38 (2012), 669–685.
- [47] M. Stingl, On the solution of nonlinear semidefinite programs by augmented Lagrangian methods. PhD thesis, Institute of Applied Mathematics II, Friedrich-Alexander University of Erlangen-Nuremberg, Erlangen, Germany, 2006.
- [48] D. Noll, Local convergence of an augmented Lagrangian method for matrix inequality constrained programming, *Optim. Methods Softw.* 22 (2007), 777–802.
- [49] Y. Li, L. Zhang, A new nonlinear Lagrangian method for nonconvex semidefinite programming, *J. Appl. Anal.* 15 (2009), 149–172.
- [50] L. Zhang, Y. Li, J. Wu, Nonlinear rescaling Lagrangians for nonconvex semidefinite programming, *Optim.* 63 (2014), 899–920.
- [51] H. Yamashita, H. Yabe, A survey of numerical methods for nonlinear semidefinite programming, *J. Oper. Res. Soc. Jpn.*, 58 (2015), 24–60.
- [52] R. Polyak, Modified barrier functions (theory and methods), *Math. Program.* 54 (1992), 177–222.
- [53] R. Polyak, Nonlinear rescaling vs. smoothing technique in convex optimization, *Math. Program.* 92 (2002), 197–235.
- [54] I. Griva, R. A. Polyak, Primal-dual nonlinear rescaling method with dynamic parameter update, *Math. Program.* 106 (2006), 237–259.
- [55] C.-Y. Wang, D. Li, Unified theory of augmented Lagrangian methods for constrained global optimization, *J. Glob. Optim.* 44 (2009), 433–458.
- [56] D. Sun, J. Sun, Lowner’s operator and spectral functions in Euclidean Jordan algebras, *Math. Oper. Res.* 33 (2008), 421–445.
- [57] A. Shapiro, On differentiability of symmetric matrix valued functions. Technical report, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA, 2002.
- [58] X. Chen, H. Qi, P. Tseng, Analysis of nonsmooth symmetric-matrix-valued functions with applications to semidefinite complementarity problems, *SIAM J. Optim.* 13 (2003), 960–985.
- [59] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.
- [60] P. Tseng, D. P. Bertsekas, On the convergence of the exponential multiplier method for convex programming, *Math. Program.* 60 (1993), 1–19.
- [61] R. T. Rockafellar, Augmented Lagrange multiplier functions and duality in nonconvex programming, *SIAM J. Control Optim.* 12 (1974), 268–285.
- [62] E. G. Birgin, J. M. Martinez, *Practical Augmented Lagrangian Methods for Constrained Optimization*, SIAM, Philadelphia, 2014.
- [63] R. T. Rockafellar, Lagrange multipliers and optimality, *SIAM Rev.* 35 (1993), 183–238.
- [64] S. Xu, Smoothing method for minimax problems, *Comput. Optim. Appl.* 20 (2001), 267–279.
- [65] E. Polak, J. O. Royset, R. S. Womersley, Algorithms with adaptive smoothing for finite minimax problems, *J. Optim. Theory Appl.* 119 (2003), 459–484.