Augmented Lagrangian Functions for Cone Constrained Optimization: the Existence of Global Saddle Points and Exact Penalty Property

M.V. Dolgopolik

November 16, 2018

Abstract

In this article we present a general theory of augmented Lagrangian functions for cone constrained optimization problems that allows one to study almost all known augmented Lagrangians for these problems within a unified framework. We develop a new general method for proving the existence of global saddle points of augmented Lagrangian functions, called the localization principle. The localization principle unifies, generalizes and sharpens most of the known results on the existence of global saddle points, and, in essence, reduces the problem of the existence of global saddle points to a local analysis of optimality conditions. With the use of the localization principle we obtain first necessary and sufficient conditions for the existence of a global saddle point of an augmented Lagrangian for cone constrained minimax problems via both second and first order optimality conditions. In the second part of the paper, we present a general approach to the construction of globally exact augmented Lagrangian functions. The general approach developed in this paper allowed us not only to sharpen most of the existing results on globally exact augmented Lagrangians, but also to construct first globally exact augmented Lagrangian functions for equality constrained optimization problems, for nonlinear second order cone programs and for nonlinear semidefinite programs. These globally exact augmented Lagrangians can be utilized in order to design new superlinearly (or even quadratically) convergent optimization methods for cone constrained optimization problems.

1 Introduction

The main goal of this article is to present a general theory of augmented Lagrangian functions for cone constrained optimization problems that provides a theoretical foundation for the development of augmented Lagrangian methods for these problems. In recent years, several attempts were made to develop a general theory of Lagrangian functions. A general duality theory for nonlinear Lagrangian functions for mathematical programming problems was developed in [69, 84]. In [35, 47, 100, 101], the image space analysis was applied to the study of duality theory for augmented Lagrangian functions, while in [5, 85, 81] some general classes of augmented Lagrangian functions constructed from the Rockafellar-Wets augmented Lagrangian ([68], Section 11.K) were studied. A

unified theory of augmented Lagrangian methods for mathematical programming problems was presented in [83]. However, there is no satisfactory and general enough theory of the existence of global saddle points that can be applied to various augmented Lagrangians for various cone constrained optimization problems. Furthermore, there are no results on *exact* augmented Lagrangian functions for cone constrained optimization problems. Our aim is to fill these gaps and develop a general theory containing simple and easily verifiable necessary and sufficient conditions for the existence of global saddle points of augmented Lagrangian functions, and for the global exactness of these function.

In this paper, instead of utilizing a modification of the Rockafellar-Wets augmented Lagrangian, we consider a more straightforward approach to the theory of augmented Lagrangian functions (that is very similar to the one used in the image space analysis), in which an augmented Lagrangian is defined simply as the sum of the objective function and a convolution function depending on constraints, multipliers and penalty parameter, and satisfying some general assumptions (axioms). The main advantage of this approach is the fact that it does not rely on the particular structure of the augmented Lagrangian, and allows one to include almost all particular cases into the general theory.

The axiomatic approach that we use was inspired by [50]. It provides one with a simple and unified framework for the study of various augmented Lagrangian functions, such as the Hestenes-Powell-Rockafellar augmented Lagrangian [2, 66, 67], the cubic augmented Lagrangian [42], Mangasarian's augmented Lagrangian [59, 89], the exponential penalty function [1, 80, 79, 50, 83], the Log-Sigmoid Lagrangian [64, 65], the penalized exponential-type augmented Lagrangians [1, 79, 50, 83], the modified barrier functions [63, 79, 50, 83], the p-th power augmented Lagrangian [43, 44, 91, 46, 45, 88, 50], He-Wu-Meng's augmented Lagrangian [36], extensions of the Hestenes-Powell-Rockafellar augmented Lagrangian to the case of nonlinear second order cone programs [51, 52, 97], nonlinear semidefinite programs [40, 78, 76, 96, 77, 86, 57, 87, 90, 92], and semi-infinite programs [70, 41, 73, 6], as well as extensions of the exponential penalty function and the modified barrier functions to the case of nonlinear second order cone programs [93], and nonlinear semidefinite programs [74, 60, 48, 94, 92, 55].

The first part of the paper is devoted to the problem of the existence of global saddle points of augmented Lagrangian functions, which is important for convergence analysis of augmented Lagrangian methods, since the existence of a global saddle point is, usually, necessary for the global convergence of these methods. The problem of the existence of global saddle points was studied for general cone constrained optimization problems in [72, 99], for mathematical programming problems in [50, 49, 56, 98, 89, 82, 79, 81], for nonlinear second order cone programming problems in [97], for nonlinear semidefinite programming problems in [90, 55], and for semi-infinite programming problems in [70, 6]. The analysis of the known results on the existence of global saddle points of augmented Lagrangian functions indicates that the same, in essence, results are proved and reproved multiple times in different settings.

In this article, we propose a unified approach to the study of global saddle points of augmented Lagrangian functions, called *the localization principle.* The localization principle was first developed by the author for the study of exact linear penalty functions [27, 30] and augmented Lagrange multipliers [28]. A modification of this principle presented in this paper provides first simple nec-

essary and sufficient conditions for the existence of global saddle points that unify, generalize and sharpen almost all known results in this area (see Remark 10). Furthermore, the localization principle reduces the study of global saddle points to a local analysis of sufficient optimality conditions. With the use of this principle we obtained simple necessary and sufficient conditions for the existence of global saddle points for cone constrained *minimax* problems via both second order and first order (the so-called *alternance* conditions; see [58, 10, 9, 11, 13, 12]) sufficient optimality conditions. To the best of authors knowledge, the problem of the existence of saddle points for cone constrained minimax problems has never been studied before. We also provide simple sufficient conditions for the existence of local saddle points that unify and sharpen many existing result (see Remark 16).

As it is well known, standard augmented Lagrangian methods converge only linearly (see, e.g., [2, 74, 60, 52, 76, 93, 94] and references therein), and in order to apply them one has to minimize an augmented Lagrangian function numerous times. In order to overcome this difficulty, Di Pillo and Grippo [14] proposed to consider the so-called *exact* augmented Lagrangian functions. Exact augmented Lagrangian functions are constructed in such a way that it is necessary to minimize them only once (but *simultaneously* in primal and dual variables) in order to recover KKT-points corresponding to globally optimal solutions of the initial optimization problem. Furthermore, one can design superlinearly and even quadratically convergent methods for minimizing exact augmented Lagrangian functions. Thus, the use of the exact augmented Lagrangians allows one to overcome main disadvantages of standard augmented Lagrangian methods. Exact augmented Lagrangian functions and numerical methods based on the use of these functions were studied in [16, 15, 53, 21, 22, 19, 18, 32, 31, 54, 23, 17, 24, 25, 20, 26]). However, it should be noted that all existing exact augmented Lagrangian functions were constructed from the Hestenes-Powell-Rockafellar augmented Lagrangian, and were only considered for mathematical programming problems, except for the exact augmented Lagrangian function for nonlinear semidefinite programming problem from a recent paper [33].

In the second part of the paper, we develop a general theory of globally exact augmented Lagrangian functions. We present a simple method for constructing exact augmented Lagrangian functions, and provide first simple necessary and sufficient conditions for the global exactness of these functions in the form of the localization principle. We also demonstrate that globally exact augmented Lagrangians for mathematical programming problems can be constructed not only from the Hestenes-Powell-Rockafellar augmented Lagrangian, but also from many other augmented Lagrangian functions. Moreover, we, for the first time, propose globally exact augmented Lagrangian functions for equality constrained problems and for nonlinear second order cone programs. We also propose new globally exact augmented Lagrangian functions for nonlinear semidefinite programming problems. These exact augmented Lagrangian functions can be utilized in order to design new efficient superlinearly (or even quadratically) convergent optimization methods.

The paper is organized as follows. In Section 2, we introduce a general augmented Lagrangian function for a cone constrained optimization problem, and present main assumptions on this function that are utilized throughout the article. In Section 3, we provide many particular examples of augmented Lagrangian functions for general cone constrained, mathematical programming,

nonlinear second order cone programming, nonlinear semidefinite programming and semi-infinite programming problems, and point out that all these augmetned Lagrangian functions satisfy the main assumptions of this article. Some general results on the existence of global sadde points are presented in Section 4, while in Section 5 we study the localization principle. Applications of the localization principle to cone constrained minimax problems are given in Section 6. The general theory of globally exact augmented Lagrangian functions is developed in Section 7, while some applications of this theory as well as particular examples of globally exact augmented Lagrangians are presented in Section 8.

2 An Augmented Lagrangian Function for Cone Constrained Optimization Problems

Let X be a finite dimensional normed space, and $A \subset X$ be a nonempty set. Let also Y be a normed space, and $K \subset Y$ be a nonempty closed convex cone. As usual, denote by Y^* the topological dual of Y, and by $\langle \cdot, \cdot \rangle$ the standard coupling function between Y and its dual or the inner product in $\mathbb{R}^s$, $s \in \mathbb{N}$, depending on the context.

Throughout this article, we study the following cone constrained optimization problem

$$\min f(x) \quad \text{subject to} \quad G(x) \in K, \quad x \in A, \qquad (\mathcal{P})$$

where $f\colon X \to \mathbb{R} \cup \{+\infty\}$ and $G\colon X \to Y$ are given functions. Hereinafter, we suppose that there exists a feasible point x of $(\mathcal{P})$ such that $f(x) < +\infty$, and there exists a globally optimal solution of the problem $(\mathcal{P})$.

Let us introduce an *augmented Lagrangian function* for the problem $(\mathcal{P})$. Choose a function $\Phi\colon Y \times Y^* \times (0, +\infty) \to \mathbb{R} \cup \{+\infty\}$, and define

$$\mathscr{L}(x, \lambda, c) = f(x) + \Phi(G(x), \lambda, c)$$

where $\lambda \in Y^*$ is a Lagrange multiplier, and $c > 0$ is a penalty parameter.

Remark 1. Note that only the constraint $G(x) \in K$ is incorporated into the augmented Lagrangian function, while the constraint $x \in A$ has to be taken into account directly. This approach allows one to choose what constraints of an optimization problem under consideration are handled via the augmented Lagrangian function, and what constraints are handled by other methods.

Our main goal is to obtain simple necessary and sufficient conditions for the existence of global saddle points of the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$ that can be easily applied to various cone constrained optimization problems, and various augmented Lagrangians for these problems.

Before we proceed to the study of saddle points, let us list main assumptions on the function Φ that we utilise throughout the article. In order to include several particular cases into the general theory, we formulate our assumptions (as well as all definitions and results below) with respect to a given closed convex cone $\Lambda \subseteq Y^*$. In particular, one can define $\Lambda = Y^*$ or $\Lambda = K^*$, where $K^* = \{y^* \in Y^* \mid \langle y^*, y \rangle \leq 0 \text{ for all } y \in K\}$ is the polar cone of K. Let us point out that a proper choice of the cone Λ is necessary to ensure that the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$ has desirable properties. Some remarks on how to choose the cone Λ are given throughout the text.

For any $y \in Y$ and $r > 0$ denote $B(y, r) = \{z \in Y \mid \|z - y\| \leq r\}$ and $\operatorname{dist}(y, K) = \inf_{z \in K} \|y - z\|$. In the subsequent sections, the following assumptions on the function Φ are utilised:

(A1) $\forall y \in K \ \forall c > 0$ one has $\Phi(y, 0, c) \geq 0$;

(A2) $\forall y \in K \ \forall \lambda \in \Lambda \ \forall c > 0$ one has $\Phi(y, \lambda, c) \leq 0$;

(A3) $\forall y \notin K \ \forall c > 0 \ \exists \lambda \in \Lambda$ such that $\lim_{t \to +\infty} \Phi(y, t\lambda, c) = +\infty$;

(A4) $\forall y \in Y \ \forall \lambda \in \Lambda$ the function $\Phi(y, \lambda, c)$ is non-decreasing in c;

(A5) $\forall \lambda \in \Lambda \ \forall c > 0$ the function $\Phi(\cdot, \lambda, c)$ is monotone with respect to the order generated by K, i.e. $\Phi(y_1, \lambda, c) \leq \Phi(y_2, \lambda, c)$ if $y_1 - y_2 \in K$;

(A6) $\forall y \notin K \ \forall \lambda \in \Lambda \ \forall c_0 > 0 \ \exists r > 0$ such that

$$\lim_{c \to +\infty} \inf \{\Phi(z, \lambda, c) - \Phi(z, \lambda, c_0) \mid z \in B(y, r) \colon \Phi(z, \lambda, c_0) < +\infty\} = +\infty;$$

(A7) $\forall \lambda \in \Lambda \ \forall c_0 > 0 \ \forall r > 0$ one has

$$\lim_{c \to +\infty} \inf \left\{ \Phi(y, \lambda, c) - \Phi(y, \lambda, c_0) \ \middle| \right.$$
$$\left. y \in Y \colon \operatorname{dist}(y, K) \geq r, \ \Phi(y, \lambda, c_0) < +\infty \right\} = +\infty.$$

(A8) $\forall y \in K \ \forall \lambda \in \Lambda \setminus K^* \ \forall c > 0$ one has $\Phi(y, \lambda, c) < 0$;

(A9) $\forall y \in K \ \forall \lambda \in \Lambda \ \forall c > 0$ such that $\langle \lambda, y \rangle \neq 0$ one has $\Phi(y, \lambda, c) < 0$;

(A10) $\forall y \in K \ \forall \lambda \in K^* \ \forall c > 0$ such that $\langle \lambda, y \rangle = 0$ one has $\Phi(y, \lambda, c) = 0$;

(A11) $\forall y \in K \ \forall \lambda \in K^* \ \forall c > 0$ such that $\langle \lambda, y \rangle = 0$ the function $\Phi(\cdot, \lambda, c)$ is Fréchet differentiable at y, and $D_y \Phi(y, \lambda, c) = \Phi_0(\lambda)$, where D_y stands for the Fréchet derivative in y, and $\Phi_0 \colon K^* \to K^*$ is a surjective mapping that does not depend on y and c, and such that $\langle \Phi_0(\lambda), y \rangle = 0$ iff $\langle \lambda, y \rangle = 0$;

(A12) $\forall y \in K \ \forall \lambda \in \Lambda$ one has $\Phi(y, \lambda, c) \to 0$ as $c \to \infty$.

Remark 2. Since K is closed, $(A7)$ implies $(A6)$. Furthermore, note that the bigger is the cone Λ, the more restrictive are the above assumptions (apart from assumption $(A3)$).

3 Examples of Augmented Lagrangian Functions

Below, we provide many particular examples of the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$ for various cone constrained optimization problems, and point out whether assumptions above are satisfied in these examples.

3.1 General Cone Constrained Problems

We start with an augmented Lagrangian function introduced by Rockafellar and Wets ([68], Section 11.K), which is the only augmented Lagrangian function for the general cone constrained optimization problem known to the author (apart from its direct generalizations, as in [5, 85, 81]). For more details on this augmented Lagrangian, see [38, 72, 39, 28] and references therein.

Example 1. Let a function $\sigma \colon Y \to [0, +\infty]$ be such that $\sigma(0) = 0$ and $\sigma(y) > 0$ for any $y \neq 0$. Define

$$\Phi(y, \lambda, c) = \inf_{p \in K - y} \big(- \langle \lambda, p \rangle + c\sigma(p) \big).$$

Here we suppose that the function σ is such that $\Phi(y, \lambda, c) > -\infty$ for all $y \in Y$, $\lambda \in Y^*$ and $c > 0$. In particular, one can set $\sigma(y) = \|y\|^2/2$.

Let $\Lambda = Y^*$. Then assumptions $(A1)$–$(A5)$ and $(A10)$ are satisfied in the general case. Assumptions $(A6)$ and $(A7)$ hold true, if the function σ has *a valley at zero* (i.e. for any neighbourhood U of zero there exists $\delta > 0$ such that $\sigma(y) \geq \delta$ for all $y \in Y \setminus U$). Assumptions $(A8)$ and $(A9)$ are valid, if $\sigma(ty)/t \to 0$ as $t \to 0$ for any $y \in Y$. Assumption $(A11)$ is satisfied with $\Phi_0(\lambda) \equiv \lambda$, in particular, if Y is a Hilbert space and $\sigma(y) = \|y\|^2/2$ (see [72]). Finally, assumption $(A12)$ is satisfied, provided $\sigma(y) \geq \omega(\|y\|)$ for some non-negative continuous function ω such that $\omega(t) = 0$ iff $t = 0$, and $\liminf_{t \to +\infty} \omega(t)/t > 0$.

3.2 Mathematical Programming

Consider the following mathematical programming problem:

$$\min f(x) \quad \text{subject to} \quad g_i(x) \leq 0, \quad i \in I, \quad g_j(x) = 0, \quad j \in J, \quad x \in A, \quad (1)$$

where $g_i \colon X \to \mathbb{R}$ are given functions, $I = \{1, \ldots, l\}$ and $J = \{l+1, \ldots, l+s\}$. Denote $Y = \mathbb{R}^{l+s}$, $G(\cdot) = (g_1(\cdot), \ldots, g_{l+s}(\cdot))$ and $K = \mathbb{R}^l_- \times \{0_s\}$, where $\mathbb{R}_- = (-\infty, 0]$. Then problem (1) is equivalent to the problem $(\mathcal{P})$. Note that in this case $K^* = \mathbb{R}^l_+ \times \mathbb{R}^s$ and $Y^* = \mathbb{R}^{l+s}$, where $\mathbb{R}_+ = [0, +\infty)$.

Below, we only provide examples of *separable* augmented Lagrangian functions for problem (1), i.e. such augmented Lagrangians that

$$\Phi(y, \lambda, c) = \sum_{i=1}^{l+s} \Phi_i(y_i, \lambda_i, c), \quad y = (y_1, \ldots, y_{l+s}), \quad \lambda = (\lambda_1, \ldots, \lambda_{l+s}),$$

where $\Phi_i \colon \mathbb{R}^2 \times (0, +\infty) \to \mathbb{R} \cup \{+\infty\}$ are some functions. Let us note that most of (if not all) augmented Lagrangian functions for mathematical programming problems appearing in applications are indeed separable.

Example 2. Suppose that $J = \emptyset$, i.e. suppose that there are no equality constraints. Let $\phi \colon \mathbb{R} \to \mathbb{R}$ be a twice continuously differentiable and strictly convex function such that $\phi(0) = 0$, $\phi'(0) = 0$ and the derivative $\phi'(\cdot)$ is surjective. For any $s, t \in \mathbb{R}$ define

$$P(t, \lambda) = \begin{cases} \lambda t + \phi(t), & \text{if } \lambda + \phi'(t) \geq 0, \\ \min_{\tau \in \mathbb{R}}[\lambda \tau + \phi(\tau)], & \text{otherwise,} \end{cases}$$

and set

$$\Phi_i(y_i, \lambda_i, c) = \frac{1}{c} P(cy_i, \lambda_i) \quad \forall i \in I$$

(see [1], Section 5.1.2, Example 1). In this case, the function $\mathscr{L}(x, \lambda, c)$ is called the *essentially quadratic* augmented Lagrangian function for problem (1) [79, 50, 83]. If $\phi(t) = t^2/2$, then the essentially quadratic augmented Lagrangian function coincides with the well-known Hestenes-Powell-Rockafellar Lagrangian function [66, 67, 2], which is a particular case of the augmented Lagrangian from Example 1.

Let $\Lambda = Y^* = \mathbb{R}^l$. Then one can verify that all assumptions $(A1)$–$(A12)$ are satisfied, and $\Phi_0(\lambda) \equiv \lambda$.

Example 3. Suppose that $J = \emptyset$. Define

$$\Phi_i(y_i, \lambda_i, c) = \frac{1}{3c} \Big[\max \big\{ \operatorname{sign}(\lambda_i) \sqrt{|\lambda_i|} + cy_i, 0 \big\}^3 - |\lambda_i|^{3/2} \Big] \quad \forall i \in I.$$

Then $\mathscr{L}(x, \lambda, c)$ coincides with *the cubic augmented Lagrangian* [42].

Let $\Lambda = Y^* = \mathbb{R}^l$. Then it is not difficult to check that all assumptions $(A1)$–$(A12)$ are valid, and $\Phi_0(\lambda) \equiv \lambda$.

Example 4. Let $\phi \colon \mathbb{R} \to \mathbb{R}$ be a twice differentiable and strictly convex function such that $\phi(0) = \phi'(0) = 0$, and $\phi'(\cdot)$ is surjective. Define

$$\Phi_i(y_i, \lambda_i, c) = \frac{1}{c} \Big[\phi\Big(\max\{cy_i + \lambda_i, 0\} \Big) - \phi(\lambda_i) \Big] \quad \forall i \in I,$$

$$\Phi_j(y_j, \lambda_j, c) = \frac{1}{c} \Big[\phi(cy_j + \lambda_j) - \phi(\lambda_j) \Big] \quad \forall j \in J.$$

Then $\mathscr{L}(x, \lambda, c)$ coincides with the augmented Lagrangian function proposed by Mangasarian [59] (see also [89]).

Let $\Lambda = Y^* = \mathbb{R}^{l+s}$. Then one can check that all assumptions $(A1)$–$(A12)$ are satisfied, and $\Phi_0(\lambda) \equiv \phi'(\lambda) = (\phi'(\lambda_1), \ldots, \phi'(\lambda_{l+s}))$.

Example 5. Suppose that $J = \emptyset$. Let $\phi \colon \mathbb{R} \to \mathbb{R}$ be a twice differentiable and strictly increasing function such that $\phi(0) = 0$. Define

$$\Phi_i(y_i, \lambda_i, c) = \frac{\lambda_i}{c} \phi(cy_i) \quad \forall i \in I.$$

If $\phi(t) = e^t - 1$, then the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$ coincides with *the exponential penalty function* [1, 80, 79, 50, 83]. In the case $\phi(t) = 2(\ln(1 + e^t) - \ln 2)$ it coincides with *the Log-Sigmoid Lagrangian* [64, 65]

Let $\Lambda = K^* = \mathbb{R}_+^l$. Then one can check that assumptions $(A1)$–$(A3)$, $(A5)$ and $(A8)$–$(A10)$ are satisfied. Assumption $(A4)$ is valid, provided ϕ is convex. Assumption $(A11)$ is satisfied iff $\phi'(0) \neq 0$, and in this case $\Phi_0(\lambda) \equiv \phi'(0)\lambda$. Assumption $(A12)$ is satisfied iff $\phi(t)/t \to 0$ as $t \to -\infty$, while assumptions $(A6)$ and $(A7)$ are not valid, regardless of the choice of Λ (set $\lambda = 0$). Note also that assumptions $(A2)$, $(A4)$ (if ϕ is convex), $(A5)$, $(A8)$ and $(A9)$ are satisfied iff $\Lambda \subseteq K^*$.

Example 6. Suppose that $J = \emptyset$. Let the function ϕ be as in the previous example, and let $\xi \colon \mathbb{R} \to \mathbb{R}_+$ be a twice continuously differentiable non-decreasing

function such that $\xi(t) = 0$ for all $t \le 0$, and $\xi(t) > 0$ for all $t > 0$ (in particular, one can choose $\xi(t) = \max\{0, t\}^3$). Define

$$\Phi_i(y_i, \lambda_i, c) = \frac{\lambda_i}{c}\phi(cy_i) + \frac{1}{c}\xi(cy_i) \quad \forall i \in I$$

(see [1], Section 5.1.2, Example 2). In this case, the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$ is called *the penalized exponential-type augmented Lagrangian function* [79, 50, 83].

Let $\Lambda = K^* = \mathbb{R}_+^l$. Then, as in the previous example, assumptions $(A1)$–$(A3)$, $(A5)$ and $(A8)$–$(A10)$ are satisfied in the general case, assumption $(A4)$ is valid, provided both ϕ and ξ are convex, assumption $(A11)$ is satisfied iff $\phi'(0) \ne 0$ (in this case $\Phi_0(\lambda) \equiv \phi'(0)\lambda$), and assumption $(A12)$ is satisfied iff $\phi(t)/t \to 0$ as $t \to -\infty$. However, in contrast to the previous example, assumptions $(A6)$ and $(A7)$ are satisfied if and, in the case $l > 1$, only if $\xi(t)/t \to +\infty$ as $t \to +\infty$, and ϕ is bounded below (note that if $l > 1$, then for $(A7)$ to hold true it is necessary that $\Phi_i(\cdot, \lambda_i, c)$ is bounded below on $\mathbb{R}_-$). Let us also note that for assumptions $(A2)$ and $(A4)$–$(A9)$ to hold true it is necessary that $\Lambda \subseteq K^*$.

Remark 3. As we will demonstrate below (see Example 18 and Remark 14), in the case when assumption $(A6)$ is not satisfied, it is necessary to impose some rather restrictive assumptions on the problem $(\mathcal{P})$ in order to guarantee the existence of a global saddle point of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$. In order to avoid this drawback one can introduce an additional penalty term into the definition of the function $\Phi(y, \lambda, c)$ in the same way as in example above (cf. [79]).

Example 7. Suppose that $J = \emptyset$. Let $\phi\colon (-\infty, 1) \to \mathbb{R}$ be a twice differentiable and strictly increasing function such that $\phi(0) = 0$. Define

$$\Phi_i(y_i, \lambda_i, c) = \begin{cases} \dfrac{\lambda_i}{c}\phi(cy_i), & \text{if} \quad cy_i < 1, \\ +\infty, & \text{otherwise.} \end{cases} \quad \forall i \in I$$

In this case, the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$ coincides with *the modified barrier function* introduced by Polyak [63]. In particular, if $\phi(s) = -\ln(1 - s)$ or $\phi(s) = 1/(1 - s) - 1$, then $\mathscr{L}(x, \lambda, c)$ coincides with the modified Frisch function or the modified Carroll function, respectively [63] (see also [79, 50, 83]).

Let $\Lambda = K^* = \mathbb{R}_+^l$. Then assumptions $(A1)$–$(A3)$ and $(A5)$–$(A10)$ are satisfied. Assumption $(A4)$ is satisfied, if ϕ is convex; assumption $(A11)$ is valid iff $\phi'(0) \ne 0$ ($\Phi_0(\lambda) \equiv \phi'(0)\lambda$), while assumption $(A12)$ is satisfied, provided $\phi(t)/t \to 0$ as $t \to -\infty$. Furthermore, note that the function $\Phi_i(\cdot, \lambda_i, c)$ is l.s.c. iff $\phi(t) \to +\infty$ as $t \to 1$ and $\Lambda \subseteq K^*$. Thus, in particular, the modified Frisch and the modified Carroll functions satisfy assumptions $(A1)$–$(A12)$ and are l.s.c. in y. Note, finally, that for assumptions $(A2)$, $(A4)$, $(A5)$, $(A8)$ and $(A9)$ to hold true it is necessary that $\Lambda \subseteq K^*$.

Example 8. Suppose that $J = \emptyset$. Choose $b \ge 0$ and a non-decreasing function $\phi\colon \mathbb{R} \to \mathbb{R}_+$ such that $\phi(t) > \phi(b) > 0$ for all $t > b$ (in particular, one can set $\phi(t) = \exp(t)$ or $\phi(t) = \max\{0, t\}$ with $b > 0$). Note that the inequality

$g_i(x) \leq 0$ is satisfied iff $\phi(g_i(x) + b)/\phi(b) \leq 1$. Furthermore, $\phi(g_i(x) + b) \geq 0$ for all $x \in X$ by the definition of ϕ. Define

$$\Phi_i(y_i, \lambda_i, c) = \frac{\lambda_i}{c} \left[\left(\frac{\phi(y_i + b)}{\phi(b)} \right)^c - 1 \right] \quad \forall i \in I$$

(see [45]). Then $\mathscr{L}(x, \lambda, c)$ coincides with *the p-th power augmented Lagrangian function* [43, 44, 91, 46, 45, 88, 50].

Let $\Lambda = K^* = \mathbb{R}_+^l$. Then assumptions $(A1)$–$(A5)$, $(A8)$–$(A10)$ and $(A12)$ are satisfied. Assumption $(A11)$ is satisfied provided $\phi'(b) \neq 0$ $(\Phi_0(\lambda) \equiv \phi'(b)\lambda)$, while assumptions $(A6)$ and $(A7)$ are not valid, regardless of the choice of Λ. Note also that, as in the above examples, assumptions $(A2)$, $(A4)$, $(A5)$, $(A8)$ and $(A9)$ are satisfied iff $\Lambda \subseteq K^*$.

Example 9. Let $J = \emptyset$. Following the ideas of [36], define

$$\Phi_i(y_i, \lambda_i, c) = \frac{1}{c} \int_0^{cy_i} \left(\sqrt{t^2 + \lambda_i^2} + t \right) dt \quad \forall i \in I.$$

For any $\lambda_i \neq 0$ the function Φ_i has the form

$$\Phi_i(y_i, \lambda_i, c) = \frac{y_i}{2} \sqrt{(cy_i)^2 + \lambda_i^2} + \frac{cy_i^2}{2} + \frac{\lambda_i^2}{2c} \ln \left(\sqrt{(cy_i)^2 + \lambda_i^2} + cy_i \right) - \frac{\lambda_i^2}{2c} \ln |\lambda_i|,$$

while $\Phi_i(y_i, 0, c) = cy_i(|y_i| + y_i)/2$. In this case, we refer to the function $\mathscr{L}(x, \lambda, c)$ as He-Wu-Meng's augmented Lagrangian.

Let $\Lambda = Y^* = \mathbb{R}^l$. Then one can verify that assumptions $(A1)$–$(A6)$ and $(A9)$–$(A12)$ are satisfied, and $\Phi_0(\lambda) \equiv \lambda$. Assumption $(A8)$ holds true iff $\Lambda \subseteq K^*$, since, in essence, He-Wu-Meng's augmented Lagrangian function does not distinguish multipliers λ and $-\lambda$. Assumption $(A7)$ is not valid in the case $l > 1$ due to the fact that $\Phi_i(y_i, \lambda_i, c) \to -\infty$ as $y_i \to -\infty$.

3.3 Nonlinear Second Order Cone Programming

Consider the following nonlinear second order cone programming problem:

$$\min f(x) \quad \text{subject to} \quad g_i(x) \in Q_{l_i+1}, \quad i \in I, \quad h(x) = 0, \quad x \in A, \quad (2)$$

where $g_i \colon X \to \mathbb{R}^{l_i+1}$, $I = \{1, \ldots, r\}$, and $h \colon X \to \mathbb{R}^s$ are given functions, and

$$Q_{l_i+1} = \left\{ y = (y^0, \overline{y}) \in \mathbb{R} \times \mathbb{R}^{l_i} \mid y^0 \geq \|\overline{y}\| \right\}$$

is the second order (Lorentz) cone of dimension $l_i + 1$ (here $\|\cdot\|$ is the Euclidean norm). Denote

$$Y = \mathbb{R}^{l_1+1} \times \ldots \times \mathbb{R}^{l_r+1} \times \mathbb{R}^s, \quad K = Q_{l_1+1} \times \ldots \times Q_{l_r+1} \times \{0_s\},$$

and $G(\cdot) = (g_1(\cdot), \ldots, g_r(\cdot), h(\cdot))$. Then problem (2) is equivalent to the problem $(\mathcal{P})$. Note that in this case $K^* = (-Q_{l_1+1}) \times \ldots \times (-Q_{l_r+1}) \times \mathbb{R}^s$ and $Y^* = Y$.

Example 10. For any $y = (y_1, \ldots, y_r, z) \in Y$ and $\lambda = (\lambda_1, \ldots, \lambda_r, \mu) \in Y$ define

$$\Phi(y, \lambda, c) = \frac{c}{2} \sum_{i=1}^r \left[\operatorname{dist}^2 \left(y_i + \frac{1}{c} \lambda_i, Q_{l_i+1} \right) - \frac{1}{c^2} \|\lambda_i\|^2 \right] + \langle \mu, z \rangle + \frac{c}{2} \|z\|^2.$$

9

(see [51, 52, 97]). The function $\Phi(y, \lambda, c)$ defined above is a particular case of the function $\Phi(y, \lambda, c)$ for the general cone constrained optimization problem from Example 1 with $\sigma(y) = \|y\|^2/2$. Therefore it satisfies all assumptions $(A1)$–$(A12)$ with $\Lambda = Y^* = \mathbb{R}^{l_1+1} \times \ldots \times \mathbb{R}^{l_r+1} \times \mathbb{R}^s$, and $\Phi_0(\lambda) \equiv \lambda$.

Following the ideas of [93] we can define another augmented Lagrangian function for problem (2), which is a generalization of the augmented Lagrangians from Examples 5 and 7 to the case of nonlinear second order cone programming problems. To this end, recall that for any function $\psi\colon \mathbb{R} \to \mathbb{R}$ and for any $y = (y^0, \overline{y}) \in \mathbb{R}^{l+1}$ *Löwner's operator* associated with ψ is defined as

$$\Psi(y) = \frac{1}{2} \begin{pmatrix} \psi(y^0 + \|\overline{y}\|) + \psi(y^0 - \|\overline{y}\|) \\ \left(\psi(y^0 + \|\overline{y}\|) - \psi(y^0 - \|\overline{y}\|)\right) \dfrac{\overline{y}}{\|\overline{y}\|} \end{pmatrix},$$

if $\overline{y} \neq 0$, and $\Psi(y) = (\psi(y^0), 0_l)$ otherwise (see [93, 7, 75] for more details). It is easy to check that if ψ is non-decreasing and $\psi(0) = 0$, then $y \in K \implies \Psi(-y) \in -K$, while $y \notin K \implies \Psi(-y) \notin -K$.

Example 11. Suppose, for the sake of simplicity, that there are no equality constraints. Let $\psi\colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a non-decreasing convex function such that $\mathrm{dom}\, \psi = (-\infty, \varepsilon_0)$ for some $\varepsilon_0 \in (0, +\infty]$, $\psi(t) \to +\infty$ as $t \to \varepsilon_0$, $\psi(t)/t \to +\infty$ as $t \to +\infty$ if $\varepsilon_0 = +\infty$, ψ is twice continuously differentiable on $\mathrm{dom}\, \psi$, $\psi(0) = 0$ and $\psi'(0) = 1$. Then for any $y = (y_1, \ldots, y_r) \in Y$ and $\lambda = (\lambda_1, \ldots, \lambda_r) \in Y$ define

$$\Phi(y, \lambda, c) = -\frac{1}{c} \sum_{i=1}^{r} \langle \lambda_i, \Psi(-cy_i) \rangle, \tag{3}$$

if $y_i^0 + \|\overline{y}_i\| < \varepsilon_0$ for all $i \in I$, and $\Phi(y, \lambda, c) = +\infty$ otherwise. It is easy to see that the function $\Phi(y, \lambda, c)$ is lower semicontinuous.

Let $\Lambda = K^*$. Then one can easily verify that assumptions $(A1)$–$(A4)$, $(A8)$–$(A10)$ are satisfied in the general case. Assumption $(A11)$ is satisfied with $\Phi_0(\lambda) \equiv \lambda$ by [93], Lemma 3.1. Assumptions $(A6)$ and $(A7)$ are valid, if $\varepsilon_0 < +\infty$, while assumption $(A12)$ is valid, provided $\psi(t)/t \to 0$ as $t \to -\infty$. Finally, assumption $(A5)$ is never satisfied for the function Φ defined above, since for $(A5)$ to hold true it is necessary that Löwner's operator $y \to \Psi(-y)$ is non-increasing with respect to the order generated by the second order cone, which is not the case when ψ is non-decreasing. Let us also note that for assumptions $(A2)$, $(A4)$, $(A8)$ and $(A9)$ to hold true it is necessary that $\Lambda \subseteq K^*$.

Remark 4. Note that one can easily incorporate equality constraints into the augmented Lagrangian function from the previous example by simply adding terms corresponding to these constraints into the right-hand side of (3). One can define these additional terms in the same way as in Examples 4 or 10.

3.4 Nonlinear Semidefinite Programming

Consider the following nonlinear semidefinite programming problem:

$$\min f(x) \quad \text{subject to} \quad G_0(x) \preceq 0, \quad h(x) = 0, \quad x \in A, \tag{4}$$

where $G_0 \colon X \to \mathbb{S}^l$ and $h \colon X \to \mathbb{R}^s$ are given functions, $\mathbb{S}^l$ denotes the set of all $l \times l$ real symmetric matrices, and the relation $G_0(x) \preceq 0$ means that the matrix $G_0(x)$ is negative semidefinite. Hereinafter, we suppose that the linear space $\mathbb{S}^l$ is equipped with the inner product $\langle A, B \rangle = \mathrm{Tr}(AB)$, and the corresponding norm $\|A\|_F = \sqrt{\mathrm{Tr}(A^2)}$, which is called the Frobenius norm of a matrix $A \in \mathbb{S}^l$. Here $\mathrm{Tr}(\cdot)$ is the trace operator.

Denote by $\mathbb{S}^l_+$ the cone of $l \times l$ positive semidefinite matrices, and denote $\mathbb{S}^l_-$ by the cone of $l \times l$ negative semidefinite matrices. Define $Y = \mathbb{S}^l \times \mathbb{R}^s$, $K = \mathbb{S}^l_- \times \{0_s\}$ and $G(\cdot) = (G_0(\cdot), h(\cdot))$. Then problem (4) is equivalent to the problem $(\mathcal{P})$. Note that in this case $K^* = \mathbb{S}^l_+ \times \mathbb{R}^s$ and $Y^* = Y$.

Example 12. For any $y = (y_0, z) \in Y = \mathbb{S}^l \times \mathbb{R}^s$ and $\lambda = (\lambda_0, \mu) \in Y$ define

$$\Phi(y, \lambda, c) = \frac{1}{2c}\Big(\mathrm{Tr}\left([cy_0 + \lambda_0]^2_+\right) - \mathrm{Tr}(\lambda_0^2)\Big) + \langle \mu, z \rangle + \frac{c}{2}\|z\|^2,$$

where $[\cdot]_+$ denotes the projection of a matrix onto the cone $\mathbb{S}^l_+$ (see [78, 76, 96, 77, 86, 57, 87, 90, 92] for more details on this augmented Lagrangian). One can check that the function $\Phi(y, \lambda, c)$ defined above is a particular case of the function $\Phi(y, \lambda, c)$ from Example 1 with $\sigma(y) = (\|y_0\|_F^2 + \|z\|^2)/2$. Therefore it satisfies all assumptions $(A1)$–$(A12)$ with $\Lambda = Y^* = \mathbb{S}^l \times \mathbb{R}^s$ and $\Phi_0(\lambda) \equiv \lambda$.

As in the case of the second order cone programs, one can extend the augmented Lagrangians from Examples 5 and 7 to the case of nonlinear semidefinite programming problems with the use of Löwner's operator (see [75, 71, 8]). For any function $\psi \colon \mathbb{R} \to \mathbb{R}$ and any $y \in \mathbb{S}^l$ the matrix function (Löwner's operator) associated with ψ is defined as

$$\Psi(y) = E \, \mathrm{diag}\Big(\psi(\rho_1(y)), \dots, \psi(\rho_l(y))\Big) E^T,$$

where $y = E \, \mathrm{diag}(\rho_1(y), \dots, \rho_l(y)) E^T$ is a spectral decomposition of $y \in \mathbb{S}^l$, and $\rho_1(y), \dots, \rho_l(y)$ are the eigenvalues of y listed in the decreasing order. Note that the projection operator $[\cdot]_+$ is simply the matrix function associated with the function $\psi(t) = \max\{0, t\}$.

Example 13. Let $\psi \colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a non-decreasing convex function such that $\mathrm{dom}\,\psi = (-\infty, \varepsilon_0)$ for some $\varepsilon_0 \in (0, +\infty]$, $\psi(t) \to +\infty$ as $t \to \varepsilon_0$, $\psi(t)/t \to +\infty$ as $t \to +\infty$ if $\varepsilon_0 = +\infty$, ψ is twice continuously differentiable on $\mathrm{dom}\,\psi$, $\psi(0) = 0$ and $\psi'(0) = 1$. For any $y = (y_0, z) \in Y = \mathbb{S}^l \times \mathbb{R}^s$ and $\lambda = (\lambda_0, \mu) \in Y$ define

$$\Phi(y, \lambda, c) = \frac{1}{c}\langle \lambda_0, \Psi(cy_0)\rangle + \langle \mu, z \rangle + \frac{c}{2}\|z\|^2,$$

if $\sigma_1(y_0) < \varepsilon_0$, and $\Phi(y, \lambda, c) = +\infty$ otherwise (see [74, 60, 48, 94, 92, 55]). Note that the function $\Phi(y, \lambda, c)$ is lower semicontinuous.

Let $\Lambda = K^* = \mathbb{S}^l_+ \times \mathbb{R}^s$. Then assumptions $(A1)$–$(A4)$ and $(A8)$–$(A10)$ hold true. Furthermore, assumption $(A11)$ is satisfied with $\Phi_0(\lambda) \equiv \lambda$ by [55], Proposition 4.2. Assumption $(A5)$ is satisfied, if the matrix function $\Psi(y)$ is monotone (see, e.g., [37], Def. 6.6.33). Assumptions $(A6)$ and $(A7)$ hold true, provided $\varepsilon_0 < +\infty$, while assumption $(A12)$ is valid iff $\psi(t)/t \to 0$ as $t \to -\infty$. Note also that for assumptions $(A2)$, $(A4)$, $(A5)$, $(A8)$ and $(A9)$ to hold true it is necessary that $\Lambda \subseteq K^*$.

Following the ideas of [55] we can also extend the penalized exponential-type augmented Lagrangian function from Example 6 to the case of nonlinear semidefinite programming problems.

Example 14. Let the function ψ be as in the previous example with $\varepsilon_0 = +\infty$, and let $\xi \colon \mathbb{R} \to \mathbb{R}$ be a twice continuously differentiable non-decreasing convex function such that $\xi(t) = 0$ for all $t \leq 0$, and $\xi(t) > 0$ for all $t > 0$. Denote by $\Xi(\cdot)$ the matrix function associated with $\xi(\cdot)$.

For any $y = (y_0, z) \in Y$ and $\lambda = (\lambda_0, \mu) \in Y$ define

$$\Phi(y, \lambda, c) = \frac{1}{c} \langle \lambda_0, \Psi(cy_0) \rangle + \frac{1}{c} \operatorname{Tr} \left(\Xi(cy_0) \right) + \langle \mu, z \rangle + \frac{c}{2} \|z\|^2$$

(see [55]). Let $\Lambda = K^* = \mathbb{S}_+^l \times \mathbb{R}^s$. Then assumptions $(A1)$–$(A4)$ and $(A8)$–$(A11)$ are satisfied. Assumption $(A5)$ hold true, provided both matrix functions Ψ and Ξ are monotone, assumptions $(A6)$ and $(A7)$ are satisfied if and, in the case $l > 1$, only if $\xi(t)/t \to +\infty$ as $t \to +\infty$, and ψ is bounded below, while assumption $(A12)$ is valid iff $\psi(t)/t \to 0$ as $t \to -\infty$.

3.5 Semi-Infinite Programming

Consider the following semi-infinite programming problem:

$$\min f(x) \quad \text{subject to} \quad g_i(x, t) \leq 0, \quad t \in T, \quad i \in I, \quad h(x) = 0, \quad x \in A, \quad (5)$$

where T is a compact metric space, the functions $g_i \colon X \times T \to \mathbb{R}$, $I = \{1, \ldots, l\}$, are continuous, and $h \colon X \to \mathbb{R}^s$. Let $C(T)$ be the space of all real-valued continuous functions defined on T endowed with the uniform norm, and denote by $C_+(T)$ the closed convex subcone of $C(T)$ consisting of all nonnegative functions. As it is well-known, the topological dual space of $C(T)$ is isometrically isomorphic to the space of signed (i.e. real-valued) regular Borel measures on T, which we denote by $rca(T)$, while the space of regular Borel measures on T is denoted as $rca_+(T)$.

Define $Y = (C(T))^l \times \mathbb{R}^s$ and $K = (-C_+(T))^l \times \{0_s\}$, and introduce the function $x \to G(x) = (g_1(x, \cdot), \ldots, g_l(x, \cdot), h(x))$ mapping X to Y. Then problem (5) is equivalent to the problem $(\mathcal{P})$. Note that in this case Y^* is isometrically isomorphic to (and, thus, can be identified with) the space $(rca(T))^l \times \mathbb{R}^s$, while K^* can be identified with the cone $(rca_+(T))^l \times \mathbb{R}^s$.

To the best of author's knowledge, the only augmented Lagrangian function for semi-infinite programming problems studied in the literature is a particular case of the Rockafellar-Wets augmented Lagrangian from Example 1 (see [70, 41, 73, 6]).

Example 15. Let a function $\sigma \colon C(T) \to \mathbb{R}_+$ be such that $\sigma(0) = 0$ and $\sigma(y) > 0$ for all $y \neq 0$. Suppose also that σ has a valley at zero. For any $y = (y_1, \ldots, y_l, z) \in Y$ (i.e. $y_i \in C(T)$ and $z \in \mathbb{R}^s$) and $\lambda = (\lambda_1, \ldots, \lambda_l, \mu) \in Y^*$ (i.e. $\lambda_i \in rca(T)$ and $\mu \in \mathbb{R}^s$) define

$$\Phi(y, \lambda, c)$$

$$= \sum_{i=1}^{l} \inf \left\{ -\int_T p \, d\lambda_i + c\sigma(p) \ \Big| \ p \in C(T), \ p(\cdot) + y_i(\cdot) \leq 0 \right\} + \langle \mu, z \rangle + \frac{c}{2} \|z\|^2.$$

Let $\Lambda \subseteq Y^*$ (in particular, one can define Λ as the set of all measures $\lambda \in rca(T)$ with finite support, see [70, 6]). Then, as it was pointed out in Example 1, assumptions $(A1)$–$(A7)$ and $(A10)$ are satisfied. Assumptions $(A8)$ and $(A9)$ are valid, if $\sigma(ty)/t \to 0$ as $t \to 0$ for any $y \in Y$, while assumption $(A12)$ is satisfied, provided $\sigma(y) \geq \omega(\|y\|)$ for all $y \in Y$, where ω is a non-negative continuous function such that $\omega(t) = 0$ iff $t = 0$ and $\liminf_{t \to +\infty} \omega(t)/t > 0$. Finally, it seems that there are no natural general assumptions on the function σ and the cone Λ which can guarantee the validity of $(A11)$.

Let us also present a new augmented Lagrangian function for problem (5), which is a simple extension of the augmented Lagrangian functions from Examples 5 and 7 to the case of semi-infinite programming problems.

Example 16. Suppose, for the sake of simplicity, that there are no equality constraints. Let $\phi \colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a non-decreasing convex function such that $\operatorname{dom} \phi = (-\infty, \varepsilon_0)$ for some $\varepsilon_0 \in (0, +\infty]$, $\phi(t) \to +\infty$ as $t \to \varepsilon_0$, $\phi(t)/t \to +\infty$ as $t \to +\infty$ if $\varepsilon_0 = +\infty$, ϕ is twice continuously differentiable on $\operatorname{dom} \phi$, $\phi(0) = 0$ and $\phi'(0) = 1$. For any $y = (y_1, \dots, y_l) \in Y$ and $\lambda = (\lambda_1, \dots, \lambda_l) \in Y^*$ define

$$\Phi(y, \lambda, c) = \frac{1}{c} \sum_{i=1}^{l} \int_T \phi(cy_i) \, d\lambda_i, \tag{6}$$

if $\|y_i\| < \varepsilon_0$ for all $i \in I$, and $\Phi(y, \lambda, c) = +\infty$ otherwise.

Let $\Lambda \subseteq K^* = (rca_+(T))^l$. Then assumptions $(A1)$–$(A5)$ and $(A8)$–$(A11)$ are satisfied $(\Phi_0(\lambda) \equiv \lambda)$. Assumptions $(A6)$ and $(A7)$ are valid, provided $\varepsilon_0 < +\infty$, while assumption $(A12)$ is valid iff $\phi(t)/t \to 0$ as $t \to -\infty$.

Remark 5. (i) Note that if one defines function (6) for $\lambda \in K^*$, then there is no need to separate the case when $\|y_i\| \geq \varepsilon_0$ for some $i \in I$.
(ii) If $T \subset \mathbb{R}^q$ for some $q \in \mathbb{N}$, then one can also extend the penalized exponential-type augmented Lagrangian function from Example 6 to the case of semi-infinite programming problems by simply adding the term $c^{-1} \int_T \xi(cy_i(t)) \, dt$ to the right-hand side of (6), where the function ξ is the same as in Example 6.

4 Saddle Points of Augmented Lagrangian Functions

Let us turn to the study of saddle points of the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$. Recall that we formulate all definitions and results with respect to a given closed convex cone $\Lambda \subseteq Y^*$.

Definition 1. A pair $(x_*, \lambda_*) \in A \times \Lambda$ is called a *global saddle point* of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$ if there exists $c_0 > 0$ such that $\mathscr{L}(x_*, \lambda_*, c) < +\infty$ for all $c \geq c_0$ and

$$\sup_{\lambda \in \Lambda} \mathscr{L}(x_*, \lambda, c) \leq \mathscr{L}(x_*, \lambda_*, c) \leq \inf_{x \in A} \mathscr{L}(x, \lambda_*, c) \qquad \forall c \geq c_0.$$

The greatest lower bound of all such c_0 is denoted by $c^*(x_*, \lambda_*)$ and is referred to as *the least exact penalty parameter* for the global saddle point (x_*, λ_*).

Definition 2. A pair $(x_*, \lambda_*) \in A \times \Lambda$ is called a *local saddle point* of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$ if there exist $c_0 > 0$ and a neighbourhood U of x_* such that $\mathscr{L}(x_*, \lambda_*, c) < +\infty$ for all $c \geq c_0$ and

$$\sup_{\lambda \in \Lambda} \mathscr{L}(x_*, \lambda, c) \leq \mathscr{L}(x_*, \lambda_*, c) \leq \inf_{x \in U \cap A} \mathscr{L}(x, \lambda_*, c) \qquad \forall c \geq c_0. \qquad (7)$$

The greatest lower bound of all such c_0 is denoted by $c_{loc}^*(x_*, \lambda_*)$ and is referred to as *the least local exact penalty parameter* for the local saddle point (x_*, λ_*).

Observe a direct connection between saddle points of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$ and optimal solutions of the problem $(\mathcal{P})$.

Proposition 1. *Suppose that assumptions $(A1)$–$(A3)$ are satisfied. If (x_*, λ_*) is a local (global) saddle point of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$, then x_* is a locally (globally) optimal solution of the problem $(\mathcal{P})$.*

Proof. Let (x_*, λ_*) be a local saddle point of $\mathscr{L}(x, \lambda, c)$. Then $\mathscr{L}(x_*, \lambda_*, c)$ is finite for all $c > 0$, and there exist $c_0 > 0$ and a neighbourhood U of x_* such that (7) holds true. Let us prove, at first, that x_* is a feasible point of the problem $(\mathcal{P})$. Indeed, suppose that x_* is infeasible, i.e. that $G(x_*) \notin K$. Then by $(A3)$ for any $c > 0$ there exists $\lambda_0 \in \Lambda$ such that $\lim_{t \to +\infty} \Phi(G(x_*), t\lambda_0, c) = +\infty$. From (7) it follows that for any $c \geq c_0$ one has

$$\mathscr{L}(x_*, \lambda_*, c) \geq \mathscr{L}(x_*, t\lambda_0, c) = f(x_*) + \Phi(G(x_*), t\lambda_0, c) \quad \forall t > 0.$$

Passing to the limit as $t \to +\infty$ one obtains that $\mathscr{L}(x_*, \lambda_*, c) = +\infty$, which contradicts the definition of local saddle point. Thus, $G(x_*) \in K$.

Let $x \in U$ be a feasible point of the problem $(\mathcal{P})$. Then applying the second inequality in (7), and $(A2)$ one obtains that

$$\mathscr{L}(x_*, \lambda_*, c) \leq \mathscr{L}(x, \lambda_*, c) \leq f(x) \quad \forall c \geq c_0. \qquad (8)$$

Applying $(A2)$ again one gets that $\Phi(G(x_*), \lambda_*, c) \leq 0$. On the other hand, from the first inequality in (7) and $(A1)$ it follows that $\Phi(G(x_*), \lambda_*, c) \geq \Phi(G(x_*), 0, c) \geq 0$ for all $c \geq c_0$, which yields that $\Phi(G(x_*), \lambda_*, c) = 0$ for all $c \geq c_0$. Hence and from (8) one gets that $f(x_*) \leq f(x)$ for any feasible point $x \in U$, which implies that x_* is a locally optimal solution of the problem $(\mathcal{P})$.

Repeating the same argument as above with U replaced by A one obtains that if (x_*, λ_*) is a global saddle point of $\mathscr{L}(x, \lambda, c)$, then x_* is a globally optimal solution of the problem $(\mathcal{P})$. $\qquad\square$

Remark 6. Let assumptions $(A1)$–$(A3)$ be satisfied. From the proof of the proposition above it follows that if $(x_*, \lambda_*) \in A \times \Lambda$ is such that

$$\sup_{\lambda \in \Lambda} \mathscr{L}(x_*, \lambda, c_0) \leq \mathscr{L}(x_*, \lambda_*, c_0) < +\infty$$

for some $c_0 > 0$, then x_* is feasible point of the problem $(\mathcal{P})$, and $\mathscr{L}(x_*, \lambda_*, c_0) = f(x_*)$. If, additionally, assumption $(A4)$ holds true, then $\mathscr{L}(x_*, \lambda_*, c) = f(x_*)$ for all $c \geq c_0$. In particular, if (x_*, λ_*) is a local saddle point of $\mathscr{L}(x, \lambda, c)$, and assumptions $(A1)$–$(A3)$ are valid, then $\mathscr{L}(x_*, \lambda_*, c) = f(x_*)$ for all $c > c_{loc}^*(x_*, \lambda_*)$.

Below we will show that there is a direct connection between local (global) saddle points of the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$ and KKT-pairs of the problem $(\mathcal{P})$ corresponding to locally (globally) optimal solutions of this problem. Since for any KKT-pair (x_*, λ_*) one has $\lambda_* \in K^*$ and $\langle \lambda_*, G(x_*) \rangle = 0$, it is natural to ask when these conditions are satisfied for local or global saddle points. The two following propositions give an answer to this question.

Proposition 2. *Let assumptions* (A1), (A3) *and* (A8) *be valid, and let a pair* $(x_*, \lambda_*) \in A \times \Lambda$ *satisfy the inequalities*

$$\sup_{\lambda \in \Lambda} \mathscr{L}(x_*, \lambda, c) \leq \mathscr{L}(x_*, \lambda_*, c) < +\infty \tag{9}$$

for some $c > 0$. *Then* $\lambda_* \in K^*$.

Proof. Arguing by reductio ad absurdum, suppose that $\lambda_* \notin K^*$. Repeating the first part of the proof of Proposition 1 one can easily verify that x_* is a feasible point of the problem $(\mathcal{P})$. Then applying (A1) and (A8) one gets that $\Phi(G(x_*), \lambda_*, c) < 0 \leq \Phi(G(x_*), 0, c)$ for all $c > 0$. Hence for any $c > 0$ one has $\mathscr{L}(x_*, \lambda_*, c) < \mathscr{L}(x_*, 0, c)$, which contradicts (9). $\square$

Let us demonstrate that the proposition above does not hold true in the case when assumption (A8) is not valid. To this end, let us consider the exponential penalty function from Example 5 with $\Lambda = Y^* = \mathbb{R}^l$ (note that in this case assumption (A8) is not satisfied).

Example 17. Let $X = A = \mathbb{R}^2$. Consider the following optimization problem

$$\min f(x_1, x_2) = x_1^2 + x_2^2 \qquad \text{subject to}$$

$$g_1(x_1, x_2) = x_1 + x_2 + 2 \leq 0, \quad g_2(x_1, x_2) = \frac{1}{2}(x_1 + 2)^2 + \frac{1}{2}(x_2 + 2)^2 - 1 \leq 0.$$

Let $x_* = (-1, -1)$ and $\lambda_* = (-1, 3)$. Let us show that the pair (x_*, λ_*) is a global saddle point of the exponential penalty function $\mathscr{L}(x, \lambda, c)$ despite the fact that $\lambda_* \notin K^* = \mathbb{R}_+^2$. Indeed, by definition

$$\mathscr{L}(x, \lambda, c) = f(x) + \frac{\lambda_1}{c}\left(e^{cg_1(x)} - 1\right) + \frac{\lambda_2}{c}\left(e^{cg_2(x)} - 1\right).$$

The equalities $g_1(x_*) = g_2(x_*) = 0$ imply that $\mathscr{L}(x_*, \cdot, c) \equiv f(x_*) = 2$, which yields that

$$\sup_{\lambda \in \mathbb{R}^2} \mathscr{L}(x_*, \lambda, c) = \mathscr{L}(x_*, \lambda_*, c) = 2 \quad \forall c > 0$$

Let us prove that x_* is a point of global minimum of the function $\mathscr{L}(\cdot, \lambda_*, c)$ for any $c > 0$. Then one gets that (x_*, λ_*) is a global saddle point of the exponential penalty function for the above problem.

Note that $g_2(x) \geq g_1(x)$ for all $x \in \mathbb{R}^2$. Therefore taking into account the fact that $\lambda^* = (-1, 3)$ one gets that for any $x \in \mathbb{R}^2$ and $c > 0$ one has

$$\mathscr{L}(x, \lambda_*, c) = f(x) + \frac{2}{c}\left(e^{cg_2(x)} - 1\right) + \frac{1}{c}\left(e^{cg_2(x)} - e^{cg_1(x)}\right) \geq f_c(x), \tag{10}$$

where $f_c(x) = f(x) + 2(e^{cg_2(x)} - 1)/c$. It is easy to see that for any $c > 0$ the function $f_c(x)$ is convex, and $\nabla f_c(x_*) = 0$, which implies that x_* is a point of global minimum of f_c for all $c > 0$. Hence taking into account inequality (10), and the fact that $\mathscr{L}(x_*, \lambda_*, c) = f_c(x_*) = 2$ one obtains the required result.

Remark 7. As the previous example shows, one needs to properly choose the cone Λ in order to ensure the validity of the inclusion $\lambda_* \in K^*$. Alternatively, one can guarantee the validity of the inclusion $\lambda_* \in K^*$ with the use of a suitable constraint qualification. In particular, it is easy to check that if (x_*, λ_*) is a local saddle point of the exponential penalty function with $\Lambda = \mathbb{R}^l$, and LICQ holds true at x_*, then $\lambda_* \in K^*$. On the other hand, note that if at least one of the inequality constraints is not active at x_*, say $g_i(x_*) < 0$, then for (x_*, λ_*) to be a local saddle point of the exponential penalty function it is necessary that $\lambda_i \geq 0$ for all $\lambda \in \Lambda$.

Arguing in a similar way to the proof of Proposition 2 one can check that the following result holds true.

Proposition 3. *Let assumptions* (A1), (A3) *and* (A9) *hold true, and let a pair* $(x_*, \lambda_*) \in A \times \Lambda$ *satisfy inequalities* (9) *for some* $c > 0$. *Then* $\langle \lambda_*, G(x_*) \rangle = 0$.

Remark 8. Let us note that if

$$\Phi(y, \lambda, c) = \langle \lambda, \Phi_1(y, c) \rangle + \Phi_2(y, c) \tag{11}$$

for some functions Φ_1 and Φ_2, and (x_*, λ_*) is a local saddle point of $\mathscr{L}(x, \lambda, c)$ with $\Lambda = Y^*$, then

$$\langle \lambda, \Phi_1(G(x_*), c) \rangle \leq \langle \lambda_*, \Phi_1(G(x_*), c) \rangle \quad \forall \lambda \in Y^* \quad \forall c > c^*_{loc}(x_*, \lambda_*),$$

which in all particular examples presented above implies that $\langle \lambda_*, G(x_*) \rangle = 0$. Note also that assumption (A9) is satisfied with $\Lambda = Y^*$ for all particular augmented Lagrangian functions that cannot be represented in the form (11).

Let us now provide several useful characterizations of global saddle points of the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$. Global saddle points of $\mathscr{L}(x, \lambda, c)$ can be characterized in terms of solutions of the *augmented dual problem*. Recall that the augmented dual problem of $(\mathcal{P})$ associated with the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$ has the form

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

where $\Theta(\lambda, c) = \inf_{x \in A} \mathscr{L}(x, \lambda, c)$. Note that if assumption (A2) is satisfied, then *the weak duality* between $(\mathcal{P})$ and the augmented dual problem holds, i.e.

$$\Theta(\lambda, c) \leq f(x) \quad \forall (\lambda, c) \in \Lambda \times (0, +\infty) \quad \forall x \in \Omega,$$

where Ω is the feasible set of the problem $(\mathcal{P})$. One says that *the zero duality gap property* holds true for the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$, if

$$\inf_{x \in \Omega} f(x) = \sup \big\{ \Theta(\lambda, c) \,\big|\, \lambda \in \Lambda, \ c > 0 \big\}.$$

Now, we can obtain a simple and well-known characterization of global saddle points in terms of globally optimal solutions of the augmented dual problem.

Proposition 4. *Suppose that assumptions* (A1)–(A4) *are satisfied. If* (x_*, λ_*) *is a global saddle point of the augmented Lagrangian* $\mathscr{L}(x, \lambda, c)$, *then for any* $c > c^*(x_*, \lambda_*)$ *the pair* (λ_*, c) *is a globally optimal solution of the augmented dual problem, and the zero duality gap property holds true. Conversely, if* (λ_*, c_*) *is a globally optimal solution of the augmented dual problem, and the zero duality gap property holds true, then for any globally optimal solution* x_* *of the problem* $(\mathcal{P})$ *the pair* (x_*, λ_*) *is a global saddle point of* $\mathscr{L}(x, \lambda, c)$ *and* $c_* \geq c^*(x_*, \lambda_*)$.

Proof. Let (x_*, λ_*) be a global saddle point of $\mathscr{L}(x, \lambda, c)$. Applying Remark 6 one obtains that for any $c > c^*(x_*, \lambda_*)$ the following inequalities hold true

$$\sup_{\lambda \in \Lambda} \mathscr{L}(x_*, \lambda, c) \leq \mathscr{L}(x_*, \lambda_*, c) = f(x_*) \leq \inf_{x \in A} \mathscr{L}(x, \lambda_*, c).$$

Consequently, taking into account the definition of Θ one gets that

$$\Theta(\lambda, c) \leq f(x_*) \leq \Theta(\lambda_*, c) \quad \forall \lambda \in \Lambda \quad \forall c > c^*(x_*, \lambda_*).$$

Assumption $(A4)$ implies that $\Theta(\lambda, c)$ is non-decreasing in c for any $\lambda \in \Lambda$. Therefore $\Theta(\lambda, c) \leq f(x_*)$ for all $(\lambda, c) \in \Lambda \times (0, +\infty)$. Hence taking the supremum over all $(\lambda, c) \in \Lambda \times (0, +\infty)$ one obtains that

$$\sup \big\{ \Theta(\lambda, c) \bigm| (\lambda, c) \in \Lambda \times (0, +\infty) \big\} \leq f(x_*) \leq \Theta(\lambda_*, c) \quad \forall c > c^*(x_*, \lambda_*).$$

Consequently, (λ_*, c) is a globally optimal solution of the augmented dual problem for any $c > c^*(x_*, \lambda_*)$, and the zero duality gap property holds true, since x_* is a globally optimal solution of the problem $(\mathcal{P})$ by virtue of Proposition 1.

Let, now, (λ_*, c_*) be a globally optimal solution of the augmented dual problem, x_* be a globally optimal solution of the problem $(\mathcal{P})$, and suppose that the zero duality gap property holds true. Then with the use of the fact that Θ is non-decreasing in c due to $(A4)$ one gets that $\Theta(\lambda_*, c) = f(x_*)$ for all $c \geq c_*$. Applying assumption $(A2)$ (recall that x_* is a feasible) one obtains that $\mathscr{L}(x_*, \lambda_*, c) \leq f(x_*)$ for any $c > 0$. Therefore taking into account the definition of Θ one gets that

$$\mathscr{L}(x_*, \lambda_*, c) \leq f(x_*) = \Theta(\lambda_*, c) := \inf_{x \in A} \mathscr{L}(x, \lambda_*, c) \quad \forall c \geq c_*, \qquad (12)$$

which implies that $\mathscr{L}(x_*, \lambda_*, c) = f(x_*)$. Applying $(A2)$ again one obtains that

$$\mathscr{L}(x_*, \lambda, c) \leq f(x_*) = \mathscr{L}(x_*, \lambda_*, c) \quad \forall (\lambda, c) \in \Lambda \times [c_*, +\infty). \qquad (13)$$

Combining (13) and (12) one gets that (x_*, λ_*) is a global saddle point of $\mathscr{L}(x, \lambda, c)$ and $c_* \geq c^*(x_*, \lambda_*)$. $\qquad \square$

Corollary 1. *Let assumptions $(A1)$–$(A4)$ be satisfied, and suppose that the zero duality gap property holds true for $\mathscr{L}(x, \lambda, c)$. Then a global saddle point of $\mathscr{L}(x, \lambda, c)$ exists if and only if there exists a globally optimal solution of the augmented dual problem.*

Corollary 2. *Let assumptions $(A1)$–$(A4)$ be satisfied, and let (x_*, λ_*) be a global saddle point of $\mathscr{L}(x, \lambda, c)$. Then for any globally optimal solution z_* of $(\mathcal{P})$ the pair (z_*, λ_*) is a global saddle point of $\mathscr{L}(x, \lambda, c)$ as well.*

Let us also obtain a characterization of global saddle points of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$ in terms of the behaviour of *the optimal value function* (or *the perturbation function*)

$$\beta(p) = \inf \big\{ f(x) \bigm| x \in A, \ G(x) - p \in K \big\} \quad \forall p \in Y$$

of the problem $(\mathcal{P})$ (cf. [68], Section 11.K, and [72, 85]).

Proposition 5. *Let assumptions* $(A1)$–$(A5)$ *be satisfied. Then a pair* $(x_*, \lambda_*) \in A \times \Lambda$ *is a global saddle point of* $\mathscr{L}(x, \lambda, c)$ *if and only if* x_* *is a globally optimal solution of the problem* $(\mathcal{P})$, *and there exists* $c_0 > 0$ *such that*

$$\beta(p) \geq \beta(0) - \Phi(p, \lambda_*, c) \quad \forall p \in Y \quad \forall c \geq c_0. \tag{14}$$

Proof. Let (x_*, λ_*) be a global saddle point of $\mathscr{L}(x, \lambda, c)$. By Proposition 1 the point x_* is a globally optimal solution of the problem $(\mathcal{P})$, which implies that $f(x_*) = \beta(0)$.

Fix an arbitrary $p \in Y$. If there is no $x \in A$ such that $G(x) - p \in K$, then $\beta(p) = +\infty$, and inequality (14) is valid. Otherwise, for any such x one has

$$\mathscr{L}(x, \lambda_*, c) \geq \mathscr{L}(x_*, \lambda_*, c) = f(x_*) = \beta(0) \quad \forall c > c^*(x_*, \lambda_*)$$

by virtue of the fact that (x_*, λ_*) is a global saddle point of $\mathscr{L}(x, \lambda, c)$ and Remark 6. Hence and from $(A5)$ it follows that $f(x) \geq \beta(0) - \Phi(p, \lambda_*, c)$ for all $c > c^*(x_*, \lambda_*)$. Taking the infimum over all $x \in A$ for which $G(x) - p \in K$ one obtains that (14) holds true.

Let, now, x_* be a globally optimal solution of $(\mathcal{P})$, and let (14) be valid. Fix arbitrary $c \geq c_0$ and $x \in A$. Define $p = G(x)$. Then (14) implies that

$$f(x) \geq \beta(p) \geq f(x_*) - \Phi(G(x), \lambda_*, c)$$

or, equivalently, $\mathscr{L}(x, \lambda_*, c) \geq f(x_*)$. Taking the infimum over all $x \in A$ one gets that $\Theta(\lambda_*, c) \geq f(x_*) = \inf_{x \in \Omega} f(x)$. Hence and from the fact that the weak duality between $(\mathcal{P})$ and the augmented dual problem holds by $(A2)$, it follows that the zero duality gap property holds true for $\mathscr{L}(x, \lambda, c)$, and (λ_*, c) is a globally optimal solution of the augmented dual problem. Consequently, applying Proposition 4 one gets that (x_*, λ_*) is a global saddle point. $\qquad\square$

Remark 9. It should be noted that under some additional assumptions inequality (14) holds true if and only if there exist $c_0 > 0$ and a neighbourhood of zero, $U \subset Y$, such that $\beta(p) \geq \beta(0) - \Phi(p, \lambda_*, c)$ for all $p \in U$ and $c \geq c_0$ (cf. [68], Theorem 11.61, and [72], Lemma 3.1). However, we do not present an exact formulation of this result here, and leave it to the interested reader.

5 The Localization Principle for Saddle Points

In this section, we develop a new general method for proving the existence of global saddle points of augmented Lagrangian functions. According to this method, one has to verify that there exists a multiplier λ_* such that for any globally optimal solution x_* of the problem $(\mathcal{P})$ the pair (x_*, λ_*) is a local saddle point of $\mathscr{L}(x, \lambda, c)$ in order to prove the existence of a global saddle point. In turn, the existence of such multiplier λ_* can usually be proved with the use of sufficient optimality conditions. Thus, the general method for proving the existence of global saddle points of augmented Lagrangian functions, that we discuss in this section, allows one to reduce the problem of the existence of global saddle points to a local analysis of sufficient optimality conditions. That is why we call this method *the localization principle*.

Let x_* be a globally optimal solution of the problem $(\mathcal{P})$. Denote by $\Lambda_{loc}(x_*)$ (resp. $\Lambda(x_*)$) the set of all $\lambda_* \in \Lambda$ for which the pair (x_*, λ_*) is a local (resp.

global) saddle point of the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$. Also, denote by Ω_* the set of all globally optimal solutions of the problem $(\mathcal{P})$. Define

$$\Lambda_{loc}(\mathcal{P}) = \bigcap_{x_* \in \Omega_*} \Lambda_{loc}(x_*), \quad \Lambda(\mathcal{P}) = \bigcap_{x_* \in \Omega_*} \Lambda(x_*)$$

Corollary 2 implies that, in actuality, $\Lambda(\mathcal{P}) = \Lambda(x_*)$ for any $x_* \in \Omega_*$, provided assumptions $(A1)$–$(A4)$ are valid.

Observe that $\Lambda(\mathcal{P}) \subseteq \Lambda_{loc}(\mathcal{P})$ by virtue of the fact that every global saddle point of $\mathscr{L}(x, \lambda, c)$ is a local one. Therefore for the existence of a global saddle point of $\mathscr{L}(x, \lambda, c)$ it is necessary that $\Lambda_{loc}(\mathcal{P}) \neq \emptyset$. The localization principle states that under some additional assumptions the condition $\Lambda_{loc}(\mathcal{P}) \neq \emptyset$ is also sufficient for the existence of a global saddle point.

We need an auxiliary result in order to prove the localization principle.

Lemma 1. *Let $\lambda_* \in \Lambda_{loc}(\mathcal{P})$, A be closed, and assumptions $(A2)$, $(A4)$ and $(A6)$ be valid. Suppose also that G is continuous on A, $\mathscr{L}(\cdot, \lambda_*, c)$ is l.s.c. on A for all $c > 0$, and $\mathscr{L}(\cdot, \lambda_*, c_0)$ is bounded from below on A for some $c_0 > 0$. Let, finally, sequences $\{c_n\} \subset [c_0, +\infty)$ and $\{\varepsilon_n\} \subset (0, +\infty)$ be such that $c_n \to +\infty$ and $\varepsilon_n \to 0$ as $n \to \infty$. Then any cluster point of a sequence $\{x_n\} \subset A$ such that*

$$\mathscr{L}(x_n, \lambda_*, c_n) \leq \inf_{x \in A} \mathscr{L}(x, \lambda_*, c_n) + \varepsilon_n \quad \forall n \in \mathbb{N}. \tag{15}$$

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

Proof. Let x_* be a cluster point of a sequence $\{x_n\} \subset A$ satisfying (15). Replacing, if necessary, the sequences $\{c_n\}$, $\{\varepsilon_n\}$ and $\{x_n\}$ by their subsequences, one can suppose that $\{x_n\}$ converges to x_*. Note that $x_* \in A$ due to the fact that A is closed.

Fix an arbitrary $x_0 \in \Omega_*$. Note that $\lambda_* \in \Lambda$, since $\lambda_* \in \Lambda_{loc}(\mathcal{P})$, which implies that $\mathscr{L}(x_0, \lambda_*, c) \leq f(x_0)$ for all $c > 0$ by assumption $(A2)$. Hence and from (15) it follows that

$$\mathscr{L}(x_n, \lambda_*, c_n) \leq f(x_0) + \varepsilon_n \quad \forall n \in \mathbb{N}, \tag{16}$$

which yields that

$$f(x_n) < +\infty, \quad \Phi(G(x_n), \lambda_*, c_n) < +\infty \quad \forall n \in \mathbb{N}. \tag{17}$$

Choose an arbitrary $c > 0$. Since $\lim c_n = +\infty$, there exists $n_0 \in \mathbb{N}$ such that $c_n \geq c$ for all $n \geq n_0$. Hence with the use of $(A4)$ one obtains that $\mathscr{L}(x_n, \lambda_*, c) \leq f(x_0) + \varepsilon_n$ for all $n \geq n_0$. Passing to the limit inferior as $n \to \infty$ one gets that $\mathscr{L}(x_*, \lambda_*, c) \leq f(x_0)$ for all $c > 0$. Let us verify that x_* is a feasible point of the problem $(\mathcal{P})$. Then applying $(A2)$ one obtains that $f(x_*) \leq f(x_0)$, which implies that $x_* \in \Omega_*$ due to the definition of x_0.

Arguing by reductio ad absurdum, suppose that x_* is infeasible, i.e. that $y_* := G(x_*) \notin K$. Then by $(A6)$ there exists $r > 0$ such that

$$\lim_{c \to \infty} \inf\{\Phi(y, \lambda_*, c) - \Phi(y, \lambda_*, c_0) \mid y \in B(y_*, r) \colon \Phi(y, \lambda_*, c_0) < +\infty\} = +\infty. \tag{18}$$

Since G is continuous and $x_n \to x_*$ as $n \to \infty$, there exists $k \in \mathbb{N}$ such that $G(x_n) \in B(y_*, r)$ for all $n \geq k$. Consequently, taking into account (17) and assumption $(A4)$ one gets that for all $n \geq k$ the following inequalities hold true

$$\mathscr{L}(x_n, \lambda_*, c_n) = \mathscr{L}(x_n, \lambda_*, c_0) + \Phi(G(x_n), \lambda_*, c_n) - \Phi(G(x_n), \lambda_*, c_0)$$
$$\geq \gamma + \inf\{\Phi(y, \lambda_*, c_n) - \Phi(y, \lambda_*, c_0) \mid y \in B(y_*, r) \colon \Phi(y, \lambda_*, c_0) < +\infty\},$$

where $\gamma = \inf_{x \in A} \mathscr{L}(x, \lambda_*, c_0) > -\infty$. Hence applying (18) one obtains that $\mathscr{L}(x_n, \lambda_*, c_n) \to +\infty$ as $n \to \infty$, which contradicts (16). Thus, x_* is a feasible point of the problem $(\mathcal{P})$, which completes the proof. $\qquad\square$

Now, we are ready to derive the localization principle for global saddle points of the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$. Denote by $f_* = \inf_{x \in \Omega} f(x)$ the optimal value of the problem $(\mathcal{P})$.

Theorem 1 (Localization Principle). *Let A be closed, G be continuous on A, and $\mathscr{L}(\cdot, \lambda, c)$ be l.s.c. on A for all $\lambda \in \Lambda$ and $c > 0$. Suppose also that assumptions $(A1)$–$(A4)$ and $(A6)$ are satisfied. Then a global saddle point of $\mathscr{L}(x, \lambda, c)$ exists if and only if $\Lambda_{loc}(\mathcal{P}) \neq \emptyset$, and there exist $\lambda_* \in \Lambda_{loc}(\mathcal{P})$ and $c_0 > 0$ such that the set*

$$S(\lambda_*, c_0) := \{x \in A \mid \mathscr{L}(x, \lambda_*, c_0) < f_*\}$$

is either bounded or empty. Furthermore, $\Lambda(\mathcal{P})$ consists of all those $\lambda_* \in \Lambda_{loc}(\mathcal{P})$ for which there exists $c_0 > 0$ such that the set $S(\lambda_*, c_0)$ is either bounded or empty.

Proof. Suppose that there exists a global saddle point (x_*, λ_*) of $\mathscr{L}(x, \lambda, c)$. By Corollary 2, for any globally optimal solution x_0 of $(\mathcal{P})$ the pair (x_0, λ_*) is a global saddle point of $\mathscr{L}(x, \lambda, c)$, which yields that $\lambda_* \in \Lambda_{loc}(\mathcal{P})$. Furthermore, from the definition of global saddle point and Remark 6 it follows that for any $c > c^*(x_*, \lambda_*)$ the set $S(\lambda_*, c)$ is empty.

Suppose, now, that $\lambda_* \in \Lambda_{loc}(\mathcal{P})$, and there exists $c_0 > 0$ such that the set $S(\lambda_*, c_0)$ is either bounded or empty. From $(A4)$ it follows that

$$S(\lambda_*, c) \subseteq S(\lambda_*, c_0) \quad \forall c \geq c_0, \tag{19}$$

which implies, in particular, that $S(\lambda_*, c)$ is bounded (or empty) for any $c \geq c_0$.

If there exists $\overline{c} \geq c_0$ such that $S(\lambda_*, \overline{c}) = \emptyset$, then $\inf_{x \in A} \mathscr{L}(x, \lambda_*, c) \geq f_*$ for all $c \geq \overline{c}$ due to $(A4)$. Hence applying the definition of local saddle point and Remark 6 one gets that for any globally optimal solution x_* of $(\mathcal{P})$, and for all $c > \max\{\overline{c}, c^*_{loc}(x_*, \lambda_*)\}$ the following inequalities hold true

$$\sup_{\lambda \in \Lambda} \mathscr{L}(x_*, \lambda, c) \leq \mathscr{L}(x_*, \lambda_*, c) = f(x_*) = f_* \leq \inf_{x \in A} \mathscr{L}(x, \lambda_*, c),$$

i.e. (x_*, λ_*) is a global saddle point of $\mathscr{L}(x, \lambda, c)$.

Thus, it remains to consider the case when $S(\lambda_*, c) \neq \emptyset$ for all $c \geq c_0$. As it was noted above (see inclusion (19)), the set $S(\lambda_*, c)$ is bounded for all $c \geq c_0$. Therefore, taking into account the fact that $\mathscr{L}(\cdot, \lambda_*, c)$ is l.s.c. on A, one obtains that $\mathscr{L}(\cdot, \lambda_*, c)$ attains a global minimum on A at a point $x(c)$ for all $c \geq c_0$. Choose an increasing unbounded sequence $\{c_n\} \subset [c_0, +\infty)$, and denote

$x_n = x(c_n)$. Inclusion (19) implies that the sequence $\{x_n\}$ is contained in the bounded set $S(\lambda_*, c_0)$. Therefore there exists a subsequence $\{x_{n_k}\}$ converging to a point x_* that belongs to the set A due to the fact that A is closed.

By the definition of x_n one has $\mathscr{L}(x_n, \lambda_*, c_n) = \inf_{x \in A} \mathscr{L}(x, \lambda_*, c_n)$. Consequently, applying Lemma 1 one obtains that x_* is a globally optimal solution of the problem $(\mathcal{P})$. Hence taking into account the fact that $\lambda_* \in \Lambda_{loc}(\mathcal{P})$ and applying Remark 6 one gets that there exist $\overline{c} > 0$ and a neighbourhood U of x_* such that

$$\sup_{\lambda \in \Lambda} \mathscr{L}(x_*, \lambda, c) \le \mathscr{L}(x_*, \lambda_*, c) = f(x_*) \le \inf_{x \in U \cap A} \mathscr{L}(x, \lambda_*, c) \quad \forall c \ge \overline{c}. \quad (20)$$

Since $\{x_{n_k}\}$ converges to x_*, and $\{c_n\}$ is an increasing unbounded sequence, there exists $k_0 \in \mathbb{N}$ such that $x_{n_k} \in U$ and $c_{n_k} \ge \overline{c}$ for all $k \ge k_0$. Hence applying (20) and the fact that $x_* \in \Omega_*$ one obtains that $\mathscr{L}(x_{n_k}, \lambda_*, c_{n_k}) \ge f(x_*) = f_*$ for all $k \ge k_0$, which contradicts our assumption that $S(\lambda_*, c) \ne \emptyset$ for all $c \ge c_0$ due to the fact that x_n is a global minimizer of $\mathscr{L}(\cdot, \lambda_*, c_n)$ on the set A. $\quad \square$

Corollary 3 (Localization Principle). *Let all assumptions of the theorem above be satisfied, and suppose that the set A is compact. Then $\Lambda(\mathcal{P}) = \Lambda_{loc}(\mathcal{P})$.*

Remark 10. The localization principle for global saddle points of augmented Lagrangian functions for cone constrained programming problems (Theorem 1) unifies and sharpens most of the known results on the existence of global saddle points of augmented Lagrangians for mathematical programming problems ([50], Thm. 3.3; [49], Thm. 3; [56], Thm. 4.1; [98], Thm. 3.3; [89], Thm. 2; [82], Thms. 3.3 and 3.4; [79], Thms. 3.1, 3.3 and 3.4; [81], part 3 of Cor. 4.1, 4.3 and 4.4, part 2 of Cor. 4.2 and 4.5, and parts 3 and 4 of Cor. 4.6), nonlinear second order cone programming problems ([97], Thm. 3.1) and nonlinear semidefinite programming ([90], Thm. 4; [55], Thm. 4.4). Furthermore, the localization principle provides first simple *necessary and sufficient* conditions for the existence of global saddle points.

Let us note that if, instead of assumption $(A6)$, stronger assumption $(A7)$ holds true, then instead of $S(\lambda_*, c)$ one can use a smaller set.

Proposition 6. *Let assumptions $(A4)$ and $(A7)$ be valid, and let the function $\mathscr{L}(\cdot, \lambda_*, c_0)$ be bounded from below on A for some $\lambda_* \in \Lambda$ and $c_0 > 0$. Then the set $S(\lambda_*, c_1)$ is bounded for some $c_1 > 0$ iff there exist $c_2 > 0$ and $\alpha > 0$ such that the set $Q(\lambda_*, c_2, \alpha) := \{x \in A \mid \mathscr{L}(x, \lambda_*, c_2) < f_*, \, d(G(x), K) < \alpha\}$ is bounded.*

Proof. The validity of the "only if" part of the proposition follows directly from the obvious inclusion $Q(\lambda_*, c, \alpha) \subseteq S(\lambda_*, c)$ that holds true for all $c > 0$ and $\alpha > 0$. Therefore, let us prove the "if" part. Namely, let us show that for any $\alpha > 0$ there exists $\overline{c} > 0$ such that $S(\lambda_*, c) \subseteq Q(\lambda_*, c, \alpha)$ for all $c \ge \overline{c}$. Then taking into account the inclusion $Q(\lambda_*, c', \alpha) \subseteq Q(\lambda_*, c'', \alpha)$, that holds true for all $c' \ge c''$ due to $(A4)$, one obtains the required result.

Fix $\alpha > 0$. Let $x \in A$ be such that $d(G(x), K) \ge \alpha$ and $\Phi(G(x), \lambda_*, c_0) < +\infty$. Then for all $c \ge c_0$ one has

$$\mathscr{L}(x, \lambda_*, c) = \mathscr{L}(x, \lambda_*, c_0) + \Phi(G(x), \lambda_*, c) - \Phi(G(x), \lambda_*, c_0)$$
$$\ge \gamma + \inf\{\Phi(y, \lambda_*, c) - \Phi(y, \lambda_*, c_0) \mid y \in Y : d(y, K) \ge \alpha, \, \Phi(y, \lambda_*, c_0) < +\infty\},$$

where $\gamma = \inf_{x \in A} \mathscr{L}(x, \lambda_*, c_0) > -\infty$. Applying $(A7)$ one gets that there exists $\overline{c} \geq c_0$ (that depends only on λ_*, c_0 and $\alpha > 0$) such that $\mathscr{L}(x, \lambda_*, c) \geq f_*$ for all $c \geq \overline{c}$. In other words, $S(\lambda_*, c) = Q(\lambda_*, c, \alpha)$ for all $c \geq \overline{c}$. $\qquad\square$

Remark 11. Let the following assumption be valid:

(A13) $\forall \lambda \in \Lambda \; \forall \varepsilon > 0 \; \forall c_0 > 0 \; \exists c \geq c_0 \; \exists \alpha > 0$ such that $\Phi(G(x), \lambda, c) > -\varepsilon$ for all $x \in A$: $\operatorname{dist}(G(x), K) < \alpha$;

Then one can easily verify that the set $Q(\lambda, c, \alpha)$ from the proposition above is bounded for some $c > 0$ and $\alpha > 0$, provided there exists $\gamma > 0$ such that the set

$$\left\{ x \in A \mid f(x) < f_* + \gamma, \; d(G(x), K) < \gamma \right\} \tag{21}$$

is bounded. Note that the assumption on the boundedness of set (21) is utilized in most of the known results on existence of global saddle points of augmented Lagrangian functions (cf., e.g., [79], Thms. 3.1–3.4; [50], Thm. 3.3; [89], Thm. 2; [98], Thm. 3.3; [99], Thms. 3.1 and 3.2; [97], Thm. 3.1, etc.). Observe also that assumption $(A13)$ is satisfied in Examples 1 and 15, provided $\sigma(y) \geq \omega(\|y\|)$ for some non-negative continuous function ω such that $\liminf_{t \to +\infty} \omega(t)/t > 0$. This assumptions is always valid in Examples 2–4, 8, 10 and 12. Finally, $(A13)$ is satisfied in Examples 5–7, 9, 11, 13, 14 and 16 if either the function ϕ (or ψ) is bounded from below or the sets $g_i(A) \cap K_i$, $i \in I$, are bounded, where K_i is either $\mathbb{R}_-$ or the second order cone or the cone of negative semidefinite matrices or the cone of non-positive continuous functions, depending on the context. Let us finally note that the assumption on the boundedness of the sets $g_i(A) \cap K_i$ was utilized in various paper on augmented Lagrangian functions (see, e.g., [83, 50, 81])

Remark 12. Note that if assumption $(A7)$ is valid, and the set $Q(\lambda_*, c, \alpha)$ is bounded for some $c > 0$ and $\alpha > 0$, but the function $\mathscr{L}(\cdot, \lambda_*, c)$ is not bounded from below for any $c > 0$, then one can redefine the function $\Phi(y, \lambda, c)$ in order to guarantee the existence of global saddle points. Namely, define

$$\widehat{\Phi}(y, \lambda, c) = \begin{cases} \left(\alpha - \operatorname{dist}(y, K)^{\varkappa}\right) \cdot \Phi\left(\dfrac{1}{\alpha - \operatorname{dist}(y, K)^{\varkappa}} y, \lambda, c\right), & \text{if } d(y, K)^{\varkappa} < \alpha, \\ +\infty, & \text{otherwise} \end{cases}$$

for some $\alpha > 0$ and $\varkappa > 0$. One can verify that such transformation not only guarantees the boundedness from below of the function $\mathscr{L}(\cdot, \lambda, c)$, but also preserves local and global saddle points, the boundedness of the set $Q(\lambda_*, c, \alpha)$, and general properties of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$ (such as continuity, differentiability if $\varkappa \geq 2$, etc.) in all particular examples presented above, except for Example 11. Finally, it should be noted that there is no need for such transformation of the function $\Phi(y, \lambda, c)$ in Example 7, and Examples 11, 13 and 16 in the case when $\varepsilon_0 < +\infty$, since the augmented Lagrangian functions from these examples are bounded from below iff the set $Q(\lambda_*, c, \alpha)$ is bounded for some $c > 0$ and $\alpha > 0$.

The following proposition contains a simple reformulation of the boundedness condition on the set $S(\lambda_*, c)$ that will allow us to give an illuminating interpretation of the localization principle.

Proposition 7. *Let $\lambda_* \in \Lambda_{loc}(\mathcal{P})$, A be closed, $\mathscr{L}(\cdot, \lambda_*, c)$ be l.s.c. on A for all $c > 0$, and G be continuous on A. Suppose also that assumptions $(A2)$, $(A4)$ and $(A6)$ are satisfied. Then for the existence of $c_0 > 0$ such that the set $S(\lambda_*, c_0)$ is either bounded or empty it is necessary and sufficient that there exist $c_0 > 0$ and $R > 0$ such that for any $c \geq c_0$ there exists $x(c) \in \arg\min_{x \in A} \mathscr{L}(x, \lambda_*, c)$ with $\|x(c)\| \leq R$.*

Proof. Suppose that exists $c_0 > 0$ such that the set $S(\lambda_*, c_0)$ is either bounded or empty. Note that $(A4)$ implies that (19) holds true. If there exists $c_1 \geq c_0$ such that $S(\lambda_*, c_1) = \emptyset$, then taking into account $(A2)$ and $(A4)$ one gets that $x_* \in \arg\min_{x \in A} \mathscr{L}(x, \lambda_*, c)$ for all $c \geq c_1$, where x_* is a globally optimal solution of $(\mathcal{P})$. On the other hand, if $S(\lambda_*, c) \neq \emptyset$ for all $c \geq c_0$, then taking into account (19) and the fact that $\mathscr{L}(\cdot, \lambda_*, c)$ is l.s.c. on A one obtains that for any $c \geq c_0$ there exists $x(c) \in \arg\min_{x \in A} \mathscr{L}(x, \lambda_*, c)$ such that $x(c) \in S(\lambda_*, c_0)$. Since $S(\lambda_*, c_0)$ is bounded, there exists $R > 0$ such that $\|x(c)\| \leq R$ for all $c \geq c_0$.

Suppose, now, that there exist $c_0 > 0$ and $R > 0$ such that for any $c \geq c_0$ there exists $x(c) \in \arg\min_{x \in A} \mathscr{L}(x, \lambda_*, c)$ with $\|x(c)\| \leq R$. Choose an increasing unbounded sequence $\{c_n\} \subset [c_0, +\infty)$, and define $x_n = x(c_n)$. Without loss of generality one can suppose that the sequence $\{x_n\}$ converges to a point x_*. By Lemma 1, the point x_* is a globally optimal solution of $(\mathcal{P})$.

Taking into account Remark 6 and the fact that $\lambda_* \in \Lambda_{loc}(\mathcal{P})$ one gets that there exist $r > 0$ and a neighbourhood U of x_* such that

$$\sup_{\lambda \in \Lambda} \mathscr{L}(x_*, \lambda, c) \leq \mathscr{L}(x_*, \lambda_*, c) = f(x_*) \leq \inf_{x \in U \cap A} \mathscr{L}(x, \lambda_*, c) \quad \forall c \geq r.$$

Since $x_n \to x_*$ and $c_n \to +\infty$ as $n \to \infty$, there exists $n \in \mathbb{N}$ for which $x_n \in U$ and $c_n \geq r$. For any such n one has $\mathscr{L}(x_n, \lambda_*, c_n) \geq f(x_*) = f_*$, which implies that $S(\lambda_*, c_n) = \emptyset$ due to the definition of x_n. $\qquad\square$

Remark 13. With the use of the proposition above we can give the following interpretation of the localization principle. Roughly speaking, according to the localization principle a global saddle point of the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$ exists if and only if there exists a multiplier λ_* such that for any globally optimal solution x_* of $(\mathcal{P})$ the pair (x_*, λ_*) is a local saddle point of $\mathscr{L}(x, \lambda, c)$, the function $\mathscr{L}(\cdot, \lambda_*, c)$ attains a global minimum on A for any sufficiently large $c > 0$, and global minimizers of $\mathscr{L}(\cdot, \lambda_*, c)$ on A do not escape to infinity as $c \to +\infty$.

Note that, in particular, the exponential penalty function (Example 5) and the p-th power augmented Lagrangian (Example 8) do not satisfy assumption $(A6)$. Therefore one cannot apply the localization principle in order to obtain necessary and sufficient conditions for the existence of global saddle points of these augmented Lagrangian functions. Let us show that global saddle points of these augmented Lagrangian functions exist only under a rather restrictive assumption on the problem $(\mathcal{P})$.

Example 18. Consider the following nonlinear programming problem

$$\min f(x) \quad \text{subject to} \quad g_1(x) \leq 0, \quad g_2(x) \leq 0. \tag{22}$$

In this case $K = \mathbb{R}_-^2$. Let $\Lambda = K^* = \mathbb{R}_+^2$, and let

$$\mathscr{L}(x, \lambda, c) = f(x) + \frac{\lambda_1}{c}\left(e^{cg_1(x)} - 1\right) + \frac{\lambda_2}{c}\left(e^{cg_2(x)} - 1\right)$$

be the exponential penalty function for this problem. Suppose that a pair (x_*, λ_*) with $\lambda_* \in \mathbb{R}^2_+$ is a global saddle point of $\mathscr{L}(x, \lambda, c)$, i.e.

$$\sup_{\lambda \in \mathbb{R}^2_+} \mathscr{L}(x_*, \lambda, c) \leq \mathscr{L}(x_*, \lambda_*, c) \leq \inf_{x \in X} \mathscr{L}(x, \lambda_*, c) \qquad (23)$$

for any sufficiently large $c > 0$. Suppose, in addition, that $g_2(x_*) < 0$. Then $(\lambda_*)_2 = 0$ by Proposition 3, which implies that $\mathscr{L}(x, \lambda_*, c)$ does not depend on $g_2(x)$. Consequently, taking the supremum in (23) over only those $\lambda \in \mathbb{R}^2_+$ for which $\lambda_2 = 0$ one gets that $(x_*, (\lambda_*)_1)$ is a global saddle point of the reduced augmented Lagrangian function $\mathscr{L}_0(x, \lambda_1, c) = \mathscr{L}(x, (\lambda_1, 0), c)$. Hence applying Proposition 1 one gets that x_* is a globally optimal solution of problem (22), and of the problem of minimizing $f(x)$ subject to $g_1(x) \leq 0$.

Thus, if a global saddle point of the exponential penalty function exists, then the removal of inequality constraints that are inactive at a globally optimal solution must not change the global optimality of the given solution. After inactive inequality constraints have been removed, the exponential penalty function will satisfy assumptions $(A6)$ and $(A7)$ for a given multiplier $\lambda_* \in \Lambda_{loc}(\mathcal{P})$, provided all components of λ_* are positive (or, equivalently, if λ_* satisfies the strict complementarity condition at all globally optimal solutions of the problem $(\mathcal{P})$). Then arguing in the same way as in the proof of the localization principle one can obtain simple necessary and sufficient conditions for the existence of a global saddle point of the exponential penalty function.

Namely, one can verify that the following result holds true

Proposition 8. *Let A be closed, and $\mathscr{L}(x, \lambda, c)$ be the exponential penalty function for the problem*

$$\min f(x) \quad subject\ to \quad g_i(x) \leq 0, \quad i \in I, \quad x \in A. \qquad (\mathcal{P}_1)$$

Let also f be l.s.c. on A, and g_i, $i \in I$, be continuous on A. Then a global saddle point of $\mathscr{L}(x, \lambda, c)$ exists if and only if there exist $\lambda_* \in \Lambda_{loc}(\mathcal{P}_1)$ and $c_0 > 0$ such that the set $S(\lambda_*, c_0)$ is either bounded or empty, every globally optimal solutions of the problem $(\mathcal{P}_1)$ is a globally optimal solution of the problem

$$\min f(x) \quad subject\ to \quad g_i(x) \leq 0, \quad i \in I(\lambda_*), \quad x \in A, \qquad (\mathcal{P}_2)$$

and $\lambda_* \in \Lambda_{loc}(\mathcal{P}_2)$, where $I(\lambda_*) = \{i \in I \mid (\lambda_*)_i > 0\}$.

Remark 14. (i) It should be noted that the example and the proposition above were inspired by "the separation condition" (3.11) from [79], where it was used in order to obtain sufficient conditions for the existence of a global saddle point of the exponential penalty function ([79], Theorem 3.2). Let us note that Proposition 8 significantly sharpens Theorem 3.2 from [79], since we do not assume that A is compact or that a globally optimal solution of the problem $(\mathcal{P}_1)$ is unique, and obtain *necessary and sufficient* conditions for the existence of a global saddle point, in contrast to only sufficient conditions in [79].

(ii) Observe that every globally optimal solution of the problem $(\mathcal{P}_1)$ is a globally optimal solution of the problem $(\mathcal{P}_2)$ and $\Lambda_{loc}(\mathcal{P}_1) = \Lambda_{loc}(\mathcal{P}_2)$ in the case when the problem $(\mathcal{P}_1)$ is convex. However, note also that in the convex case one obviously has $\Lambda(\mathcal{P}) = \Lambda_{loc}(\mathcal{P}) = \Lambda_{loc}(x_*)$ for any $x_* \in \Omega_*$, and the localization principle holds trivially.

(iii) The interested reader can extend Proposition 8 to case of other augmented Lagrangians, such as the ones from Examples 11, 13 and 16 with $\varepsilon_0 = +\infty$.

6 Applications of the Localization Principle for Global Saddle Points

The main goal of this section is to demonstrate that the localization principle allows one to easily prove the existence of global saddle points of augmented Lagrangian functions with the use of sufficient optimality conditions. Below, we suppose that $X = \mathbb{R}^d$. For the sake of simplicity, in this section we also suppose that the set A is convex.

At first, we study the existence of local saddle points. Our aim is to establish a connection between KKT-pairs of the problem $(\mathcal{P})$ and local saddle points of the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$.

Let $(\mathcal{P})$ be a cone constrained minimax problem, i.e. let $f(x)$ have the form $f = \max_{k \in M} f_k$, where $f_k \colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ are given functions, and $M = \{1, \ldots, m\}$. Denote by $L(x, \lambda) = f(x) + \langle \lambda, G(x) \rangle$ the standard Lagrangian function for the problem $(\mathcal{P})$. For any $\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{R}_+^m$ denote $L_0(x, \lambda, \alpha) = \sum_{k=1}^m \alpha_k f_k(x) + \langle \lambda, G(x) \rangle$. Finally, denote $M(x) = \{k \in M \mid f_k(x) = f(x)\}$.

Suppose that the functions f_k, $k \in M$, and G are twice differentiable at a point $x_* \in A$. Recall that a pair (x_*, λ_*) is called a *KKT-pair* of the problem $(\mathcal{P})$ if $G(x_*) \in K$, $\lambda_* \in K^*$, $\langle \lambda_*, G(x_*) \rangle = 0$ and

$$\big[L(\cdot, \lambda_*)\big]'(x_*, h) \geq 0 \quad \forall h \in T_A(x_*).$$

Here $[L(\cdot, \lambda_*)]'(x_*, h)$ is the directional derivative of the function $L(\cdot, \lambda_*)$ at the point x_* in a direction h, and $T_A(x_*)$ is the contingent cone to A at x_*. Any $\lambda_* \in K^*$ such that (x_*, λ_*) is a KKT-pair of $(\mathcal{P})$ is called *a Lagrange multiplier* of $(\mathcal{P})$ at x_*.

One can easily verify that a pair (x_*, λ_*) is a KKT-pair of $(\mathcal{P})$ iff $G(x_*) \in K$, $\lambda_* \in K^*$, $\langle \lambda_*, G(x_*) \rangle = 0$, and there exists a vector $\alpha \in \mathbb{R}_+^m$ (that is sometimes called *a Danskin-Demyanov multiplier*) such that $\alpha_k = 0$ for any $k \notin M(x_*)$, and

$$\sum_{k=1}^m \alpha_k = 1, \quad \langle D_x L_0(x_*, \lambda_*, \alpha), h \rangle \geq 0 \quad \forall h \in T_A(x_*). \tag{24}$$

The set of all such α is denoted by $\alpha(x_*, \lambda_*)$.

Let us also recall the second order necessary optimality conditions (cf. [4], Theorem 3.45 and Proposition 3.46). One says that a KKT-pair (x_*, λ_*) satisfies *the second order necessary optimality condition* if

$$\sup_{\alpha \in \alpha(x_*, \lambda_*)} \langle h, D_{xx}^2 L_0(x_*, \lambda_*, \alpha) h \rangle - \sigma(\lambda_*, \mathcal{T}(h)) \geq 0$$

$$\forall h \in C(x_*, \lambda_*) \colon [L(\cdot, \lambda_*)]'(x_*, h) = 0. \tag{25}$$

Here $\sigma(\lambda_*, \mathcal{T}(h)) = \sup_{y \in \mathcal{T}(h)} \langle \lambda_*, y \rangle$, $\mathcal{T}(h) = T_K^2(G(x_*), DG(x_*)h)$ is *the outer second order tangent set* to the set K at the point $G(x_*)$ in the direction $DG(x_*)h$ (see [4], Definition 3.28), and

$$C(x_*, \lambda_*) = \Big\{ h \in T_A(x_*) \mid DG(x_*)h \in T_K\big(G(x_*)\big), \ \langle \lambda_*, DG(x_*)h \rangle = 0 \Big\}.$$

is *the critical cone* at the point x_*. Note that if the cone K is polyhedral, then $\sigma(\lambda_*, \mathcal{T}(h)) = 0$ for any $h \in C(x_*, \lambda_*)$, and optimality condition (25) is reduced

to the standard optimality condition. On the other hand, in the general case $\sigma(\lambda_*, \mathcal{T}(h)) \le 0$ for all $h \in C(x_*, \lambda_*)$, which means that optimality condition (25) is weaker than the standard one. It should be mentioned that the term $\sigma(\lambda_*, \mathcal{T}(h))$, in a sense, represents the contribution of the curvature of the cone K at the point $G(x_*)$.

In order to study a connection between local saddle points and KKT-pairs, we need to introduce an additional assumption on differentiability properties of the function Φ.

Definition 3. Let assumption $(A11)$ hold true, and G be twice Fréchet differentiable at a feasible point $x_* \in A$. Let also $\lambda_* \in K^*$ be such that $\langle \lambda_*, G(x_*) \rangle = 0$. One says that the function $\Phi(G(x), \lambda, c)$ admits *the second order expansion* in x at (x_*, λ_*), if for all $c > 0$ there exists a positively homogeneous of degree 2 function $\omega_c \colon X \to \mathbb{R}$ such that for any h in a neighbourhood of zero and $c > 0$ one has

$$
\Phi(G(x_* + h), \lambda_*, c) - \Phi(G(x_*), \lambda_*, c) = \langle \mu_*, DG(x_*)h \rangle
$$
$$
+ \frac{1}{2}\langle \mu_*, D^2 G(x_*)(h, h) \rangle + \frac{1}{2}\omega_c(h) + o(\|h\|^2),
$$

where $o(\|h\|^2)/\|h\|^2 \to 0$ as $h \to 0$, and

1. $\mu_* = \Phi_0(\lambda_*)$ and $\Phi_0(\lambda) = D_y \Phi(y, \lambda, c)$ (see $(A11)$);

2. $\omega_c(h) \to -\sigma(\mu_*, \mathcal{T}(h))$ as $c \to +\infty$ for any $h \in C(x_*, \mu_*)$;

3. if $\limsup_{[h,c] \to [h_*, +\infty]} \omega_c(h)$ is finite for some $h_* \in T_A(x_*)$, then $h_* \in C(x_*, \mu_*)$, and $\limsup_{[h,c] \to [h_*, +\infty]} \omega_c(h) \ge -\sigma(\mu_*, \mathcal{T}(h_*))$.

Note that the function $\Phi(G(\cdot), \lambda_*, c)$ need not be twice differentiable at x_* to admit the second order expansion in x at (x_*, λ_*).

Remark 15. Let us discuss when the assumption that the function $\Phi(G(x), \lambda, c)$ admits the second order expansion in x at (x_*, λ_*) is satisfied for the previously analysed examples. This assumption is satisfied in Example 1, provided Y is finite dimensional, $\sigma(y) = \|y\|^2/2$, (x_*, λ_*) is a KKT-pair, and the restriction of the function $\sigma(\lambda_*, T_K^2(G(x_*), \cdot))$ to its effective domain is u.s.c. (see [72], formulae (3.23) and (3.25)). In this case one has

$$
\omega_c(h) = \min_{z \in \mathscr{C}(x_*, \lambda_*)} \left(c\|DG(x_*)h - z\|^2 - \sigma\big(\lambda_*, T_K^2(G(x_*), z)\big) \right),
$$

where $\mathscr{C}(x_*, \lambda_*) = \{z \in T_K(G(x_*)) \mid \langle \lambda_*, z \rangle = 0\}$. In particular, the assumption holds true in Examples 10 and 12, if (x_*, λ_*) is a KKT-pair.

The assumption is satisfied in Example 2 with

$$
\omega_c(h) = \sum_{i \in I_+(x_*, \lambda_*)} c\phi''(0)\|\nabla g_i(x_*)h\|^2 + \sum_{i \in I_0(x_*, \lambda_*)} c\phi''(0) \max\{0, \langle \nabla g_i(x_*), h \rangle\}^2,
$$

provided $\phi''(0) > 0$, where $I_+(x_*, \lambda_*) = \{i \in I(x_*) \mid (\lambda_*)_i > 0\}$, $I_0(x_*, \lambda_*) = \{i \in I(x_*) \mid (\lambda_*)_i = 0\}$ and $I(x_*) = \{i \in I \mid g_i(x_*) = 0\}$. The assumption is satisfied in Example 3 iff the strict complementarity (s.c.) condition holds true, and it is satisfied in Example 4 iff $\phi''(t) > 0$ for all $t \in \mathbb{R}$. This assumptions

is valid in Examples 5–7 iff $\phi'(0) \neq 0$, $\phi''(0) > 0$ and s.c. condition holds true. The assumptions is valid in Example 8 iff $\phi'(b) \neq 0$, $\phi''(b) > -\phi'(b)^2$ and s.c. condition holds true, and it is always satisfied in Example 9.

The assumption is valid in Example 11, provided the constraint nondegeneracy and s.c. conditions are satisfied ([93], Cor. 3.1 and Prp. 3.1). Similarly, this assumption is satisfied in Examples 13 and 14, if s.c. condition holds true ([74], Thm. 5.1, and [55], Prp. 4.2).

Finally, one can check that the function $\Phi(x, \lambda, c)$ from Example 16 never admits the second order expansion in x due to the fact that in this example $\omega_c(h) \to 0$ as $c \to +\infty$ for any $h \in C(x_*, \lambda_*)$, while the sigma term $\sigma(\lambda_*, \mathcal{T}(h))$ for semi-infinite programming problems is not identically equal to zero in the general case (see [4], Section 5.4.3). Apparently, the same conclusion can be drawn for Example 15.

Let us show that a local saddle point of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$ must be a KKT-pair of the problem $(\mathcal{P})$ that under some additional assumptions satisfies the second order necessary optimality condition.

Proposition 9. *Let (x_*, λ_*) be a local saddle point of $\mathscr{L}(x, \lambda, c)$, and the functions f_k, $k \in M$, and G be differentiable at x_*. Suppose also that assumptions (A1)–(A3), (A8), (A9) and (A11) are satisfied. Then (x_*, μ_*) is a KKT-pair of the problem $(\mathcal{P})$, where $\mu_* = \Phi_0(\lambda_*)$. If, in addition, $x_* \in \operatorname{int} A$, the functions f_k, $k \in M$, and G are twice continuously differentiable at x_*, the function $\Phi(G(x), \lambda, c)$ admits the second order expansion in x at (x_*, λ_*), and either $m = 1$ or $\Phi(G(\cdot), \lambda_*, c)$ is twice continuously differentiable at x_*, then the KTT-pair (x_*, μ_*) satisfies the second order necessary optimality condition.*

Proof. Proposition 1 implies that x_* is a feasible point of the problem $(\mathcal{P})$, i.e. $G(x_*) \in K$, while Propositions 2 and 3, and (A11) imply that $\mu_* \in K^*$ and $\langle \mu_*, G(x_*) \rangle = 0$.

Taking into account (A11) one gets that the function $\mathscr{L}(\cdot, \lambda_*, c)$ is Hadamard directionally differentiable at x_*, and $[\mathscr{L}(\cdot, \lambda_*, c)]'(x_*, h) = [L(\cdot, \mu_*)]'(x_*, h)$ for all $h \in X$ and $c > 0$. Note that by the definition of local saddle point x_* is a local minimizer of $\mathscr{L}(\cdot, \lambda_*, c)$ on the set A for any $c > c_{loc}^*(x_*, \lambda_*)$. Therefore

$$[\mathscr{L}(\cdot, \lambda_*, c)]'(x_*, h) = [L(\cdot, \mu_*)]'(x_*, h) \geq 0 \quad \forall h \in T_A(x_*) \tag{26}$$

for any $c > c_{loc}^*(x_*, \lambda_*)$. Thus, (x_*, μ_*) is a KKT-pair of the problem $(\mathcal{P})$.

Let us now turn to second order necessary optimality conditions. Suppose, at first, that $m = 1$. In this case, the function $L(\cdot, \mu_*)$ is twice differentiable at x_*, and $D_x L(x_*, \mu_*) = 0$ due to the fact that $x_* \in \operatorname{int} A$. Fix arbitrary $c > c_{loc}^*(x_*, \lambda_*)$ and $h \in C(x_*, \mu_*)$, and choose a sequence $\{\gamma_n\} \subset (0, 1)$ such that $\gamma_n \to 0$ as $n \to \infty$. Since $x_* \in \operatorname{int} A$, one can suppose that $x_n := x_* + \gamma_n h \in A$ for all $n \in \mathbb{N}$.

As it was noted above, x_* is a local minimizer of $\mathscr{L}(\cdot, \lambda_*, c)$ on the set A. Therefore $\mathscr{L}(x_n, \lambda_*, c) \geq \mathscr{L}(x_*, \lambda_*, c)$ for any sufficiently large n. Hence taking into account (26), the equality $D_x L(x_*, \mu_*) = 0$, and the fact that $\Phi(G(x), \lambda, c)$ admits the second order expansion in x at (x_*, λ_*) one obtains that

$$0 \leq \mathscr{L}(x_n, \lambda_*, c) - \mathscr{L}(x_*, \lambda_*, c)$$
$$= \frac{\gamma_n^2}{2} \langle h, D_{xx}^2 L(x_*, \mu_*) h \rangle + \frac{\gamma_n^2}{2} \omega_c(h) + o(\|x_n - x_*\|^2)$$

for any n large enough. Dividing this inequality by γ_n^2, passing to the limit as $n \to +\infty$, and then passing to the limit as $c \to \infty$ with the use of Def. 3 one obtains that the KKT-pair (x_*, μ_*) satisfies the second order necessary optimality condition.

Suppose, now, that $\Phi(G(\cdot), \lambda_*, c)$ is twice continuously differentiable at x_*. Fix an arbitrary $c > c_{loc}^*(x_*, \lambda_*)$. As it was pointed out above, x_* is a point of local minimum of $\mathscr{L}(\cdot, \lambda_*, c)$. Consequently, the point $(x_*, 0) \in \mathbb{R}^{d+1}$ is a locally optimal solution of the problem

$$\min_{(x,z)} z \quad \text{subject to} \quad g_k(x, z) \le 0, \quad k \in M, \tag{27}$$

where $g_k(x, z) = f_k(x) + \Phi(G(x), \lambda_*, c) - z$. The Lagrangian function for this problem has the form $\mathcal{L}(x, z, \alpha) = z + \sum_{k=1}^m \alpha_k g_k(x, z)$. Observe that the set of Lagrange multipliers of problem (27) at the point $(x_*, 0)$ coincides with $\alpha(x_*, \mu_*)$.

From the fact that $\Phi(G(\cdot), \lambda_*, c)$ is twice continuously differentiable at x_* it follows that the functions $g_k(x, z)$ are twice continuously differentiable at the point $(x_*, 0)$. Furthermore, the Mangasarian-Fromovitz constraint qualification obviously holds at this point. Therefore, applying the second order necessary optimality conditions (see, e.g., [4], Theorem 3.45) one can easily verify that for any $h \in \mathbb{R}^d$ such that $\langle D_x g_k(x_*, 0), h \rangle \le 0$ for all $k \in M(x_*)$ one has

$$\sup_{\alpha \in \alpha(x_*, \mu_*)} \langle h, D_{xx} \mathcal{L}(x_*, 0, \alpha) h \rangle \ge 0.$$

Hence applying $(A11)$ and Def. 3 one gets that

$$\sup_{\alpha \in \alpha(x_*, \mu_*)} \langle h, D_{xx} L_0(x_*, \mu_*, \alpha) h \rangle + \omega_c(h) \ge 0 \quad \forall c > c_{loc}^*(x_*, \lambda_*)$$

for any $h \in \mathbb{R}^d$ such that $[L(\cdot, \mu_*)]'(x_*, h) = \max_{k \in M(x_*)} \langle D_x g_k(x_*, 0), h \rangle \le 0$. Passing to the limit as $c \to \infty$, and taking into account Def. 3 one obtains that the KKT-pair (x_*, μ_*) satisfies optimality condition (25). $\qquad \square$

Let us prove that under some additional assumptions any KKT-pair satisfying the second order *sufficient* optimality condition is a local saddle point of the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$.

Suppose that the functions f_k, $k \in M$ and G are twice Fréchet differentiable at a point $x_* \in A$. One says that a KKT-pair (x_*, λ_*) satisfies *the second order sufficient optimality condition* (cf. [4], Theorem 3.86) if

$$\sup_{\alpha \in \alpha(x_*, \lambda_*)} \langle h, D_{xx}^2 L_0(x_*, \lambda_*, \alpha) h \rangle - \sigma(\lambda_*, \mathcal{T}(h)) > 0$$

$$\forall h \in C(x_*, \lambda_*) \setminus \{0\} : [L(\cdot, \lambda_*)]'(x_*, h) = 0. \tag{28}$$

The following result holds true.

Theorem 2. *Let x_* be a locally optimal solution of the problem $(\mathcal{P})$, the functions f_k, $k \in M$, and G be twice Fréchet differentiable at x_*, and (x_*, μ_*) be a KKT-pair of the problem $(\mathcal{P})$ satisfying the second order sufficient optimality condition. Suppose also that assumptions $(A2)$, $(A4)$, $(A10)$ and $(A11)$ hold true, and the function $\Phi(G(x), \lambda, c)$ admits the second order expansion in x at (x_*, λ_*) for some $\lambda_* \in \Phi_0^{-1}(\mu_*)$. Then (x_*, λ_*) is a local saddle point of $\mathscr{L}(x, \lambda, c)$.*

Proof. Taking into account the fact that (x_*, μ_*) is a KKT-pair of the problem $(\mathcal{P})$ and utilizing assumption $(A2)$, $(A10)$ and $(A11)$ one can easily verify that for all $c > 0$ one has $\sup_{\lambda \in \Lambda} \mathscr{L}(x_*, \lambda, c) \leq \mathscr{L}(x_*, \lambda_*, c)$.

Applying $(A11)$ and Def. 3 one obtains that for any $c > 0$ there exists a neighbourhood U_c of x_* such that for all $x \in U_c$ one has

$$
\Big| \mathscr{L}(x, \lambda_*, c) - \mathscr{L}(x_*, \lambda_*, c)
$$
$$
- \max_{k \in M(x_*)} \Big(\langle \nabla f_k(x_*), x - x_* \rangle + \frac{1}{2} \langle x - x_*, \nabla^2 f_k(x_*)(x - x_*) \rangle \Big)
$$
$$
- \Big\langle \mu_*, DG(x_*)(x - x_*) + \frac{1}{2} D^2 G(x_*)(x - x_*, x - x_*) \Big\rangle
$$
$$
- \frac{1}{2} \omega_c(x - x_*) \Big| < \frac{1}{2c} \|x - x_*\|^2. \quad (29)
$$

Arguing by reductio ad absurdum, suppose that (x_*, λ_*) is not a local saddle point of $\mathscr{L}(x, \lambda, c)$. Then for any $n \in \mathbb{N}$ there exists $x_n \in A \cap U_n$ such that $\mathscr{L}(x_n, \lambda_*, n) < \mathscr{L}(x_*, \lambda_*, n)$. Then taking into account (29) one obtains that

$$
0 > \mathscr{L}(x_n, \lambda_*, n) - \mathscr{L}(x_*, \lambda_*, n) \geq \max_{k \in M(x_*)} \Big(\langle \nabla f_k(x_*), z_n \rangle + \frac{1}{2} \langle z_n, \nabla^2 f_k(x_*) z_n \rangle \Big)
$$
$$
+ \Big\langle \mu_*, DG(x_*) z_n + \frac{1}{2} D^2 G(x_*)(z_n, z_n) \Big\rangle + \frac{1}{2} \omega_n(z_n) - \frac{1}{2n} \|z_n\|^2 \quad (30)
$$

for any $n \in \mathbb{N}$, where $z_n = x_n - x_*$.

Recall that if $\alpha \in \alpha(x_*, \mu_*)$, then $\alpha \in \mathbb{R}_+^m$, $\alpha_k = 0$ for all $k \notin M(x_*)$, and $\sum_{k=1}^m \alpha_k = 1$. Therefore with the use of (30) one obtains that for any $\alpha \in \alpha(x_*, \mu_*)$ and $n \in \mathbb{N}$ one has

$$
0 > \mathscr{L}(x_n, \lambda_*, n) - \mathscr{L}(x_*, \lambda_*, n) \geq \big\langle D_x L_0(x_*, \mu_*, \alpha), z_n \big\rangle
$$
$$
+ \frac{1}{2} \big\langle z_n, D_{xx}^2 L_0(x_*, \mu_*, \alpha) z_n \big\rangle + \frac{1}{2} \omega_n(z_n) - \frac{1}{2n} \|z_n\|^2. \quad (31)
$$

Denote $h_n = z_n / \|z_n\|$. Since X is a finite dimensional space, without loss of generality one can suppose that the sequence h_n converges to a point h_* with $\|h_*\| = 1$. Furthermore, since $x_n \in A$ for all $n \in \mathbb{N}$ and A is convex, $h_* \in T_A(x_*)$ and $z_n \in T_A(x_*)$ for all $n \in \mathbb{N}$.

Let us check that $[L(\cdot, \mu_*)]'(x_*, h_*) = 0$. Indeed, from the facts that (x_*, μ_*) is a KKT-pair and $h_* \in T_A(x_*)$ it follows that $[L(\cdot, \mu_*)]'(x_*, h_*) \geq 0$. Suppose that $[L(\cdot, \mu_*)]'(x_*, h_*) > 0$. Then by virtue of $(A11)$ for any $c > 0$ one has

$$
\lim_{n \to \infty} \frac{\mathscr{L}(x_* + \gamma_n h_n, \lambda_*, c) - \mathscr{L}(x_*, \lambda_*, c)}{\gamma_n} = [L(\cdot, \mu_*)]'(x_*, h_*) > 0,
$$

where $\gamma_n = \|x_n - x_*\|$. Note that $x_* + \gamma_n h_n = x_n$. Therefore, in particular, there exists $n_0 \in \mathbb{N}$ such that $\mathscr{L}(x_n, \lambda_*, 1) > \mathscr{L}(x_*, \lambda_*, 1)$ for all $n \geq n_0$. Observe that $\mathscr{L}(x_*, \lambda_*, c) = f(x_*)$ for all $c > 0$ due to $(A10)$ and the fact that (x_*, μ_*) is a KKT-pair, while $\mathscr{L}(x_n, \lambda_*, c) \geq \mathscr{L}(x_n, \lambda_*, 1)$ for all $c \geq 1$ by $(A4)$. Consequently, $\mathscr{L}(x_n, \lambda_*, n) > \mathscr{L}(x_*, \lambda_*, n)$ for any $n \geq n_0$, which contradicts the definition of x_n. Thus, $[L(\cdot, \mu_*)]'(x_*, h_*) = 0$.

From (31) and the fact that (x_*, μ_*) is a KKT-pair (see (24)) it follows that

$$0 > \left\langle z_n, D^2_{xx} L_0(x_*, \mu_*, \alpha) z_n \right\rangle + \omega_n(z_n) - \frac{1}{n} \|z_n\|^2.$$

Dividing this inequality by $\|z_n\|^2$, passing to the limit superior as $n \to \infty$ with the use of Def. 3, and taking the supremum over all $\alpha \in \alpha(x_*, \mu_*)$ one obtains that

$$0 \geq \sup_{\alpha \in \alpha(x_*, \mu_*)} \left\langle h_*, D^2_{xx} L_0(x_*, \mu_*, \alpha) h_* \right\rangle - \sigma(\mu_*, \mathcal{T}(h_*)), \quad h_* \in C(x_*, \mu_*),$$

which contradicts the fact that the KKT-pair (x_*, μ_*) satisfies the second order sufficient optimality condition due to the fact that $[L(\cdot, \mu_*)]'(x_*, h_*) = 0$. $\qquad \square$

Remark 16. (i) Let us note that $\Phi_0(\lambda) \equiv \lambda$ for most particular augmented Lagrangian functions appearing in applications. Therefore, usually, there is a direct connection between local saddle points and KKT-pairs corresponding to locally optimal solutions, i.e. if (x_*, λ_*) is a local saddle point, then x_* is a locally optimal solution, and (x_*, λ_*) is a KKT-pair satisfying the second order necessary optimality condition (Proposition 9), and, conversely, if (x_*, λ_*) is a KKT-pair satisfying the second order sufficient optimality condition, then (x_*, λ_*) is a local saddle point (Theorem 2).
(ii) As it was noted above, existing augmented Lagrangian functions for semi-infinite programming problems do not admit the second order expansion in x in the sense of Def. 3 due to the absence of the sigma term $\sigma(\lambda_*, \mathcal{T}(h_*))$ in their second order expansions. Therefore, the second order sufficient optimality condition (28) cannot be utilized in order to prove the existence of a local saddle point in the case of semi-infinite programming problems. More important, one can easily provide an example of a semi-infinite programming problem such that there exists a KKT-pair of this problem satisfying (28), but which is not a local saddle point of augmented Lagrangian functions from Example 16 and Example 15 with $\sigma(y)$ defined as in [70, 6] (such KKT-pair must not satisfy optimality conditions without the sigma term).
(iii) Theorem 2 unifies and sharpens many known results on existence of local saddle points of augmented Lagrangian for mathematical programming problems (see, e.g., [95], Thm. 3.3; [98], Thm. 2.8; [89], Thm. 2; [49], Thm. 2; [50], Thm. 3.2; [79], Thms. 2.1, 2.3 and 2.4), nonlinear second order cone programming problems ([97], Thm. 2.3) and nonlinear semidefinite programming problems ([55], Thm. 4.2). Furthermore, Theorem 2 also extends the aforementioned results to the case of minimax cone constrained optimization problem. To the best of author's knowledge, Theorem 2 contains first simple sufficient conditions for the existence of local saddle points in the case of cone constrained *minimax* problems. Let us also note that Theorem 2 provides a correct proof of Theorem 4 in [28].

In the theorem above, we demonstrated that the existence of a local saddle point of the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$ can be easily proved via second order sufficient optimality conditions. Let us show that one can utilize *first* order sufficient optimality conditions for constrained minimax problems [58, 10, 9, 11, 13, 12] in order to obtain a similar result.

Suppose that the functions f_k, $k \in M$ and G are differentiable at a point $x_* \in A$, and let (x_*, λ_*) be a KKT-pair of the problem $(\mathcal{P})$. Then $[L(\cdot, \lambda_*)]'(x_*, h) \geq 0$

for any $h \in T_A(x_*)$. The natural "no gap" first order *sufficient* optimality condition for the problem $(\mathcal{P})$ has the form

$$[L(\cdot, \lambda_*)]'(x_*, h) > 0 \quad \forall h \in C(x_*, \lambda_*) \setminus \{0\} \tag{32}$$

(cf. [4], Section 3.1.4). The main drawbacks of this sufficient optimality condition consist in the facts that this condition is difficult to verify, and it rarely holds true for smooth problems. However, sufficient optimality condition (32) often holds for constrained minimax problems, and, furthermore, it can be re-formulated in a more convenient form of the so-called *alternance conditions* that are independent of the Lagrange multiplier λ_*.

Let us introduce *alternance optimality conditions* for the problem $(\mathcal{P})$. Let $Z \subset \mathbb{R}^d$ be a set consisting of d linearly independent vectors. Denote by $N_A(x_*) = \{z \in \mathbb{R}^d \mid \langle z, h \rangle \le 0 \, \forall h \in T_A(x_*)\}$ the normal cone to A at x_*. For any linear subspace $Y_0 \subset Y$ denote by $Y_0^\perp = \{y^* \in Y^* \mid \langle y^*, y \rangle = 0 \, \forall y \in Y_0\}$ the annihilator of Y_0. For the sake of correctness, for any linear operator $U \colon \mathbb{R}^d \to Y$ denote by $[U]^*$ the composition of the natural isomorphism between $(\mathbb{R}^d)^*$ and $\mathbb{R}^d$ and the adjoint operator $U^* \colon Y^* \to (\mathbb{R}^d)^*$.

One says that a *p-point alternance* exists at x_* with $p \in \{1, \ldots, d+1\}$, if there exist $k_0 \in \{1, \ldots, p\}$, $i_0 \in \{k_0 + 1, \ldots, p\}$, vectors

$$V_1, \ldots, V_{k_0} \in \left\{ \nabla f_k(x_*) \,\Big|\, k \in M(x_*) \right\},$$

$$V_{k_0+1}, \ldots, V_{i_0} \in \big[DG(x_*)\big]^* \left(K^* \cap \mathrm{span}(G(x_*))^\perp \right), \quad V_{i_0+1}, \ldots, V_p \in N_A(x_*),$$

and vectors $V_{p+1}, \ldots, V_{d+1} \in Z$ such that the dth-order determinants Δ_s of the matrices composed of the columns $V_1, \ldots, V_{s-1}, V_{s+1}, \ldots V_{d+1}$ satisfy the following conditions

$$\Delta_s \ne 0, \quad s \in \{1, \ldots, p\}, \quad \mathrm{sign}\,\Delta_s = -\,\mathrm{sign}\,\Delta_{s+1}, \quad s \in \{1, \ldots, p-1\},$$
$$\Delta_s = 0, \quad s \in \{p+1, \ldots d+1\}.$$

One can verify that a p-point alternance exists at x_* for some $p \in \{1, \ldots, d+1\}$ iff there exists $\lambda_* \in K^*$ such that (x_*, λ_*) is a KKT-pair of the problem $(\mathcal{P})$. Moreover, the existence of a *complete* (i.e. $d+1$-point) alternance is a first order sufficient optimality condition for the problem $(\mathcal{P})$ (see [58, 10, 9, 11, 13, 12] for more details). Note that in the case of complete alternance one has

$$\Delta_s \ne 0 \quad s \in \{1, \ldots, d+1\}, \quad \mathrm{sign}\,\Delta_s = -\,\mathrm{sign}\,\Delta_{s+1} \quad s \in \{1, \ldots, d\},$$

i.e. the determinants Δ_s, $s \in \{1, \ldots, d+1\}$ are not equal to zero and have *alternating* signs, which explains the term *alternance optimality conditions*. Finally, it should be mentioned that usually alternance optimality conditions can only be applied in the case when the cardinality of $M(x_*)$ is greater than 1, i.e. when the objective function $f(x)$ is nonsmooth at x_*.

Remark 17. Let us point out that there is a close connection between standard and alternance optimality conditions for minimax optimization problems. In particular, if one considers the unconstrained problem (i.e. $G(x) \equiv 0$, $K = \{0\}$ and $A = \mathbb{R}^d$), then the standard first order necessary optimality condition $0 \in \partial f(x_*) = \{\nabla f_k(x_*) \mid k \in M(x_*)\}$ is equivalent to the existence of a p-point alternance at x_* for some $p \in \{1, \ldots, d+1\}$. Similarly, the natural first order

sufficient optimality condition $0 \in \operatorname{int} \partial f(x_*)$, that is equivalent to the first order growth condition at x_* (i.e. there exists $\gamma > 0$ such that $f(x) \geq f(x_*) + \gamma \|x - x_*\|$ for any x in a neighbourhood of x_*), is also equivalent to the existence of a complete alternance. Let us also note that the existence of complete alternance is a natural assumption for many particular minimax problems. For more details on alternance optimality conditions see [58, 10, 9, 11, 13, 12].

Our aim is to prove that the existence of a complete alternance at a locally optimal solution of the problem $(\mathcal{P})$ guarantees the existence of a local saddle point of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$.

Theorem 3. *Let x_* be a locally optimal solution of the problem $(\mathcal{P})$, the functions f_k, $k \in M$ and G, be twice Fréchet differentiable at x_*, and let a complete alternance exists at x_*. Suppose also that assumptions $(A2)$, $(A4)$, $(A10)$ and $(A11)$ are satisfied, and the function $\Phi(G(x), \lambda, c)$ admits the second order expansion in x at (x_*, λ) for any $\lambda \in K^*$ such that $\langle \lambda, G(x_*) \rangle = 0$. Then there exists $\lambda_* \in K^*$ such that (x_*, λ_*) is a local saddle point of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$. Furthermore, under the assumptions of the theorem for any KKT-pair (x_*, μ_*) of the problem $(\mathcal{P})$, and for all $\lambda_* \in \Phi_0^{-1}(\mu_*)$ the pair (x_*, λ_*) is a local saddle point of $\mathscr{L}(x, \lambda, c)$.*

Proof. Let us verify, at first, that there exists $\mu_* \in K^*$ such that (x_*, μ_*) is a KKT-pair of the problem $(\mathcal{P})$. Indeed, let $k_0 \in \{1, \ldots, d+1\}$, $i_0 \in \{k_0 + 1, \ldots, d+1\}$ and vectors $V_1, \ldots, V_{d+1} \in \mathbb{R}^d$ be from the definition of complete alternance. Applying Cramer's rule to the system $\sum_{s=2}^{d+1} \beta_s V_s = -V_1$ one obtains that there exist unique $\beta_s > 0$, $s \in \{2, \ldots, d+1\}$ such that

$$V_1 + \sum_{s=2}^{d+1} \beta_s V_s = 0, \quad \beta_s = (-1)^{s-1} \frac{\Delta_s}{\Delta_1} > 0 \quad s \in \{2, \ldots, d+1\}. \tag{33}$$

Denote $\widehat{\beta} = 1 + \beta_2 + \ldots + \beta_{k_0}$, $\gamma_1 = 1/\widehat{\beta}$, and $\gamma_s = \beta_s / \widehat{\beta} > 0$, $s \in \{2, \ldots, d+1\}$. Define $\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{R}_+^m$ as follows

$$\alpha_k = \begin{cases} \gamma_s, & \text{if } k \in M(x_*) \text{ and } \exists s \in \{1, \ldots, k_0\} \colon V_s = \nabla f_k(x_*), \\ 0, & \text{otherwise}. \end{cases}$$

Then one can easily see that $\alpha_1 + \ldots + \alpha_m = 1$, $\alpha_k = 0$ if $k \notin M(x_*)$, and

$$\sum_{s=1}^{k_0} \gamma_i V_i = \sum_{k=1}^{m} \alpha_k \nabla f_k(x_*).$$

Denote $W = \sum_{k_0+1}^{i_0} \gamma_s V_s$. From the fact that $K^* \cap \operatorname{span}(G(x_*))^{\perp}$ is a convex cone it follows that $W \in [DG(x_*)]^*(K^* \cap \operatorname{span}(G(x_*))^{\perp})$. Hence $W = [DG(x_*)]^* \mu_*$ for some $\mu_* \in K^*$ such that $\langle \mu_*, G(x_*) \rangle = 0$ (in the case $k_0 = d+1$ one has $W = 0$ and $\mu_* = 0$). Therefore for any $h \in \mathbb{R}^d$ one has

$$\left\langle D_x L_0(x_*, \mu_*, \alpha), h \right\rangle = \left\langle \sum_{s=1}^{i_0} \gamma_i V_i, h \right\rangle = -\gamma_s \sum_{s=i_0+1}^{d+1} \langle V_s, h \rangle.$$

By definition, $V_s \in N_A(x_*)$ for any $s \in \{i_0 + 1, \ldots, d+1\}$. Consequently,

$$\left\langle D_x L_0(x_*, \mu_*, \alpha), h \right\rangle \geq 0 \quad \forall h \in T_A(x_*)$$

32

(note that in the case $k_0 = d+1$ or $i_0 = d+1$ one has $D_x L_0(x_*, \mu_*, \alpha) = 0$, and the above inequality holds trivially). Thus, (x_*, μ_*) is a KKT-pair of the problem $(\mathcal{P})$ and $\alpha \in \alpha(x_*, \mu_*)$.

Let, now, $\mu_* \in K^*$ be such that (x_*, μ_*) is a KKT-pair of the problem $(\mathcal{P})$, and $\lambda_* \in \Phi_0^{-1}(\mu_*)$ be arbitrary. With the use of $(A2)$, $(A10)$ and $(A11)$ one obtains that $\sup_{\lambda \in \Lambda} \mathscr{L}(x_*, \lambda, c) \le \mathscr{L}(x_*, \lambda_*, c)$ for all $c > 0$.

Applying $(A11)$ and Def. 3 one gets that for any $c > 0$ there exists a neighbourhood U_c of x_* such that for any $x \in U_c$ one has

$$
\begin{aligned}
\Big| \mathscr{L}&(x, \lambda_*, c) - \mathscr{L}(x_*, \lambda_*, c) \\
&- \max_{k \in M(x_*)} \Big(\langle \nabla f_k(x_*), x - x_* \rangle + \frac{1}{2} \langle x - x_*, \nabla^2 f_k(x_*)(x - x_*) \rangle \Big) \\
&- \langle \mu_*, DG(x_*)(x - x_*) \rangle - \frac{1}{2} \langle \mu_*, D^2 G(x_*)(x - x_*, x - x_*) \rangle \\
&\qquad\qquad\qquad\qquad - \frac{1}{2} \omega_c(x - x_*) \Big| < \frac{1}{2c} \| x - x_* \|^2. \quad (34)
\end{aligned}
$$

Recall that our aim is to show that (x_*, λ_*) is a local saddle point of $\mathscr{L}(x, \lambda, c)$. Arguing by reductio ad absurdum, suppose that this claim is false. Then for any $n \in \mathbb{N}$ there exists $x_n \in A \cap U_n$ such that $\mathscr{L}(x_n, \lambda_*, n) < \mathscr{L}(x_*, \lambda_*, n)$. Taking into account (34) one obtains that for any $n \in \mathbb{N}$ the following inequality holds true

$$
\begin{aligned}
0 > \max_{k \in M(x_*)} \Big(&\langle \nabla f_k(x_*), z_n \rangle + \frac{1}{2} \langle z_n, \nabla^2 f_k(x_*) z_n \rangle \Big) + \langle \mu_*, DG(x_*) z_n \rangle \\
&+ \frac{1}{2} \langle \mu_*, D^2 G(x_*)(z_n, z_n) \rangle + \frac{1}{2} \omega_n(z_n) - \frac{1}{2n} \| z_n \|^2, \quad (35)
\end{aligned}
$$

where $z_n = x_n - x_*$. For any $n \in \mathbb{N}$ denote $h_n = z_n / \| z_n \|$. Without loss of generality, one can suppose that the sequence $\{ h_n \}$ converges to a vector $h_* \in T_A(x_*)$ such that $\| h_* \| = 1$.

Since A is convex, $z_n \in T_A(x_*)$ for all $n \in \mathbb{N}$. Hence taking into account (35) and the fact that (x_*, μ_*) is a KKT-pair one obtains that

$$
0 > \langle z_n, D_{xx}^2 L_0(x_*, \mu_*, \alpha) z_n \rangle + \omega_n(z_n) - \frac{1}{n} \| z_n \|^2
$$

for any $n \in \mathbb{N}$ and $\alpha \in \alpha(x_*, \mu_*)$. Dividing this inequality by $\| z_n \|^2$ and passing to the limit superior as $n \to \infty$ with the use of Def. 3 one gets that $h_* \in C(x_*, \mu_*)$, which implies that $\langle \mu_*, DG(x_*) h_* \rangle = 0$ due to the definition of the cone $C(x_*, \mu_*)$.

Dividing (35) by $\| z_n \|$, passing to the limit superior as $n \to \infty$ with the use of Def. 3, and taking into account the facts that $\omega_n(z_n) / \| z_n \| = \omega_n(z_n / \sqrt{\| z_n \|})$, $z_n / \sqrt{\| z_n \|} \to 0$ as $n \to \infty$, and $\sigma(\mu_*, \mathcal{T}(0)) = 0$ one obtains

$$
0 \ge \max_{k \in M(x_*)} \langle \nabla f_k(x_*), h_* \rangle. \quad (36)
$$

Here we used the equality $\langle \mu_*, DG(x_*) h_* \rangle = 0$.

Introduce the matrix $V = (V_1, \dots, V_{d+1})$, and define $\beta = (1, \beta_2, \dots, \beta_{d+1})^T$, where the vectors V_s are from the definition of complete alternance, and β_s are

from (33). The first equality in (33) implies that

$$\langle V\beta, h_* \rangle = \langle \beta, V^T h_* \rangle = \langle V_1, h_* \rangle + \sum_{s=2}^{d+1} \beta_s \langle V_s, h_* \rangle = 0. \tag{37}$$

As it was shown above, $h_* \in C(x_*, \mu_*)$. Therefore $h_* \in T_A(x_*)$ and $\langle V_s, h_* \rangle \le 0$ for any $s \in \{i_0 + 1, \ldots, d+1\}$. Furthermore, from the facts that $h_* \in C(x_*, \mu_*)$ and K is convex it follows that $DG(x_*)h_* \in T_K(G(x_*)) = \operatorname{cl}\operatorname{cone}(K - G(x_*))$ and $\langle y^*, DG(x_*)h_* \rangle \le 0$ for any $y^* \in K^* \cap \operatorname{span}(G(x_*))^\perp$, which yields that $\langle V_s, h_* \rangle \le 0$ for any $s \in \{k_0 + 1, \ldots, i_0\}$. Thus, $\langle V_s, h_* \rangle \le 0$ for all $s \in \{k_0 + 1, \ldots, d+1\}$.

The definition of complete alternance implies that the matrix V has full rank. Consequently, $V^T h_* \ne 0$ due to the fact that $h_* \ne 0$. Hence taking into account (37) and the fact that $\beta_s > 0$ for all $s \in \{2, \ldots, d+1\}$ one obtains that there exists $s_0 \in \{1, \ldots, k_0\}$ such that $\langle V_{s_0}, h_* \rangle > 0$. Recall that $V_s \in \{\nabla f_k(x_*) \mid k \in M(x_*)\}$ for any $s \in \{1, \ldots, k_0\}$. Therefore there exists $k \in M(x_*)$ such that $V_{s_0} = \nabla f_k(x_*)$. Hence one has $\max_{k \in M(x_*)} \langle \nabla f_k(x_*), h_* \rangle > 0$, which contradicts (36). Thus, (x_*, λ_*) is a local saddle point of $\mathscr{L}(x, \lambda, c)$. $\square$

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

Recall that Ω_* is the set of globally optimal solutions of the problem $(\mathcal{P})$. By Proposition 9 any local saddle point (x_*, λ_*) of the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$ must be a KKT-pair of the problem $(\mathcal{P})$. Hence taking into account Corollary 2 one obtains that for the existence of a global saddle point of $\mathscr{L}(x, \lambda, c)$ it is *necessary* that there exists $\lambda_* \in K^*$ such that (x_*, λ_*) is a KKT-pair of $(\mathcal{P})$ for any $x_* \in \Omega_*$. Therefore, if there are at least two globally optimal solutions of the problem $(\mathcal{P})$ that have disjoint sets of Lagrange multipliers, then there are no global saddle points of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$.

Under the additional assumption that the first or the second order sufficient optimality conditions hold at every $x_* \in \Omega_*$, one can demonstrate that the existence of $\lambda_* \in K^*$ such that (x_*, λ_*) is a KKT-pair of $(\mathcal{P})$ for any $x_* \in \Omega_*$ is also sufficient for the existence of a global saddle point of $\mathscr{L}(x, \lambda, c)$.

Theorem 4. *Let A be closed, G be continuous on A, and $\mathscr{L}(\cdot, \lambda, c)$ be l.s.c. on A for any $\lambda \in \Lambda$ and $c > 0$. Suppose that assumptions $(A1)$–$(A4)$, $(A6)$, $(A10)$ and $(A11)$ are satisfied. Let also the following assumptions be valid:*

1. *f_k, $k \in M$ and G are twice Fréchet differentiable at every point $x_* \in \Omega_*$;*

2. *there exists $\mu_* \in \Lambda$ such that (x_*, μ_*) is a KKT-pair of the problem $(\mathcal{P})$ for any $x_* \in \Omega_*$;*

3. *for any $x_* \in \Omega_*$ either a complete alternance exists at x_* or the KKT-pair (x_*, μ_*) satisfies the second order sufficient optimality condition;*

4. *the function $\Phi(G(x), \lambda, c)$ admits the second order expansion in x at every point (x_*, λ_*) such that $x_* \in \Omega_*$ and $\lambda_* \in \Phi_0^{-1}(\mu_*)$.*

Then for any $\lambda_* \in \Phi_0^{-1}(\mu_*)$ one has $\lambda_* \in \Lambda(\mathcal{P})$ iff there exists $c_0 > 0$ such that the set $S(\lambda_*, c_0) = \{x \in A \mid \mathscr{L}(x, \lambda_*, c_0) < f_*\}$ is either bounded or empty.

Proof. With the use of Theorems 2 and 3 one obtains that $\lambda_* \in \Lambda_{loc}(\mathcal{P})$. Then applying the localization principle we arrive at the required result. $\qquad\square$

In the case when a complete alternance exists at every globally optimal solution of $(\mathcal{P})$ one can obtain a stronger result.

Theorem 5. *Let A be closed, G be continuous on A, $\mathscr{L}(\cdot, \lambda, c)$ be l.s.c. on A for any $\lambda \in \Lambda$ and $c > 0$. Suppose that assumptions $(A1)$–$(A4)$, $(A6)$ and $(A8)$–$(A11)$ are satisfied. Let also the following assumptions be valid:*

1. *the functions f_k, $k \in M$, and G are twice Fréchet differentiable at every $x_* \in \Omega_*$;*

2. *the function $\Phi(G(x), \lambda, c)$ admits the second order expansion in x at every point $(x_*, \lambda_*) \in \Omega_* \times K^*$ such that $\langle \lambda_*, G(x_*) \rangle = 0$;*

3. *a complete alternance exists at every $x_* \in \Omega_*$.*

Then a global saddle point of $\mathscr{L}(x, \lambda, c)$ exists if and only if there exist $\mu_* \in K^*$, $\lambda_* \in \Phi_0^{-1}(\mu_*)$ and $c_0 > 0$ such that for any $x_* \in \Omega_*$ the pair (x_*, μ_*) is a KKT-pair of $(\mathcal{P})$, and the set $S(\lambda_*, c_0)$ is either bounded or empty. Furthermore, $\Lambda(\mathcal{P})$ consists of all those $\lambda_* \in \Lambda$ which satisfy the above assumptions.

Proof. Let (x_*, λ_*) be a global saddle point of $\mathscr{L}(x, \lambda, c)$, and $\mu_* = \Phi_0(\lambda_*)$. Then for any $z_* \in \Omega_*$ the pair (z_*, μ_*) is a KKT-pair of $(\mathcal{P})$ by Corollary 2 and Proposition 9. Moreover, the set $S(\lambda_*, c_0)$ is empty for any $c_0 \geq c^*(x_*, \lambda_*)$.

Let, now, $\mu_* \in K^*$ and $\lambda_* \in \Phi_0^{-1}(\mu_*)$ be such that for any $x_* \in \Omega_*$ the pair (x_*, μ_*) is a KKT-pair of $(\mathcal{P})$, and there exists $c_0 > 0$ for which the set $S(\lambda_*, c_0)$ is either bounded or empty. Since a complete alternance exists at every $x_* \in \Omega_*$, $\lambda_* \in \Lambda_{loc}(\mathcal{P})$ by Theorem 3. Then applying the localization principle one obtains the desired result. $\qquad\square$

Remark 18. To the best of author's knowledge, Theorems 4 and 5 are the first results on the existence of global saddle points of augmented Lagrangian functions for cone constrained minimax problems.

Let us demonstrate how one can apply the theorems above to semi-infinite and generalized semi-infinite min-max problems.

Example 19. Consider the following semi-infinite programming problem

$$\min f(x) \quad \text{subject to} \quad g(x, t) \leq 0 \quad \forall t \in T, \tag{38}$$

where T is a compact metric space. Suppose that the functions $f(\cdot)$ and $g(\cdot, t)$, $t \in T$, are twice continuously differentiable, and the functions $g(x, t)$, $\nabla_x g(x, t)$ and $\nabla_{xx}^2 g(x, t)$ are continuous (jointly in x and t).

Let x_* be a globally optimal solution of problem (38), and (x_*, λ_*) be a KKT-pair of this problem. Then, without loss of generality (see, e.g., [4], Lemma 5.110), one can suppose that the support of the measure λ_* consists of at most d points, i.e. λ_* has the form $\lambda_* = \sum_{i=1}^{l} \lambda_i \delta(t_i)$ for some $l \leq d$, $\lambda_i \geq 0$ and $t_i \in T$, where $\lambda_i g(x, t_i) = 0$ for all $i \in I := \{1, \ldots, l\}$, and $\delta(t_i)$ denotes the Dirac measure of mass one at the point t_i. Denote $\overline{\lambda} = (\lambda_1, \ldots, \lambda_m)$.

Let, finally, $\Phi(y, \lambda, c)$ be defined as in Example 16. Then

$$\mathscr{L}(x, \lambda_*, c) = f(x) + \frac{1}{c} \sum_{i=1}^{l} \lambda_i \phi(cg(x, t_i)) = \mathcal{L}(x, \overline{\lambda}, c) \quad \forall x \in \mathbb{R}^d, \qquad (39)$$

where $\mathcal{L}(x, \lambda, c)$ is the exponential penalty function (Example 5) for the discretized problem

$$\min f(x) \quad \text{subject to} \quad g_i(x) = g(x, t_i) \leq 0 \quad \forall i \in I. \qquad (40)$$

With the use of (39) one can easily check that (x_*, λ_*) is a global saddle point of $\mathscr{L}(x, \lambda, c)$ iff $(x_*, \overline{\lambda})$ is a global saddle point of $\mathcal{L}(x, \lambda, c)$. Therefore applying Proposition 1 one obtains that for the pair (x_*, λ_*) to be a global saddle point of $\mathscr{L}(x, \lambda, c)$ it is necesssary that x_* is a globally optimal solution of the discretized problem (40).

One can apply Theorem 4 in order to obtain necessary and sufficient conditions for the existence of a global saddle point of the exponential penalty function $\mathcal{L}(x, \lambda, c)$ for the discretized problem (40), which, in turn, can be used as necessary and sufficient conditions for the existence of a global saddle point of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$. However, this approach forces one to use sufficient optimality conditions for the discretized problem (40) that are unnatural for semi-infinite programming problems due to the absence of the sigma term $\sigma(\lambda_*, \mathcal{T}(h))$ (see [4], Section 5.4.3). Let us note that this drawback is common for all existing results on global saddle points of augmented Lagrangian functions for semi-infinite programming problems (cf. [70, 6]).

Remark 19. It should be noted that all existing augmented Lagrangian functions for semi-infinite programming problems are constructed as a straightforward generalization of corresponding augmented Lagrangian functions for mathematical programming problems. This approach leads to unsatisfactory results, since one has to utilize unnatural optimality conditions in order to prove the existence of global or local saddle points of augmented Lagrangian functions for semi-infinite programming problems. Clearly, a completely different approach to the construction of augmented Lagrangian functions for these problems is needed. The search of such an approach is a challenging topic of future research.

Example 20. Consider the following generalized semi-infinite min-max problem

$$\min_{x \in \mathbb{R}^d} \max_{z \in Z(x)} f(x, z), \quad Z(x) = \left\{ z \in A \mid G(x, z) \in K \right\}, \qquad (41)$$

where $A \subset \mathbb{R}^l$ is a nonempty set, while $f \colon \mathbb{R}^{d+l} \to \mathbb{R}$ and $G \colon \mathbb{R}^{d+l} \to Y$ are given functions. Denote $f_0(x) = \max_{z \in Z(x)} f(x, z)$, $Z^*(x) = \{z \in Z(x) \mid f_0(x) = f(x, z)\}$, and introduce the following lower level cone constrained optimization problem

$$\min_{z \in \mathbb{R}^l} (-f_x(z)) \quad \text{subject to} \quad G_x(z) \in K, \quad z \in A, \qquad (42)$$

where $f_x(z) = f(x, z)$ and $G_x(z) = G(x, z)$. Clearly, the set of globally optimal solutions of this problem coincide with $Z^*(x)$.

Being inspired by the ideas of [62], define

$$h(x, \lambda, c) = \sup_{z \in A} \Big(f(x, z) - \Phi(G(x, z), \lambda, c) \Big) \quad \forall x \in \mathbb{R}^d, \ \lambda \in Y^*, \ c > 0,$$

Let $\mathscr{L}_x(z, \lambda, c)$ be the augmented Lagrangian function for problem (42). Then $h(x, \lambda, c) = -\inf_{z \in A} \mathscr{L}_x(z, \lambda, c)$, i.e. $h(x, \lambda, c)$ is the negative of the objective function of the augmented dual problem of problem (42). Therefore applying Proposition 4 one obtains that the following result holds true.

Proposition 10. *Suppose that assumption* $(A2)$ *holds true. Then*

$$h(x, \lambda, c) \geq f_0(x) \quad \forall x \in \mathbb{R}^d, \ \lambda \in \Lambda, \ c > 0.$$

Suppose, additionally, that assumptions $(A1)$–$(A4)$ *hold true and* $Z^*(x) \neq \emptyset$. *Then* $h(x, \lambda, c) = f_0(x)$ *for some* $\lambda \in \Lambda$ *and* $c > 0$ *iff there exists* $z \in Z^*(x)$ *such that* (z, λ) *is a global saddle point of* $\mathscr{L}_x(z, \lambda, c)$, *and* $c \geq c^*(z, \lambda)$.

The proposition above motivates us to consider the following auxiliary minimax problem

$$\min_{(x, \lambda, c)} h(x, \lambda, c) \quad \text{subject to} \quad \lambda \in \Lambda, \quad c > 0. \tag{43}$$

Let x_* be a locally optimal solution of problems (41) such that $f_0(x) > f_0(x_*)$ for any $x \in U \setminus \{x_*\}$, where U is a neighbourhood of x_*. From Proposition 10 it follows that if $h(x_*, \lambda_*, c_*) = f_0(x_*)$ for some $\lambda_* \in \Lambda$ and $c_* > 0$, then (x_*, λ_*, c_*) is a locally optimal solution of problem (43) such that $h(x, \lambda, c) > h(x_*, \lambda_*, c_*)$ for any $x \in U \setminus \{x_*\}$, $\lambda \in \Lambda$ and $c > 0$. With the use of this result one can easily obtain necessary and sufficient conditions for the sets of all locally (and hence globally) optimal solutions of problems (41) and (43) to coincide. Here, we only provide such conditions for the case of globally optimal solutions.

We say that problems (41) and (43) are *equivalent* if their optimal values coincide, and x_* is a globally optimal solution of problem (41) iff (x_*, λ_*, c_*) with some $\lambda_* \in \Lambda$ and $c_* > 0$ is a globally optimal solution of problem (43). Applying Proposition 10 one can easily obtain the following result.

Proposition 11. *Suppose that assumptions* $(A1)$–$(A4)$ *hold true, and* $Z^*(x_*) \neq \emptyset$ *for any globally optimal solution* x_* *of problem* (41). *Then problems* (41) *and* (43) *are equivalent iff for any globally optimal solution* x_* *of problem* (41) *there exists a global saddle point of the augmented Lagrangian* $\mathscr{L}_{x_*}(z, \lambda, c)$.

With the use of the proposition above and Theorems 4 and 5 (or the localization principle) one can obtain simple necessary and sufficient conditions for the equivalence of problems (41) and (43) in terms of sufficient optimality conditions for lower level problem (42). The interested reader can easily formulate these conditions, which unify and significantly generalize some existing results ([62], Thrm. 3.1; [82], Thrm. 4.5).

7 Exact Augmented Lagrangian Functions

Being inspired by the ideas of Di Pillo, Grippo and Lucidi [14, 16, 15, 53, 21, 22, 19, 18] (see also [32, 31, 54, 23, 17, 24, 25, 20, 26]), in this section, we present a general method for constructing exact augmented Lagrangian functions, and obtain simple sufficient (and necessary) conditions for the exactness of these functions. In particular, we demonstrate that one can easily extend the localization principle to the case of exact augmented Lagrangian functions, thus

showing that the study of the exactness of an augmented Lagrangian function can be easily reduced to a local analysis of sufficient optimality conditions.

Introduce the penalized augmented Lagrangian function

$$\mathscr{L}_e(x, \lambda, c) = f(x) + \Phi(G(x), \lambda, c) + \eta(x, \lambda) = \mathscr{L}(x, \lambda, c) + \eta(x, \lambda),$$

where $\eta\colon X \times Y^* \to [0, +\infty]$ is a given non-negative function (the subscript "e" stands for "exact"). The function η must be defined in such a way that it penalizes the violation of the KKT optimality conditions. However, at first, we suppose that η is an arbitrary non-negative function.

Our aim is to show that under some additional assumptions the penalized augmented Lagrangian $\mathscr{L}_e(x, \lambda, c)$ is *exact*, i.e. that all points of global minimum (x_*, λ_*) of $\mathscr{L}_e(x, \lambda, c)$ in (x, λ) on the set $A \times \Lambda$ are exactly KKT-pairs of the problem $(\mathcal{P})$ corresponding to globally optimal solutions of this problem. Note that unlike the case of augmented Lagrangian functions studied in the previous sections, one must *simultaneously* minimize the penalized augmented Lagrangian $\mathscr{L}_e(x, \lambda, c)$ *both* in primal variable x and in dual variable λ in order to recover optimal solutions of the original problem.

We start by studying the behaviour of global minimizers of $\mathscr{L}_e(x, \lambda, c)$ in (x, λ) as the penalty parameter c increases unboundedly. To this end, we need to introduce a stronger version of assumption $(A6)$.

$(A6)_s$ $\forall y \notin K$ $\forall \lambda \in \Lambda$ $\forall c_0 > 0$ $\exists r > 0$ such that

$$\lim_{c \to \infty} \inf \left\{ \Phi(z, \mu, c) - \Phi(z, \mu, c_0) \,\Big|\, \right.$$
$$\left. z \in B(y, r), \ \mu \in B(\lambda, r) \cap \Lambda \colon \Phi(z, \mu, c_0) < +\infty \right\} = +\infty;$$

One can verify that this assumption is satisfied in Example 1, provided σ has a valley at zero, and hence it is valid in Examples 10, 12 and 15. This assumption is also valid in Examples 2–4, 6, 7, 9 and 14 in the general case, and in Examples 11, 13 and 16 in the case when $\varepsilon_0 < +\infty$. Finally, assumption $(A6)_s$ (as well as $(A6)$) is never satisfied in Examples 5 and 8.

The following result extends Lemma 1 to the case of the penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$.

Lemma 2. *Let A be closed, G be continuous on A, $\mathscr{L}_e(\cdot, \cdot, c)$ be l.s.c. on $A \times \Lambda$ for all $c > 0$, and $\mathscr{L}_e(\cdot, \cdot, c_0)$ be bounded from below on $A \times \Lambda$ for some $c_0 > 0$. Suppose also that assumptions $(A2)$, $(A4)$, $(A6)_s$ and $(A12)$ are satisfied, and there exists $(\overline{x}, \overline{\lambda}) \in \Omega_* \times \Lambda$ such that $\eta(\overline{x}, \overline{\lambda}) = 0$. Let, finally, a pair (x_*, λ_*) be a cluster point of a sequence $\{(x_n, \lambda_n)\} \subset A \times \Lambda$ such that*

$$\mathscr{L}_e(x_n, \lambda_n, c_n) \le \inf_{(x, \lambda) \in A \times \Lambda} \mathscr{L}_e(x, \lambda, c_n) + \varepsilon_n \quad \forall n \in \mathbb{N},$$

where $\{c_n\} \subset [c_0, +\infty)$ is an increasing unbounded sequence, and the sequence $\{\varepsilon_n\} \subset (0, +\infty)$ is such that $\varepsilon_n \to 0$ as $n \to \infty$. Then x_* is a globally optimal solution of $(\mathcal{P})$ and $\eta(x_*, \lambda_*) = 0$.

Proof. Replacing, if necessary, the sequences $\{(x_n, \lambda_n)\}$, $\{c_n\}$ and $\{\varepsilon_n\}$ with their subsequences, one can suppose that (x_*, λ_*) is a limit point of the sequence $\{(x_n, \lambda_n)\}$. Furthermore, note that $(x_*, \lambda_*) \in A \times \Lambda$ by virtue of the fact that A and Λ are closed.

Let us verify, at first, that x_* is a feasible point of the problem $(\mathcal{P})$. Indeed, with the use of $(A2)$ one gets that $\mathscr{L}_e(\overline{x}, \overline{\lambda}, c) \le f(\overline{x}) < +\infty$ for all $c > 0$. Therefore for any $n \in \mathbb{N}$ one has

$$\mathscr{L}_e(x_n, \lambda_n, c_n) \le \inf_{(x,\lambda) \in A \times \Lambda} \mathscr{L}_e(x, \lambda, c_n) + \varepsilon_n \le f(\overline{x}) + \varepsilon_n. \tag{44}$$

Hence, in particular, $f(x_n) < +\infty$ and $\Phi(G(x_n), \lambda_n, c_n) < +\infty$ for all $n \in \mathbb{N}$. Furthermore, one has

$$\limsup_{n \to \infty} \mathscr{L}_e(x_n, \lambda_n, c_n) \le f(\overline{x}) < +\infty. \tag{45}$$

Arguing by reductio ad absurdum, suppose that $G(x_*) \notin K$. Then by $(A6)_s$ there exists $r > 0$ such that

$$\lim_{c \to \infty} \inf \Big\{ \Phi(y, y^*, c) - \Phi(y, y^*, c_0) \,\Big|$$
$$y \in B(G(x_*), r), \ y^* \in B(\lambda_*, r) \cap \Lambda \colon \Phi(y, y_*, c_0) < +\infty \Big\} = +\infty. \tag{46}$$

From the facts that G is continuous on A, and (x_*, λ_*) is a limit point of the sequence $\{(x_n, \lambda_n)\}$ it follows that there exists $n_0 \in \mathbb{N}$ such that $G(x_n) \in B(G(x_*), r)$ and $\lambda_n \in B(\lambda_*, r)$ for all $n \ge n_0$. Consequently, for any $n \ge n_0$ one has

$$\mathscr{L}_e(x_n, \lambda_n, c_n) = \mathscr{L}_e(x_n, \lambda_n, c_0) + \Phi(G(x_n), \lambda_n, c_n) - \Phi(G(x_n), \lambda_n, c_0)$$
$$\ge \gamma + \inf \Big\{ \Phi(y, y^*, c) - \Phi(y, y^*, c_0) \,\Big|$$
$$y \in B(G(x_*), r), \ y^* \in B(\lambda_*, r) \cap \Lambda \colon \Phi(y, y_*, c_0) < +\infty \Big\},$$

where $\gamma = \inf\{\mathscr{L}_e(x, \lambda, c_0) \mid (x, \lambda) \in A \times \Lambda\} > -\infty$. Consequently, applying (46) one obtains that $\mathscr{L}_e(x_n, \lambda_n, c_n) \to +\infty$ as $n \to \infty$, which contradicts (45). Thus, $G(x_*) \in K$.

Observe that from $(A4)$ and (44) it follows that for any $c > 0$ there exists $n_0 \in \mathbb{N}$ such that $\mathscr{L}_e(x_n, \lambda_n, c) \le f(\overline{x}) + \varepsilon_n$ for all $n \ge n_0$. Passing to the limit inferior as $n \to \infty$, and taking into account the fact that $\mathscr{L}_e(x, \lambda, c)$ is l.s.c. in (x, λ) one obtains that $\mathscr{L}_e(x_*, \lambda_*, c) \le f(\overline{x})$ for all $c > 0$. Taking into account the facts that $G(x_*) \in K$ and $\lambda_* \in \Lambda$, and passing to the limit as $c \to \infty$ with the use of $(A12)$ one gets that $f(x_*) + \eta(x_*, \lambda_*) \le f(\overline{x})$. Therefore $x_* \in \Omega_*$ and $\eta(x_*, \lambda_*) = 0$ due to the fact that x_* is feasible, $\overline{x} \in \Omega_*$, and η is a non-negative function. $\qquad\square$

By analogy with the theory of exact penalty functions [27], let us introduce the definition of locally and globally exact augmented Lagrangian functions.

Definition 4. Let x_* be a locally optimal solution of $(\mathcal{P})$, and $\lambda_* \in \Lambda$ be such that $\eta(x_*, \lambda_*) = 0$. The penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$ is called (*locally*) *exact* at (x_*, λ_*) (with respect to the function η) if there exist $c_0 > 0$ and a neighbourhood U of (x_*, λ_*) such that

$$\mathscr{L}_e(x, \lambda, c) \ge \mathscr{L}_e(x_*, \lambda_*, c) \quad \forall (x, \lambda) \in U \cap (A \times \Lambda) \quad \forall c \ge c_0.$$

Note that if assumption $(A4)$ holds true, then $\mathscr{L}_e(x, \lambda, c)$ is locally exact at (x_*, λ_*) iff there exists $c_0 > 0$ such that the pair (x_*, λ_*) is a point of local minimum of $\mathscr{L}_e(\cdot, \cdot, c_0)$ on the set $A \times \Lambda$.

Definition 5. The penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$ is called (*globally*) *exact* (with respect to the function η) if there exists $c_0 > 0$ such that for any $c \geq c_0$ the function $\mathscr{L}_e(\cdot, \cdot, c)$ attains a global minimum on $A \times \Lambda$, and $(x_*, \lambda_*) \in \arg\min_{(x, \lambda) \in A \times \Lambda} \mathscr{L}_e(x, \lambda, c)$ if and only if x_* is a globally optimal solution of $(\mathcal{P})$ and $\eta(x_*, \lambda_*) = 0$.

Thus, if the augmented Lagrangian $\mathscr{L}_e(x, \lambda, c)$ is globally exact, then the problem of minimizing $\mathscr{L}_e(x, \lambda, c)$ in (x, λ) over the set $A \times \Lambda$ is equivalent to the original problem $(\mathcal{P})$ for any sufficiently large value of the penalty parameter c, since in this case points of global minimum of $\mathscr{L}_e(x, \lambda, c)$ on $A \times \Lambda$ are exactly those pairs (x_*, λ_*) for which $x_* \in \Omega_*$ and $\eta(x_*, \lambda_*) = 0$. In particular, in the case when $\eta(x, \lambda) = 0$ iff (x, λ) is a KKT-pair, points of global minimum of $\mathscr{L}_e(x, \lambda, c)$ on $A \times \Lambda$ coincide with KKT-pairs of the problem $(\mathcal{P})$ corresponding to globally optimal solutions of this problem.

Our aim is to show that (under some additional assumptions) the augmented Lagrangian $\mathscr{L}_e(x, \lambda, c)$ is exact if and only if it is locally exact at every pair (x_*, λ_*) such that $x_* \in \Omega_*$ and $\eta(x_*, \lambda_*) = 0$. In other words, our aim is to prove the validity of the localization principle for the penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$. As in the case of the localization principle for global saddle points, the localization principle for the penalized augmented Lagrangian $\mathscr{L}_e(x, \lambda, c)$ allows one to study a local behaviour of $\mathscr{L}_e(x, \lambda, c)$ near globally optimal solutions of the problem $(\mathcal{P})$ in order to prove the global exactness of this function.

We need to introduce a stronger version of assumption $(A4)$. We say that a non-decreasing function $h \colon (0, +\infty) \to \mathbb{R} \cup \{+\infty\}$ is strictly increasing at a point $t > 0$ such that $h(t) < +\infty$, if $h(\tau) > h(t)$ for any $\tau > t$.

$(A4)_s$ assumption $(A4)$ holds true, and $\forall y \in Y \ \forall \lambda \in \Lambda \ \forall c > 0$ such that $\Phi(y, \lambda, c) < +\infty$ either the function $\Phi(y, \lambda, \cdot)$ is strictly increasing at c or $\Phi(y, \lambda, c) = 0$ and $y \in K$.

Note that this assumption is satisfied in Example 1, if the infimum in the definition of $\Phi(y, \lambda, c)$ is attained for all $y \in Y$, $\lambda \in \Lambda$ and $c > 0$. Hence, in particular, assumption $(A4)_s$ is satisfied in Examples 10 and 12. This assumption is always valid in Examples 2–4 and 9, and it is valid in Examples 6 and 14, provided ϕ is strictly convex, and ξ is strictly convex on $\mathbb{R}_+$. Finally, observe that assumption $(A4)_s$ (unlike $(A4)$) is never satisfied in Examples 5, 7, 8, 11, 13 and 16.

Before we proceed to the localization principle, let us point out that instead of verifying that all pairs (x_*, λ_*) with $x_* \in \Omega_*$ and $\eta(x_*, \lambda_*) = 0$ are global minimizers of $\mathscr{L}_e(x, \lambda, c)$ on $A \times \Lambda$, it is sufficient to check that at least one such pair is a point of global minimum of $\mathscr{L}_e(x, \lambda, c)$ on $A \times \Lambda$ in order to prove that $\mathscr{L}_e(x, \lambda, c)$ is globally exact.

Lemma 3. *Let assumption $(A4)_s$ be valid, and suppose that*

$$\Phi(G(x_*), \lambda_*, c) = 0 \quad \forall c > 0 \quad \forall (x_*, \lambda_*) \in \Omega_* \times \Lambda \colon \eta(x_*, \lambda_*) = 0. \qquad (47)$$

Then $\mathscr{L}_e(x, \lambda, c)$ is globally exact if and only if there exist $(x_0, \lambda_0) \in \Omega_* \times \Lambda$ and $c_0 > 0$ such that $\eta(x_0, \lambda_0) = 0$ and the pair (x_0, λ_0) is a point of global minimum of $\mathscr{L}_e(\cdot, \cdot, c_0)$ on $A \times \Lambda$.

Proof. The validity of the "only if" part of the proposition follows directly from the definition of global exactness. Therefore, let us prove the "if" part of the proposition.

By the definition of (x_0, λ_0) and assumption (47) one has

$$\min_{(x,\lambda) \in A \times \Lambda} \mathscr{L}_e(x, \lambda, c_0) = \mathscr{L}_e(x_0, \lambda_0, c_0) = f(x_0) = f_*.$$

Note that $(A4)_s$ implies that the function $\mathscr{L}_e(x, \lambda, c)$ is non-decreasing in c, while (47) implies that $\mathscr{L}_e(x_*, \lambda_*, c) = f(x_*) = f_*$ for all $c > 0$ and for any $(x_*, \lambda_*) \in \Omega_* \times \Lambda$ such that $\eta(x_*, \lambda_*) = 0$. Therefore any such pair (x_*, λ_*) is a point of global minimum of $\mathscr{L}_e(\cdot, \cdot, c)$ on $A \times \Lambda$ and $\min_{(x,\lambda) \in A \times \Lambda} \mathscr{L}_e(x, \lambda, c) = f_*$ for all $c \geq c_0$.

Let, now, (x_*, λ_*) be a point of global minimum of $\mathscr{L}_e(\cdot, \cdot, c_1)$ on $A \times \Lambda$ for some $c_1 > c_0$. Since $\mathscr{L}(x, \lambda, c)$ is non-decreasing in c, $\mathscr{L}_e(x_*, \lambda_*, c) = f_*$ for all $c \in [c_0, c_1]$ (recall that $\mathscr{L}(\cdot, \cdot, c) \geq f_*$ for all $c \geq c_0$). Hence with the use of $(A4)_s$ one gets that $G(x_*) \in K$ and $\Phi(G(x_*), \lambda_*, c) = 0$ for all $c \in [c_0, c_1]$. Therefore $\mathscr{L}_e(x_*, \lambda_*, c) = f(x_*) + \eta(x_*, \lambda_*) = f_*$ for any $c \in [c_0, c_1)$, which implies that $x_* \in \Omega_*$ and $\eta(x_*, \lambda_*) = 0$ due to the fact that x_* is feasible and $\eta(\cdot)$ is non-negative.

Thus, for any $c > c_0$ the function $\mathscr{L}_e(\cdot, \cdot, c)$ attains a global minimum on $A \times \Lambda$ (at the point (x_0, λ_0)), and $(x_*, \lambda_*) \in \arg\min_{(x,\lambda) \in A \times \Lambda} \mathscr{L}_e(x, \lambda, c)$ iff $x_* \in \Omega_*$ and $\eta(x_*, \lambda_*) = 0$. In other words, $\mathscr{L}_e(x, \lambda, c)$ is globally exact. $\square$

Now, we can extend the localization principle to the case of the penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$. Let us note that the localization principle holds true only in the finite dimensional case (cf. [27], Examples 3–5 and [28], Example 4). Therefore, hereinafter, we must suppose that the space Y is finite dimensional.

Theorem 6 (Localization principle). *Let Y be finite dimensional, A be closed, G be continuous on A and $\mathscr{L}_e(\cdot, \cdot, c)$ be l.s.c. on $A \times \Lambda$ for all $c > 0$. Suppose also that assumptions $(A2)$, $(A4)_s$, $(A6)_s$ and $(A12)$ are satisfied, and*

$$\Phi(G(x_*), \lambda_*, c) = 0 \quad \forall c > 0 \quad \forall (x_*, \lambda_*) \in \Omega_* \times \Lambda \colon \eta(x_*, \lambda_*) = 0. \qquad (48)$$

Then $\mathscr{L}_e(x, \lambda, c)$ is globally exact (with respect to the function η) if and only if

1. *there exists $(\overline{x}, \overline{\lambda}) \in \Omega_* \times \Lambda$ such that $\eta(\overline{x}, \overline{\lambda}) = 0$;*

2. *$\mathscr{L}_e(x, \lambda, c)$ is locally exact at every $(x_*, \lambda_*) \in \Omega_* \times \Lambda$ such that $\eta(x_*, \lambda_*) = 0$;*

3. *there exists $c_0 > 0$ such that the set*

$$S_e(c_0) := \left\{ (x, \lambda) \in A \times \Lambda \,\middle|\, \mathscr{L}_e(x, \lambda, c_0) < f_* \right\}$$

   is either bounded or empty.

Proof. Let $\mathscr{L}_e(x, \lambda, c)$ be globally exact. Then there exists $c_0 > 0$ such that for all $c \geq c_0$ the function $\mathscr{L}_e(\cdot, \cdot, c)$ attains a global minimum on $A \times \Lambda$ and $(x_*, \lambda_*) \in \arg\min_{(x,\lambda) \in A \times \Lambda} \mathscr{L}_e(x, \lambda, c)$ iff $x_* \in \Omega_*$ and $\eta(x_*, \lambda_*) = 0$. Therefore,

in particular, there exists a pair $(x_*, \lambda_*) \in \Omega_* \times \Lambda$ such that $\eta(x_*, \lambda_*) = 0$ (otherwise, $\mathscr{L}_e(x, \lambda, c)$ would not attain a global minimum), and $\mathscr{L}_e(x, \lambda, c)$ is locally exact at any such pair (x_*, λ_*), since any such (x_*, λ_*) is a global minimizer of $\mathscr{L}_e(\cdot, \cdot, c)$ on $A \times \Lambda$ for all $c \geq c_0$. Furthermore, applying (48) one gets that $\min_{(x, \lambda) \in A \times \Lambda} \mathscr{L}_e(x, \lambda, c) = f_*$ for all $c \geq c_0$, which implies that $S_e(c_0) = \emptyset$.

Let us prove the converse statement. If $S_e(c_0) = \emptyset$, then applying (48) one obtains that a pair $(\overline{x}, \overline{\lambda}) \in \Omega_* \times \Lambda$ such that $\eta(\overline{x}, \overline{\lambda}) = 0$ (that exists by our assumption) is a point of global minimum of $\mathscr{L}_e(\cdot, \cdot, c_0)$ on $A \times \Lambda$. Consequently, $\mathscr{L}_e(x, \lambda, c)$ is globally exact by Lemma 3.

Note that the function $c \to \mathscr{L}_e(x, \lambda, c)$ is non-decreasing by $(A4)_s$. Therefore it remains to consider the case when $S_e(c) \neq \emptyset$ for all $c > 0$. Choose an increasing unbounded sequence $\{c_n\} \subset [c_0, +\infty)$. Assumption $(A4)_s$ implies that $S_e(c_n) \subseteq S_e(c_0)$ for all $n \in \mathbb{N}$. Hence taking into account the facts that $S_e(c_0)$ is bounded, and $\mathscr{L}_e(\cdot, \cdot, c)$ is l.s.c., one obtains that for any $n \in \mathbb{N}$ there exists $(x_n, \lambda_n) \in \arg\min_{(x, \lambda) \in A \times \Lambda} \mathscr{L}_e(x, \lambda, c_n)$, and, moreover, $(x_n, \lambda_n) \in S_e(c_n) \subseteq S_e(c_0)$, which implies that $\{(x_n, \lambda_n)\}$ is a bounded sequence. Consequently, without loss of generality one can suppose that it converges to some point (x_*, λ_*) that belongs to $A \times \Lambda$, since this set is closed.

Applying Lemma 2 one gets that $x_* \in \Omega_*$ and $\eta(x_*, \lambda_*) = 0$, which implies that $\mathscr{L}_e(x, \lambda, c)$ is locally exact at (x_*, λ_*). Therefore there exist $\widehat{c} > 0$ and a neighbourhood U of (x_*, λ_*) such that

$$\mathscr{L}_e(x, \lambda, c) \geq \mathscr{L}_e(x_*, \lambda_*, c) \quad \forall (x, \lambda) \in U \cap (A \times \Lambda) \quad \forall c \geq \widehat{c}.$$

From the facts that $(x_n, \lambda_n) \to (x_*, \lambda_*)$ and $c_n \to +\infty$ as $n \to \infty$ it follows that there exists $n_0 \in \mathbb{N}$ such that $(x_n, \lambda_n) \in U$ and $c_n \geq \widehat{c}$ for all $n \geq n_0$. Hence for any such $n \in \mathbb{N}$ one has $\mathscr{L}_e(x_n, \lambda_n, c_n) \geq \mathscr{L}_e(x_*, \lambda_*, c_n)$. Consequently, taking into account the definition of (x_n, λ_n) one obtains that (x_*, λ_*) is a point of global minimum of $\mathscr{L}_e(x, \lambda, c_n)$ on $A \times \Lambda$. Applying (48) one gets that $\min_{(x, \lambda) \in A \times \Lambda} \mathscr{L}_e(x, \lambda, c_n) = \mathscr{L}_e(x_*, \lambda_*, c_n) = f(x_*) = f_*$, which contradicts the assumption that $S_e(c) \neq \emptyset$ for all $c > 0$. □

Remark 20. (i) Recall that for the existence of a global saddle point of the augmented Lagrangian function $\mathscr{L}(x, \lambda, c)$ it is necessary that there exists $\lambda_* \in K^*$ such that for any globally optimal solution x_* of the problem $(\mathcal{P})$ the pair (x_*, λ_*) is a KKT-pair of this problem. In particular, a global saddle point of $\mathscr{L}(x, \lambda, c)$ *cannot* exist if the problem $(\mathcal{P})$ has two globally optimal solution with disjoint sets of Lagrange multipliers. In contrast, the penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$ *can* be globally exact even if the problem $(\mathcal{P})$ has two (or more) globally optimal solutions with disjoint sets of Lagrange multipliers.

(ii) Note that the theorem above contains some existing results as simple particular cases (see, e.g., [54], Thms. 4.3 and 4.4).

Note that, as in the case of the localization principle for global saddle point of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$, one can reformulate the assumption on the boundedness of the set $S_e(c_0)$ from the theorem above in terms of behaviour of points of global minimum of $\mathscr{L}(\cdot, \cdot, c)$ as $c \to \infty$. Namely, arguing in a similar way to the proof of Proposition 7 one can verify that the following result holds true.

Proposition 12. *Let Y be finite dimensional, A be closed, G be continuous on A, $\mathscr{L}_e(\cdot, \cdot, c)$ be l.s.c. on $A \times \Lambda$ for all $c > 0$, and $\mathscr{L}_e(x, \lambda, c)$ be locally exact at every point $(x_*, \lambda_*) \in \Omega_* \times \Lambda$ such that $\eta(x_*, \lambda_*) = 0$. Suppose also that assumptions $(A2)$, $(A4)_s$, $(A6)_s$ and $(A12)$ are satisfied, there exists $(\overline{x}, \overline{\lambda}) \in \Omega_* \times \Lambda$ such that $\eta(\overline{x}, \overline{\lambda}) = 0$, and (48) holds true. Then for the existence of $c_0 > 0$ such that $S_e(c_0)$ is either bounded or empty it is necessary and sufficient that there exist $\tau > 0$ and $R > 0$ such that for any $c \geq \tau$ there exists $(x(c), \lambda(c)) \in \arg\min_{(x, \lambda) \in A \times \Lambda} \mathscr{L}_e(x, \lambda, c)$ with $\|x(c)\|_X + \|\lambda(c)\|_Y \leq R$.*

Remark 21. Suppose that $\eta(x_*, \lambda_*) = 0$ iff (x_*, λ_*) is a KKT-pair of the problem $(\mathcal{P})$, and there exist $x_* \in \Omega_*$ and $\lambda_* \in K^*$ such that (x_*, λ_*) is a KKT-pair of $(\mathcal{P})$. From the proposition above it follows that, roughly speaking, the augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$ is globally exact if and only if it is locally exact at every KKT-pair of $(\mathcal{P})$ corresponding to a globally optimal solution of this problem, and points of global minimum of $\mathscr{L}_e(x, \lambda, c)$ in (x, λ) do not escape to infinity as $c \to +\infty$.

Sometimes, the penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$ is locally exact at every pair $(x_*, \lambda_*) \in \Omega_* \times \Lambda$ such that $\eta(x_*, \lambda_*) = 0$, but the set $S_e(c)$ is unbounded for all $c > 0$, which implies that $\mathscr{L}_e(x, \lambda, c)$ is not globally exact. In this case it is natural to ask whether the augmented Lagrangian $\mathscr{L}_e(x, \lambda, c)$ possesses an exactness property that is, in a sense, intermediate between local and global exactness. Let us show that that the answer to this question is positive, and in this case the function $\mathscr{L}_e(x, \lambda, c)$ is exact on bounded subsets of $A \times \Lambda$. We need the following definition in order to clarify this statement.

Definition 6. Let $Q \subseteq A \times \Lambda$ be a nonempty set. The penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$ is called *exact on the set Q* (with respect to η) if there exists $c_0 > 0$ such that $\mathscr{L}_e(x, \lambda, c) \geq f_*$ for all $(x, \lambda) \in Q$ and $c \geq c_0$.

Let us demonstrate that the above definition is a natural extension of the definition of global exactness.

Proposition 13. *Let $Q \subseteq A \times \Lambda$ be a nonempty set such that there exists $(\overline{x}, \overline{\lambda}) \in (\Omega_* \times \Lambda) \cap Q$ for which $\eta(\overline{x}, \overline{\lambda}) = 0$. Let also assumption $(A4)_s$ hold true, and*

$$\Phi(G(x_*), \lambda_*, c) = 0 \quad \forall c > 0 \quad \forall (x_*, \lambda_*) \in \Omega_* \times \Lambda \colon \eta(x_*, \lambda_*) = 0. \qquad (49)$$

Then $\mathscr{L}_e(x, \lambda, c)$ is exact on Q if and only if there exists $c_0 > 0$ such that for any $c \geq c_0$ the function $\mathscr{L}_e(\cdot, \cdot, c)$ attains a global minimum on Q, and $(x_*, \lambda_*) \in \arg\min_{(x, \lambda) \in Q} \mathscr{L}_e(x, \lambda, c)$ iff $x_* \in \Omega_*$ and $\eta(x_*, \lambda_*) = 0$.

Proof. Let $\mathscr{L}_e(x, \lambda, c)$ be exact on Q, and $c_0 > 0$ be from Def. 6. Then by (49) one has that for any $c \geq c_0$ the function $\mathscr{L}_e(\cdot, \cdot, c)$ attains the global minimum on Q at $(\overline{x}, \overline{\lambda})$, and $\min_{(x, \lambda) \in Q} \mathscr{L}_e(x, \lambda, c) = f_*$.

Let, now, (x_*, λ_*) be a point of global minimum of $\mathscr{L}_e(x, \lambda, c_1)$ on Q for some $c_1 > c_0$. Then $\mathscr{L}_e(x_*, \lambda_*, c_1) = f_*$. By $(A4)_s$ the function $\mathscr{L}(x_*, \lambda_*, \cdot)$ is non-decreasing, which implies that $\mathscr{L}(x_*, \lambda_*, c) = f_*$ for all $c \in [c_0, c_1]$. Therefore with the use of $(A4)_s$ one obtains that x_* is feasible, and $\Phi(G(x_*), \lambda_*, c_0) = 0$. Hence $\mathscr{L}_e(x_*, \lambda_*, c_0) = f(x_*) + \eta(x_*, \lambda_*) = f_*$, which yields that $x_* \in \Omega_*$ and

43

$\eta(x_*, \lambda_*) = 0$. Thus, for any $c > c_0$ one has $(x_*, \lambda_*) \in \arg\min_{(x,\lambda) \in Q} \mathscr{L}_e(x, \lambda, c)$ iff $x_* \in \Omega_*$ and $\eta(x_*, \lambda_*) = 0$.

It remains to note that the validity of the converse statement follows directly from (49). $\qquad\square$

Now, we can obtain simple necessary and sufficient conditions for the augmented Lagrangian $\mathscr{L}_e(x, \lambda, c)$ to be exact on every bounded subset of $A \times \Lambda$.

Theorem 7 (Localization Principle). *Let Y be finite dimensional, A be closed, G be continuous on A and $\mathscr{L}_e(\cdot, \cdot, c)$ be l.s.c. on $A \times \Lambda$ for all $c > 0$. Suppose also that assumptions $(A4)_s$, $(A6)_s$ and $(A12)$ are satisfied, and*

$$\Phi(G(x_*), \lambda_*, c) = 0 \quad \forall c > 0 \quad \forall (x_*, \lambda_*) \in \Omega_* \times \Lambda \colon \eta(x_*, \lambda_*) = 0. \qquad (50)$$

Then $\mathscr{L}_e(x, \lambda, c)$ is exact on any bounded subset of $A \times \Lambda$ if and only if $\mathscr{L}_e(x, \lambda, c)$ is locally exact at every pair $(x_*, \lambda_*) \in \Omega_* \times \Lambda$ such that $\eta(x_*, \lambda_*) = 0$.

Proof. If $\mathscr{L}_e(x, \lambda, c)$ is exact on any bounded subset of $A \times \Lambda$, then it is exact on the intersection of a bounded neighbourhood of a point $(x_*, \lambda_*) \in \Omega_* \times \Lambda$ such that $\eta(x_*, \lambda_*) = 0$ and the set $A \times \Lambda$, which with the use of Proposition 13 implies that $\mathscr{L}_e(x, \lambda, c)$ is locally exact at every such point.

Let us prove the converse statement. Let $Q \subset A \times \Lambda$ be a bounded set. Replacing, if necessary, the set Q with its closure one can suppose that Q is closed and, thus, compact (note that if $\mathscr{L}_e(x, \lambda, c)$ is exact on the closure of Q, then it is exact on Q). Choose an increasing unbounded sequence $\{c_n\} \subset (0, +\infty)$. Since $\mathscr{L}_e(\cdot, \cdot, c)$ is l.s.c., and Q is compact, for any $n \in \mathbb{N}$ the function $\mathscr{L}_e(\cdot, \cdot, c_n)$ attains a global minimum on Q at a point (x_n, λ_n). Applying the compactness of Q again one can suppose that the sequence $\{(x_n, \lambda_n)\}$ converges to a point $(x_*, \lambda_*) \in Q$.

From $(A4)_s$ it follows that the sequence $\{\mathscr{L}_e(x_n, \lambda_n, c_n)\}$ is non-decreasing. If this sequence is unbounded, then there exists $n \in \mathbb{N}$ such that $\mathscr{L}_e(x_n, \lambda_n, c_n) \geq f_*$, which implies that $\mathscr{L}_e(x, \lambda, c)$ is exact on Q by the definition of (x_n, λ_n). On the other hand, if this sequence is bounded, then arguing in the same way as in the proof of Lemma 2 one can check that x_* is a feasible point of $(\mathcal{P})$.

If $\eta(x_*, \lambda_*) > 0$, then applying $(A12)$ one gets that there exists $c_0 > 0$ such that $\mathscr{L}_e(x_*, \lambda_*, c) \geq f(x_*) + \eta(x_*, \lambda_*)/2$ for any $c \geq c_0$. Hence taking into account $(A4)_s$ and the fact that $\mathscr{L}_e(\cdot, \cdot, c)$ is l.s.c., one obtains that $\mathscr{L}_e(x_n, \lambda_n, c_n) \geq f(x_*) + \eta(x_*, \lambda_*)/4 > f_*$ for any sufficiently large $n \in \mathbb{N}$, which implies that $\mathscr{L}_e(x, \lambda, c)$ is exact on Q by the definition of (x_n, λ_n).

If $\eta(x_*, \lambda_*) = 0$, but $f(x_*) > f_*$, then applying $(A12)$ and taking into account the fact that $\mathscr{L}_e(\cdot, \cdot, c)$ is l.s.c. one can check that $\mathscr{L}_e(x_n, \lambda_n, c_n) \geq f(x_*) - \varepsilon > f_*$, provided $\varepsilon > 0$ is small enough, and $n \in \mathbb{N}$ is large enough. Thus, $\mathscr{L}_e(x, \lambda, c)$ is exact on Q in this case as well.

Finally, if $\eta(x_*, \lambda_*) = 0$ and $f(x_*) = f_*$, i.e. $x_* \in \Omega_*$, then applying the fact that $\mathscr{L}_e(x, \lambda, c)$ is locally exact at (x_*, λ_*) one can easily verify that this function is exact on Q. $\qquad\square$

Remark 22. Note that the theorem above along with Proposition 13 contain some existing results ([14], Thms. 4–7; [32], Thms. 4.1 and 6.1–6.6; [31], Thms. 7–12) as simple particular cases.

8 Applications of the Localization Principle for Exact Augmented Lagrangian Functions

In this section, we demonstrate that one can easily prove the local exactness of the penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$ with the use of sufficient optimality conditions and/or a proper constraint qualification. This result along with the localization principle allows one to reduce the study of global exactness of the penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$ to the local analysis of sufficient optimality conditions and/or constraint qualifications. In the end of the section, we provide several particular examples of globally exact augmented Lagrangian functions.

8.1 Local Exactness via Sufficient Optimality Conditions

For the sake of simplicity, suppose that $X = \mathbb{R}^d$, A is a convex set, and f is differentiable. In order to prove the local exactness of $\mathscr{L}_e(x, \lambda, c)$ via second order sufficient optimality conditions one needs to utilize a suitable second order expansion of the function $(x, \lambda) \to \mathscr{L}_e(x, \lambda, c)$ in a neighbourhood of a given KKT-pair. The following definition describes what we mean by "a suitable expansion" (cf. Def. 3).

Definition 7. Let assumption $(A11)$ be satisfied, and G be twice Fréchet differentiable at a feasible point $x_* \in A$. Let also $\lambda_* \in K^*$ be such that $\langle \lambda_*, G(x_*) \rangle = 0$. One says that the function $\Phi(G(x), \lambda, c)$ admits *the second order expansion* in (x, λ) at (x_*, λ_*), if for any $c > 0$ there exist a function $\varphi_c \colon X \times Y^* \to \mathbb{R}$ such that for any $h \in X$, $\nu \in Y^*$ and $\alpha > 0$ one has $\varphi_c(\alpha h, \alpha \nu) = \alpha^2 \varphi_c(h, \nu)$, and

$$\Phi(G(x_* + h), \lambda_* + \nu, c) - \Phi(G(x_*), \lambda_*, c) = \langle \mu_*, DG(x_*)h \rangle$$
$$+ \frac{1}{2} \langle \mu_*, D^2 G(x_*)(h, h) \rangle + \frac{1}{2} \varphi_c(h, \nu) + o(\|h\|^2 + \|\nu\|^2),$$

where $o(\|h\|^2 + \|\nu\|^2)/(\|h\|^2 + \|\nu\|^2) \to 0$ as $(h, \nu) \to (0, 0)$, $\mu_* = \Phi_0(\lambda_*)$, and if

$$\limsup_{[h, \nu, c] \to [h_*, \nu_*, +\infty]} \varphi_c(h, \nu)$$

is finite for some $h_* \in T_A(x_*)$ and $\nu_* \in Y^*$, then $h_* \in C(x_*, \mu_*)$, and the limit is greater than or equal to $-\sigma(\mu_*, \mathcal{T}(h_*))$.

Remark 23. Let us point out when the assumption that the function $\Phi(G(x), \lambda, c)$ admits the second order expansion in (x, λ) at a point (x_*, λ_*) is valid (cf. Remark 15).

This assumption is satisfied in Example 1, provided Y is finite dimensional, $\sigma(y) = \|y\|^2/2$, (x_*, λ_*) is a KKT-pair, and the restriction of $\sigma(\lambda_*, T_K^2(G(x_*), \cdot))$ to its effective domain is u.s.c. (see [72], formulae (3.7), (3.23) and (3.25), and [29], Theorem 3.3). In this case one has

$$\varphi_c(h, \mu) = \min_{z \in \mathscr{C}(x_*, \lambda_*)} \left(\left\| c(DG(x_*)h - z) + \mu \right\|^2 - \sigma\left(\lambda_*, T_K^2(G(x_*), z)\right) \right) - \frac{1}{c} \|\mu\|^2,$$

where $\mathscr{C}(x_*, \lambda_*) = \{z \in T_K(G(x_*)) \mid \langle \lambda_*, z \rangle = 0\}$. In particular, the assumption holds true in Examples 10 and 12, if (x_*, λ_*) is a KKT-pair.

The assumption is satisfied in Example 2 in the case $\phi(s) = s^2/2$, if one defines

$$\varphi_c(h, \mu) = \sum_{i \in I_+(x_*, \lambda_*)} \Big(c\|\nabla g_i(x_*)h\|^2 + \mu_i \langle \nabla g_i(x_*), h \rangle \Big) - \frac{1}{c} \sum_{i \in I \setminus I(x_*)} \mu_i^2$$

$$+ \sum_{i \in I_0(x_*, \lambda_*)} \Big(\mu_i \max \Big\{ \langle \nabla g_i(x_*), h \rangle, -\frac{\mu_i}{c} \Big\} + \frac{c}{2} \max \Big\{ \langle \nabla g_i(x_*), h \rangle, -\frac{\mu_i}{c} \Big\}^2 \Big).$$

The assumption is also valid in Example 4, provided $\phi''(t) > 0$ for all $t \in \mathbb{R}$. In all other examples, one must modify the function $\Phi(y, \lambda, c)$ in order to ensure that the function $\Phi(G(x), \cdot, c)$ is smooth enough, and $D_\lambda \Phi(G(x_*), \lambda_*, c) = 0$ (this equality is necessary for $\Phi(G(x), \lambda, c)$ to admit the second order expansion in (x, λ); see Def. 7).

Namely, suppose for the sake of shortness that there are no equality constraints, and choose a twice continuously differentiable function $\zeta \colon \mathbb{R} \to \mathbb{R}$ such that $\zeta'(0) = 0$, $\zeta(\mathbb{R}_+) = \mathbb{R}_+$, and $\zeta(t) = 0$ iff $t = 0$. Let us replace the function $\Phi(y, \lambda, c)$ with the function $\Phi(y, \zeta(\lambda), c)$, where $\zeta(\lambda) = (\zeta(\lambda_1), \ldots, \zeta(\lambda_l))$ in the case when $(\mathcal{P})$ is a mathematical programming problem, and $\zeta(\lambda)$ is Löwner's operator associated with ζ in the case when $(\mathcal{P})$ is either a second order cone or semidefinite programming problem. Then one can verify that the assumption on the function $\Phi(G(x), \zeta(\lambda), c)$ is satisfied in Example 3 with $\zeta(t) = t^2$ iff s.c. condition holds true. The assumption holds true in Examples 5–7 iff $\phi'(0) \neq 0$, $\phi''(0) > 0$, and s.c. condition holds true. This assumption is satisfied in Example 8 iff $\phi'(b) \neq 0$, $\phi''(b) > -\phi'(b)^2$, and s.c. condition holds true, and it is always satisfied in Example 9, provided $\zeta(t) = t^{1+\varepsilon}$ for some $\varepsilon > 0$.

The assumption is also satisfied in Examples 13 and 14, if s.c. condition holds true ([74], Thm. 5.1, and [55], Prp. 4.2), and $\zeta(t) = t^2$.

Finally, let us note that under some natural assumptions on the function ζ (that are satisfied, e.g., if $\zeta(t) = t^2$) the function $\Phi(y, \zeta(\lambda), c)$ has the same properties as the function $\Phi(y, \lambda, c)$. In particular, $\Phi(y, \zeta(\lambda), c)$ satisfies the same main assumption of this paper as $\Phi(y, \lambda, c)$.

Now, we can prove that under some additional assumptions the penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$ is locally exact at a KKT-pair (x_*, λ_*) of the problem $(\mathcal{P})$, provided this KKT-pair satisfies the second order sufficient optimality condition, and the function $\eta(x_*, \cdot)$ behaves like a positive definite quadratic function in a neighbourhood of λ_*.

Theorem 8. *Let Y be finite dimensional, assumptions (A4) and (A11) be satisfied, x_* be a locally optimal solution of the problem $(\mathcal{P})$, f and G be twice Fréchet differentiable at x_*, and (x_*, μ_*) be a KKT-pair of the problem $(\mathcal{P})$ satisfying the second order sufficient optimality condition. Suppose also that the function $\Phi(G(x), \lambda, c)$ admits the second order expansion in (x, λ) at (x_*, λ_*) for some $\lambda_* \in \Phi_0^{-1}(\mu_*)$, the function η is twice Fréchet differentiable at (x_*, λ_*), $D_{\lambda\lambda}^2 \eta(x_*, \lambda_*)$ is positive definite, and $\eta(x_*, \lambda_*) = 0$. Then the penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$ is locally exact at (x_*, λ_*).*

Proof. Denote $\xi = (x, \lambda)$, $\xi_* = (x_*, \lambda_*)$ and $\overline{\xi} = (x_*, \mu_*)$. Under the assumptions of the theorem, for any $c > 0$ there exists a neighbourhood U_c of ξ_* such that

for all $\xi = (x, \lambda) \in U_c$ one has

$$
\left| \mathscr{L}_e(\xi, c) - \mathscr{L}_e(\xi_*, c) - \langle D_x L(\overline{\xi}), x - x_* \rangle - \frac{1}{2} \langle x - x_*, D^2_{xx} L(\overline{\xi})(x - x_*) \rangle \right.
$$
$$
\left. - \frac{1}{2} \varphi_c(x - x_*, \lambda - \lambda_*) - \frac{1}{2} D^2 \eta(\xi_*)(\xi - \xi_*, \xi - \xi_*) \right|
$$
$$
< \frac{1}{2c} \Big(\|x - x_*\|^2 + \|\lambda - \lambda_*\|^2 \Big). \quad (51)
$$

Here we used the equalities $D_x \eta(\xi_*) = 0$ and $D_\lambda \eta(\xi_*) = 0$ that hold true due to the facts that $\eta(x, \lambda)$ is nonnegative, and $\eta(\xi_*) = 0$.

Arguing by reductio ad absurdum, suppose that $\mathscr{L}_e(x, \lambda, c)$ is not locally exact at (x_*, λ_*). Then for any $n \in \mathbb{N}$ there exists $\xi_n = (x_n, \lambda_n) \in U_n \cap (A \times \Lambda)$ such that $\mathscr{L}_e(\xi_n, n) < \mathscr{L}_e(\xi_*, n)$. Applying (51) one gets that

$$
0 > \langle D_x L(\overline{\xi}), x_n - x_* \rangle + \frac{1}{2} \langle x_n - x_*, D^2_{xx} L(\overline{\xi})(x_n - x_*) \rangle + \frac{1}{2} \varphi_n(x_n - x_*, \lambda_n - \lambda_*)
$$
$$
+ \frac{1}{2} D^2 \eta(\xi_*)(\xi_n - \xi_*, \xi_n - \xi_*) - \frac{1}{2n} \Big(\|x_n - x_*\|^2 + \|\lambda_n - \lambda_*\|^2 \Big) \quad (52)
$$

for any $n \in \mathbb{N}$. Denote $\alpha_n = \|x_n - x_*\| + \|\lambda_n - \lambda_*\|$, $h_n = (x_n - x_*)/\alpha_n$ and $\nu_n = (\lambda_n - \lambda_*)/\alpha_n$. Replacing, if necessary, the sequence $\{(x_n, \lambda_n)\}$ by its subsequence, one can suppose that the sequence $\{(h_n, \nu_n)\}$ converges to a point (h_*, ν_*) such that $\|h_*\| + \|\nu_*\| = 1$ (recall that both X and Y are finite dimensional). Note that $h_* \in T_A(x_*)$.

Suppose, at first, that $h_* = 0$. In other words, suppose that $(x_n - x_*)/\alpha_n$ converges to zero as $n \to \infty$. Observe that

$$
x_n - x_* \in T_A(x_*), \quad \langle D_x L(\overline{\xi}), x_n - x_* \rangle \geq 0 \quad \forall n \in \mathbb{N} \quad (53)
$$

due to the facts that $x_n \in A$, A is convex, and (x_*, μ_*) is a KKT-pair. Hence dividing (52) by α_n^2, and passing to the limit as $n \to \infty$ one obtains that $D^2_{\lambda\lambda} \eta(x_*, \lambda_*)(\nu_*, \nu_*) \leq 0$ by virtue of Def. 7, which contradicts our assumption that $D^2_{\lambda\lambda} \eta(x_*, \lambda_*)$ is positive definite. Thus, $h_* \neq 0$.

Recall that η is a nonnegative function and $\eta(\xi_*) = 0$, i.e. ξ_* is a point of global minimum of the function η. Therefore $D^2 \eta(\xi_*)(\xi_n - \xi_*, \xi_n - \xi_*) \geq 0$ for all $n \in \mathbb{N}$. Applying this estimate and (53) in (52), and dividing by α_n^2 one gets that

$$
0 > \langle h_n, D^2_{xx} L(\overline{\xi}) h_n \rangle + \varphi_n(h_n, \mu_n) - \frac{1}{n} \quad (54)
$$

for any $n \in \mathbb{N}$. Here we used the fact that $\alpha^2 \varphi_c(h, \mu) = \varphi_c(\alpha h, \alpha \mu)$ for all $\alpha \geq 0$ by Def. 7.

Passing to the limit superior in (54) with the use of Def. 7 one obtains that $h_* \in C(x_*, \lambda_*)$ and $0 \geq \langle h_*, D^2_{xx} L(\overline{\xi}) h_* \rangle - \sigma(\mu_*, \mathcal{T}(h_*))$, which contradicts the assumption that the KKT-pair (x_*, μ_*) satisfies the second order sufficient optimality condition. $\quad\square$

Remark 24. Note that the theorem above contains some existing results (e.g. [22], Theorem 6.3; [54], Theorem 4.1) as simple particular cases. Furthermore, it is easy to verify that under the assumptions of the theorem there exist $\gamma > 0$ and $c_0 > 0$ such that $\mathscr{L}_e(x, \lambda, c) \geq \mathscr{L}_e(x_*, \lambda_*, c) + \gamma(\|x - x_*\|^2 + \|\lambda - \lambda_*\|^2)$ for all $c \geq c_0$ and $(x, \lambda) \in A \times \Lambda$ that are sufficiently close to (x_*, λ_*).

Let us now provide several particular examples of globally exact augmented Lagrangian functions, and demonstrate how one can utilize the theorem above and the localization principle in order to prove the global exactness of these augmented Lagrangians.

8.2 Mathematical Programming

Let the problem $(\mathcal{P})$ be the mathematical programming problem of the form

$$\min f(x) \quad \text{subject to} \quad g_i(x) \le 0, \quad i \in I, \quad g_j(x) = 0, \quad j \in J, \qquad (55)$$

where $g_i \colon X \to \mathbb{R}$, $I = \{1, \ldots, l\}$ and $J = \{l+1, \ldots, l+s\}$. Suppose that the functions f and g_i, $i \in I \cup J$ are twice continuously differentiable. Define

$$\eta_1(x, \lambda) = \sum_{i=1}^{l} \Big(\langle D_x L(x, \lambda), \nabla g_i(x) \rangle + g_i(x)^2 \lambda_i \Big)^2 + \sum_{j=l+1}^{l+s} \big\langle D_x L(x, \lambda), \nabla g_j(x) \big\rangle^2$$

(see [22]). Let x_* be a locally optimal solution of problem (55), and let LICQ holds at x_*. Then one can easily verify that $\eta_1(x_*, \lambda_*) = 0$ for some $\lambda_* \in \mathbb{R}^{l+s}$ iff (x_*, λ_*) is a KKT-pair of problem (55) (note that such λ_* is unique). Furthermore, the matrix $D^2_{\lambda\lambda} \eta_1(x_*, \lambda_*)$ is positive definite. Thus, one can apply Theorem 8 in order to prove the local exactness of the penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$ for problem (55).

In order to ensure that the function $\mathscr{L}_e(\cdot, \cdot, c)$ is level-bounded and, thus, globally exact, one must add barrier terms into the definition of this function (see [22], and Remark 12 above). Choose $\alpha > 0$ and $\varkappa > 2$, and define

$$p(x, \lambda) = \frac{a(x)}{1 + \sum_{i=1}^{l} \lambda_i^2}, \quad q(x, \lambda) = \frac{b(x)}{1 + \sum_{j=l+1}^{l+s} \lambda_j^2},$$

where

$$a(x) = \alpha - \sum_{i=1}^{l} \max\{0, g_i(x)\}^\varkappa, \quad b(x) = \alpha - \sum_{j=l+1}^{l+s} g_j(x)^2.$$

Denote $\Omega_\alpha = \{x \in \mathbb{R}^d \mid a(x) > 0, b(x) > 0\}$. Then one can introduce the following penalized augmented Lagrangian function for problem (55) (see Example 2). Namely, define

$$\mathscr{L}_e(x, \lambda, c) = f(x)$$
$$+ \sum_{i=1}^{l} \Big(\lambda_i \max\Big\{ g_i(x), -\frac{p(x, \lambda)}{c} \lambda_i \Big\} + \frac{c}{2p(x, \lambda)} \max\Big\{ g_i(x), -\frac{p(x, \lambda)}{c} \lambda_i \Big\}^2 \Big)$$
$$+ \sum_{j=l+1}^{l+s} \Big(\lambda_j g_j(x) + \frac{c}{2q(x, \lambda)} g_j(x)^2 \Big) + \eta_1(x, \lambda), \quad (56)$$

if $x \in \Omega_\alpha$, and $\mathscr{L}_e(x, \lambda, c) = +\infty$ otherwise.

One can easily verify that the augmented Lagrangian function introduced above is l.s.c. jointly in (x, λ) on $\mathbb{R}^d \times \mathbb{R}^{l+s}$, and continuously differentiable in (x, λ) on its effective domain, i.e. on $\Omega_\alpha \times \mathbb{R}^{l+s}$ (cf. [22]). One can also check

that the function $\Phi(y, \lambda, c)$ corresponding to penalized augmented Lagrangian function (56) satisfies assumptions $(A2)$, $(A4)_s$, $(A6)_s$, $(A11)$ with $\Phi_0(\lambda) \equiv \lambda$, and $(A12)$. Furthermore, the function $\Phi(G(x), \lambda, c)$ admits the second order expansion in (x, λ) at every KKT-pair of problem (55), and $\Phi(G(x_*), \lambda_*, c) = 0$, if (x_*, λ_*) is a KKT-pair. Therefore one can obtain the following result.

Theorem 9. *Let f and g_i, $i \in I \cup J$, be twice continuously differentiable, LICQ hold true at every globally optimal solution of problem (55), and for any $x_* \in \Omega_*$ a unique KKT-pair (x_*, λ_*) satisfy the second order sufficient optimality condition. Then penalized augmented Lagrangian function (56) is globally exact if and only if there exists $c_0 > 0$ such that the set*

$$S_e(c_0) := \left\{ (x, \lambda) \in \mathbb{R}^d \times \mathbb{R}^{l+s} \mid \mathscr{L}_e(x, \lambda, c_0) < f_* \right\}$$

is either bounded or empty. In particular, if there exists $\gamma > 0$ such that the set

$$\Omega(\gamma, \alpha) := \left\{ x \in \mathbb{R}^d \mid f(x) < f_* + \gamma, \; a(x) > 0, \; b(x) > 0 \right\}$$

is bounded, then penalized augmented Lagrangian function (56) is globally exact in the sense that its points of global minimum in (x, λ) on $\mathbb{R}^d \times \mathbb{R}^{l+s}$ are exactly KKT-pairs of problem (55) corresponding to globally optimal solutions of this problem, provided $c > 0$ is large enough.

Proof. Under the assumptions of the theorem augmented Lagrangian (56) is locally exact at every KKT-pair corresponding to a globally optimal solution of problem (55) by Theorem 8. Then applying the localization principle (Theorem 6) one obtains that this augmented Lagrangian is globally exact iff the set $S_e(c_0)$ is either bounded or empty for some $c_0 > 0$.

Let us verify that if the set $\Omega(\gamma, \alpha)$ is bounded, then the set $S_e(c_0)$ is bounded for some $c_0 > 0$. Minimizing the function $q(t) = \lambda_i t + ct^2 / 2p(x, \lambda)$ one obtains that for any $(x, \lambda) \in \Omega_\alpha \times \mathbb{R}^{l+s}$ and $c > 0$ the following inequalities hold true:

$$\mathscr{L}_e(x, \lambda, c) \geq f(x) - \frac{p(x, \lambda)}{2c} \sum_{i=1}^{l} \lambda_i^2 - \frac{q(x, \lambda)}{2c} \sum_{j=l+1}^{l+s} \lambda_j^2 + \eta_1(x, \lambda)$$

$$\geq f(x) - \frac{\alpha}{c} + \eta_1(x, \lambda). \quad (57)$$

Consequently, for any $c > \alpha/\gamma$ and $(x, \lambda) \in S_e(c)$ one has $x \in \Omega(\gamma, \alpha)$. Hence applying (57) and the fact that the set $\Omega(\gamma, \alpha)$ is bounded one gets that there exists $\tau \in \mathbb{R}$ such that $\mathscr{L}_e(x, \lambda, c) \geq \tau$ for all $(x, \lambda) \in \mathbb{R}^d \times \mathbb{R}^{l+s}$ and $c > \alpha/\gamma$.

Arguing by reductio ad absurdum, suppose that the set $S_e(c)$ is unbounded for any $c > 0$. Choose an increasing unbounded sequence $\{c_n\} \subset \mathbb{R}_+$ such that $c_1 > \alpha/\gamma$. Then for any $n \in \mathbb{N}$ there exists $(x_n, \lambda_n) \in S_e(c_n)$ such that $\|\lambda_n\| \geq n$. By the choice of c_1 one has that $\{x_n\} \subset \Omega(\gamma, \alpha)$. Therefore without loss of generality one can suppose that the sequence $\{x_n\}$ converges to a point x_*. Let us show that x_* is a globally optimal solution of problem (55).

Denote $w_n^j = q(x_n, \lambda_n)(\lambda_n)_j g_j(x_n) + c_n g_j(x_n)^2 / 2$

$$u_n^i = p(x_n, \lambda_n)(\lambda_n)_i \max \left\{ g_i(x_n), -\frac{p(x_n, \lambda_n)}{c_n}(\lambda_n)_i \right\}$$

$$+ \frac{c_n}{2} \max \left\{ g_i(x_n), -\frac{p(x_n, \lambda_n)}{c_n}(\lambda_n)_i \right\}^2.$$

Arguing as above, one can easily verify that $\liminf_{n\to\infty} u_n/p(x_n, \lambda_n) \ge 0$ and $\liminf_{n\to\infty} w_n/q(x_n, \lambda_n) \ge 0$.

From (57) and the fact that $(x_n, \lambda_n) \in S_e(c_n)$ it follows that the sequences $\{f(x_n)\}$ and $\{\eta_1(x_n, \lambda_n)\}$ are bounded. Moreover, by the definition of $\{(x_n, \lambda_n)\}$ one has $\tau \le \mathscr{L}_e(x_n, \lambda_n, c_n) < f_*$. Therefore the sequence $\sum_{i=1}^{l} u_n^i/p(x_n, \lambda_n) + \sum_{j=l+1}^{l+s} w_n^j/q(x_n, \lambda_n)$ is bounded, which implies that $\{u_n^i/p(x_n, \lambda_n)\}$, $i \in I$, and $\{w_n^j/q(x_n, \lambda_n)\}$, $j \in J$, are bounded sequences. By definition $0 < p(x_n, \lambda_n) \le \alpha$ and $0 < q(x_n, \lambda_n) \le \alpha$ for all $n \in \mathbb{N}$. Therefore, the sequences $\{u_n^i\}$ and $\{w_n^j\}$ are bounded as well. Hence taking into account the fact that $c_n \to +\infty$ as $n \to \infty$ one can easily check that $\max\{g_i(x_n), 0\} \to 0$ and $g_j(x_n) \to 0$ as $n \to \infty$ for all $i \in I$ and $j \in J$, which implies that x_* is a feasible point.

Indeed, suppose, for instance, that $g_j(x_n)$ does not converge to zero. Then there exist $\varepsilon > 0$ and a subsequence $\{x_{n_k}\}$ such that $|g_j(x_{n_k})| \ge \varepsilon$ for all $k \in \mathbb{N}$. Consequently, $w_{n_k}^j \ge c_{n_k}^2 \varepsilon^2/2 - \alpha|g_j(x_{n_k})|$. Passing to the limit as $k \to \infty$ one obtains that $w_{n_k}^j \to +\infty$ as $k \to \infty$, which is impossible. Thus, x_* is a feasible point of problem (55).

From (57) and the fact that $(x_n, \lambda_n) \in S_e(c_n)$ it follows that $f(x_*) \le f_*$, which implies that x_* is a globally optimal solution of problem (55). Consequently, LICQ holds at this point.

Observe that the function $\eta(x, \cdot)$ is quadratic. Hence and from the fact that LICQ holds at x_* one obtains that there exist a neighbourhood U of x_*, $\delta_1 > 0$ and $\delta_2, \delta_3 \in \mathbb{R}$ such that $\eta(x, \lambda) \ge \delta_1\|\lambda\|^2 + \delta_2\|\lambda\| + \delta_3$ for all $x \in U$ and $\lambda \in \mathbb{R}^{l+s}$. Applying this estimate in (57), and taking into account the fact that $\|\lambda_n\| \to +\infty$ as $n \to \infty$ one gets that $\mathscr{L}_e(x_n, \lambda_n, c_n) \to +\infty$ as $n \to \infty$, which contradicts the definition of the sequence $\{(x_n, \lambda_n)\}$. Thus, $S_e(c)$ is bounded for some $c > 0$. □

Remark 25. (i) Note that the theorem above strengthens all existing results on global exactness of augmented Lagrangian functions (see, e.g., [22], Thm. 4.6; [26], Prp. 1; [54], Thms. 4.3 and 4.4), since it provides first *necessary and sufficient* conditions for the global exactness, and is formulated for the optimization problem with *both* equality and inequality constraints. Furthermore, to the best of author's knowledge the theorem above provides first sufficient conditions for the *global* exactness of augmented Lagrangian functions for equality constrained optimization problems (cf. [14, 16, 32, 31, 26]).
(ii) It should be mentioned that one can construct an exact augmented Lagrangian function from Mangasarian's augmented Lagrangian (Example 4). However, since $\Phi_0(\lambda) \equiv \phi'(\lambda)$ for this augmented Lagrangian, one has to consider the penalty term $\eta(x, \lambda) = \eta_1(x, \phi'(\lambda))$, and impose some rather restrictive assumptions on the function ϕ in order to ensure the global exactness of the corresponding penalized augmented Lagrangian.

Let us extend the previous theorem to the case of other augmented Lagrangian functions. As it was noted in Remark 23, only augmented Lagrangian functions from Examples 2 and 4 admit the second order expansion in (x, λ) without some modifications of the function $\Phi(y, \lambda, c)$. In order to accommodate necessary modifications of this function one needs to modify the penalty term $\eta(x, \lambda)$ as well. Below, we utilize the transformation $\lambda \to \zeta(\lambda)$ with $\zeta(t) = t^2$ in order to ensure the desired properties of the function $\Phi(y, \lambda, c)$. Therefore we

define
$$\eta_2(x,\lambda) = \sum_{i=1}^{l} \left\langle D_x L(x,\zeta(\lambda)), \nabla g_i(x) \right\rangle^2 + \sum_{i=1}^{l} g_i(x)^2 \lambda_i^2.$$

Hereinafter, $\zeta(\lambda) = (\lambda_1^2, \ldots, \lambda_l^2)$.

Let x_* be a locally optimal solution of problem (55), and let LICQ holds at x_*. Then, as in the case of $\eta_1(x,\lambda)$, one can easily verify that $\eta_2(x_*, \lambda_*) = 0$ for some $\lambda_* \in \mathbb{R}^{l+s}$ iff $(x_*, \zeta(\lambda_*))$ is a KKT-pair of problem (55). Furthermore, the matrix $D_{\lambda\lambda}^2 \eta_1(x_*, \lambda_*)$ is positive definite, provided the pair (x_*, λ_*) satisfies s.c. condition. Finally, one can check that there exists a neighbourhood U of x_* such that $\inf_{x \in U} \eta_2(x,\lambda) \to +\infty$ as $\|\lambda\| \to +\infty$.

Now, we can define the following penalized augmented Lagrangian functions. Suppose that there are no equality constraints, and define

$$\mathscr{L}_e(x,\lambda,c) = f(x) + \eta_2(x,\lambda)$$
$$+ \frac{1}{3cp(x,\zeta(\lambda))^2} \sum_{i=1}^{l} \left[\max\left\{ cg_i(x) + p(x,\zeta(\lambda))\lambda_i, 0 \right\}^3 - \left| p(x,\zeta(\lambda))\lambda_i \right|^3 \right], \quad (58)$$

if $x \in \Omega_\alpha$, and $\mathscr{L}_e(x,\lambda,c) = +\infty$ otherwise (see Example 3). Similarly, one can define

$$\mathscr{L}_e(x,\lambda,c) = f(x) + \frac{p(x,\lambda)}{c} \sum_{i=1}^{l} \left[\lambda_i^2 \phi\left(\frac{cg_i(x)}{p(x,\lambda)} \right) + \xi\left(\frac{cg_i(x)}{p(x,\lambda)} \right) \right] + \eta_2(x,\lambda), \tag{59}$$

if $x \in \Omega_\alpha$, and $\mathscr{L}_e(x,\lambda,c) = +\infty$ otherwise (see Example 6). Here the function $\phi \colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is non-decreasing, strictly convex, and such that $\operatorname{dom}\phi = (-\infty, \varepsilon_0)$ for some $\varepsilon_0 \in (0, +\infty]$, $\phi(t) \to +\infty$ as $t \to \varepsilon_0$, $\phi(t)/t \to +\infty$ as $t \to +\infty$ if $\varepsilon_0 = +\infty$, ϕ is twice continuously differentiable on $\operatorname{dom}\phi$, $\phi(0) = 0$, $\phi'(0) = 1$, and $\phi''(0) > 0$, while the function $\xi \colon \mathbb{R} \to \mathbb{R}_+$ is twice continuously differentiable, and such that ξ is strictly convex on $\mathbb{R}_+$, $\xi(t) = 0$ for all $t \le 0$, and $\xi(t) > 0$ for all $t > 0$.

Finally, one can define

$$\mathscr{L}_e(x,\lambda,c) = f(x) + \frac{1}{cp(x,\lambda)} \sum_{i=1}^{l} \int_0^{cg_i(x)} \left(\sqrt{t^2 + p(x,\lambda)^2 \lambda_i^4} + t \right) dt + \eta_2(x,\lambda), \tag{60}$$

if $x \in \Omega_\alpha$, and $\mathscr{L}_e(x,\lambda,c) = +\infty$ otherwise (see Example 9).

As in the case of augmented Lagrangian (56), one can easily verify that the penalized augmented Lagrangian functions (58), (59) and (60) are l.s.c. jointly in (x,λ) on $\mathbb{R}^d \times \mathbb{R}^{l+s}$, and continuously differentiable in (x,λ) on their effective domains. One can also check that the functions $\Phi(y,\lambda,c)$ corresponding to these augmented Lagrangians satisfy assumptions $(A2)$, $(A4)_s$, $(A6)_s$, $(A11)$ with $\Phi_0(\lambda) \equiv \zeta(\lambda)$, and $(A12)$ in the case $\Lambda = Y^* = \mathbb{R}^{l+s}$. Furthermore, the corresponding functions $\Phi(G(x),\lambda,c)$ admit the second order expansion in (x,λ) at every point (x_*,λ_*) such that x_* is feasible, and λ_* satisfies s.c. condition. Finally, $\Phi(G(x_*),\lambda_*,c) = 0$, if x_* is feasible, and $\eta(x_*,\lambda_*) = 0$.

Thus, one can apply the localization principle (Theorem 6) and Theorem 8 in order to prove the global exactness of penalized augmented Lagrangian functions

51

(58) and (59). Augmented Lagrangian (60), as one can verify, is not bounded from below, which implies that it is not globally exact.

Arguing in the same way as in the proof of Theorem 9 one can verify that the following result holds true.

Theorem 10. *Let f and g_i, $i \in I$, be twice continuously differentiable, LICQ hold true at every globally optimal solution of problem (55), and for any $x_* \in \Omega_*$ a unique KKT-pair (x_*, λ_*) satisfy the second order sufficient optimality and the strict complementarity conditions. Then penalized augmented Lagrangian functions (58) and (59) are globally exact (with respect to the function $\eta_2(x, \lambda)$) if and only if there exists $c_0 > 0$ such that the corresponding set $S_e(c_0)$ is either bounded or empty. In particular, if there exists $\gamma > 0$ such that the set $\Omega(\gamma, \alpha)$ is bounded, then penalized augmented Lagrangian functions (58) and (59) (provided ϕ is bounded from below) are globally exact in the sense that for any sufficiently large $c > 0$ a pair (x_*, λ_*) is a point of global minimum of these functions in (x, λ) on $\mathbb{R}^d \times \mathbb{R}^l$ iff $x_* \in \Omega_*$ and $(x_*, \zeta(\lambda_*))$ is a KKT-pair of problem (55).*

As it was noted above, augmented Lagrangian (60) is not globally exact. However, with the use of Theorems 7 and 8 one can provide simple sufficient contidions for the the exactness of this augmented Lagrangian on bounded sets. Namely, the following result holds true.

Theorem 11. *Let f and g_i, $i \in I$, be twice continuously differentiable, LICQ hold true at every globally optimal solution of problem (55), and for any $x_* \in \Omega_*$ a unique KKT-pair (x_*, λ_*) satisfy the second order sufficient optimality and the strict complementarity conditions. Then penalized augmented Lagrangian function (60) is exact on any bounded subset of $\mathbb{R}^d \times \mathbb{R}^l$ with respect to the function η_2.*

Remark 26. (i) To the best of author's knowledge, all exact augmented Lagrangian function studied in the literature were constructed from the Hestenes-Powell-Rockafellar augmented Lagrangian function. Thus, penalized augmented Lagrangian functions (58), (59) and (60) as well as the theorems on exactness of these functions are completely new. Furthermore, even augmented Lagrangian (56) has not been considered for the case of general equality and inequality constraints before.

(ii) Note that in order to guarantee that there is a unique point of global minimum of augmented Lagrangians (58) and (59) in (x, λ) corresponding to a globally optimal solution of problem (55) one can add the penalty term $\omega(\lambda) = \sum_{i=1}^{l} \max\{0, -\lambda_i\}^\varkappa$ to these functions. Note also that one can incorporate equality constraints into augmented Lagrangians (58), (59) and (60) in the same way as in (56).

(iii) It should be mentioned that globally exact augmented Lagrangian functions cannot be constructed from the exponential penalty function (Example 5), the modified barrier function (Example 7) and the p-th power augmented Lagrangian (Example 8) without some nontrivial transformations of these functions due to the facts that for these augmented Lagrangians one has $\mathscr{L}(x, 0, c) \equiv f(x)$, and it is very difficult (if at all possible) to construct a continuously differentiable function $\eta(x, \lambda)$ that satisfies the main assumption of this article, and such that $f(x) + \eta(x, 0) \geq f_*$ for all $x \in \mathbb{R}^d$.

8.3 Nonlinear Second Order Cone Programming

Let the problem $(\mathcal{P})$ be the nonlinear second order cone programming problem of the form

$$\min f(x) \quad \text{subject to} \quad g_i(x) \in Q_{l_i+1}, \quad i \in I, \quad h(x) = 0, \tag{61}$$

where the functions $f\colon X \to \mathbb{R}$, $g_i\colon X \to \mathbb{R}^{l_i+1}$, $I = \{1, \ldots, r\}$, and $h\colon X \to \mathbb{R}^s$ are twice continuously differentiable, and Q_{l_i+1} is the second order (Lorentz) cone of dimension $l_i + 1$.

Let x_* be a locally optimal solution of problem (61). Recall that the point x_* is called *nondegenerate* (see [4], Def. 4.70), if

$$
\begin{bmatrix} Jg_1(x_*) \\ \vdots \\ Jg_r(x_*) \\ Jh(x_*) \end{bmatrix} \mathbb{R}^d +
\begin{bmatrix} \lim T_{Q_{l_1+1}}\big(g_1(x_*)\big) \\ \vdots \\ \lim T_{Q_{l_r+1}}\big(g_r(x_*)\big) \\ \{0\} \end{bmatrix} =
\begin{bmatrix} \mathbb{R}^{l_1+1} \\ \vdots \\ \mathbb{R}^{l_r+1} \\ \mathbb{R}^s \end{bmatrix}.
$$

where $Jg_i(x)$ is the Jacobian of $g_i(x)$, and "lin" stands for the lineality subspace of a convex cone, i.e. the largest linear space contained in this cone. Let us note that the nondegeneracy condition can be expressed as a "linear independence-type" condition (see [34], Lemma 3.1, and [3], Proposition 19). Furthermore, by [4], Proposition 4.75, the nondegeneracy condition guarantees that there exists a *unique* Lagrange multiplier at x_*.

Being inspired by the ideas of [34], for any $x \in X$ and $\lambda = (\lambda_1, \ldots, \lambda_r, \mu) \in Y^* := \mathbb{R}^{l_1+1} \times \ldots \times \mathbb{R}^{l_r+1} \times \mathbb{R}^s$ define

$$\eta(x, \lambda) = \|D_x L(x, \lambda)\|^2 + \sum_{i=1}^r \Big(\langle \lambda_i, g_i(x)\rangle^2 + \|(\lambda_i)_0 \overline{g}_i(x) + (g_i)_0(x)\overline{\lambda}_i\|^2 \Big),$$

where $\lambda_i = ((\lambda_i)_0, \overline{\lambda}_i) \in \mathbb{R} \times \mathbb{R}^{l_i}$, and the same notation is used for $g_i(x)$. Suppose that x_* is a nondegenerate locally optimal solution of problem (61). Then arguing in the same way as in the proof of Proposition 3.3 in [34] one can verify that $\eta(x_*, \lambda_*) = 0$ for some λ_* iff (x_*, λ_*) is a KKT-pair, and such λ_* is unique. Furthermore, the matrix $D^2_{\lambda\lambda}\eta(x_*, \lambda_*)$ is positive definite. Therefore one can utilize Theorem 8 in order to prove the local exactness of the penalized augmented Lagrangian function $\mathscr{L}_e(x, \lambda, c)$ for problem (61).

As in the case of the mathematical programming problem, one must add barrier terms into the definition of $\mathscr{L}_e(x, \lambda, c)$ in order to ensure that it is level-bounded and globally exact. Choose $\alpha > 0$ and $\varkappa > 2$, and for any $\lambda = (\lambda_1, \ldots, \lambda_r, \mu) \in Y^*$ define

$$p(x, \lambda) = \frac{a(x)}{1 + \sum_{i=1}^r \|\lambda_i\|^2}, \quad q(x, \lambda) = \frac{b(x)}{1 + \|\mu\|^2}, \tag{62}$$

where

$$a(x) = \alpha - \sum_{i=1}^r \operatorname{dist}^\varkappa\big(g_i(x), Q_{l_i+1}\big), \quad b(x) = \alpha - \|h(x)\|^2.$$

Denote $\Omega_\alpha = \{x \in \mathbb{R}^d \mid a(x) > 0, b(x) > 0\}$. Then one can introduce a penalized augmented Lagrangian function for problem (61) as follows (see Example 10,

and Remark 23). For any $\lambda = (\lambda_1, \ldots, \lambda_r, \mu) \in Y^*$ define

$$
\begin{aligned}
\mathscr{L}_e(x, \lambda, c) = f(x) \\
+ \frac{c}{2p(x,\lambda)} \sum_{i=1}^{r} \Big[\operatorname{dist}^2 \Big(g_i(x) + \frac{p(x,\lambda)}{c}\lambda_i, Q_{l_i+1} \Big) - \frac{p(x,\lambda)^2}{c^2} \|\lambda_i\|^2 \Big] \\
+ \langle \mu, h(x) \rangle + \frac{c}{2q(x,\lambda)} \|h(x)\|^2 + \eta(x, \lambda). \quad (63)
\end{aligned}
$$

if $x \in \Omega_\alpha$, and $\mathscr{L}_e(x, \lambda, c) = +\infty$, otherwise.

Observe that the function $\mathscr{L}_e(x, \lambda, c)$ is l.s.c. jointly in (x, λ) on $\mathbb{R}^d \times Y^*$, and continuously differentiable in (x, λ) on its effective domain $\Omega_\alpha \times Y^*$ by [4], Theorem 4.13. One can also check that the function $\Phi(y, \lambda, c)$ corresponding to penalized augmented Lagrangian function (63) satisfies assumptions $(A2)$, $(A4)_s$, $(A6)_s$, $(A11)$ with $\Phi_0(\lambda) \equiv \lambda$, and $(A12)$. Furthermore, $\Phi(G(x_*), \lambda_*, c) = 0$, if (x_*, λ_*) is a KKT-pair, and arguing in the same way as in ([72], pp. 487–488) one can check that $\Phi(G(x), \lambda, c)$ admits the second order expansion in (x, λ) at every KKT-pair of problem (61) (see also [29]). Therefore one can obtain the following result.

Theorem 12. *Let the functions f, g_i, $i \in I$, and h be twice continuously differentiable. Suppose also that every globally optimal solution of problem (61) is nondegenerate, and for any $x_* \in \Omega_*$ a unique KKT-pair (x_*, λ_*) satisfies the second order sufficient optimality condition. Then penalized augmented Lagrangian function (63) is globally exact if and only if there exists $c_0 > 0$ such that the set $S_e(c_0) := \{(x, \lambda) \in \mathbb{R}^d \times Y^* \mid | \mathscr{L}_e(x, \lambda, c_0) < f_*\}$ is either bounded or empty. In particular, if the set $\Omega(\gamma, \alpha) = \{x \in \mathbb{R}^d \mid f(x) < f_* + \gamma, \ a(x) > 0, \ b(x) > 0\}$ is bounded for some $\gamma > 0$, then penalized augmented Lagrangian function (63) is globally exact.*

Proof. Under the assumptions of the theorem augmented Lagrangian (63) is locally exact at every KKT-pair corresponding to a globally optimal solution of problem (61) by Theorem 8. Then applying the localization principle (Theorem 6) one obtains that this augmented Lagrangians is globally exact iff the set $S_e(c_0)$ is either bounded or empty for some $c_0 > 0$.

Suppose, now, that the set $\Omega(\gamma, \alpha)$ is bounded for some $\gamma > 0$. Let us check that in this case the set $S_e(c)$ is bounded for sufficiently large $c > 0$. From (62) and (63) it follows that for any $(x, \lambda) \in \Omega_\alpha \times Y^*$ and $c > 0$ one has

$$
\mathscr{L}_e(x, \lambda, c) \geq f(x) - \frac{\alpha}{c} + \eta(x, \lambda). \quad (64)
$$

Hence taking into account the fact that the function $\eta(x, \lambda)$ is nonnegative one obtains that for any $c > \alpha/\gamma$ and $(x, \lambda) \in S_e(c)$ one has $x \in \Omega(\gamma, \alpha)$. Consequently, there exists $\tau > -\infty$ such that $\mathscr{L}_e(x, \lambda, c) \geq \tau$ for any $(x, \lambda) \in S_e(c)$ and $c > \alpha/\gamma$ due to the boundedness of the set $\Omega(\gamma, \alpha)$.

Arguing by reductio ad absurdum, suppose that the set $S_e(c)$ is unbounded for any $c > 0$. Choose an increasing unbounded sequence $\{c_n\} \subset \mathbb{R}_+$ such that $c_1 > \alpha/\gamma$. By our assumption for any $n \in \mathbb{N}$ there exists $(x_n, \lambda_n) \in S_e(c_n)$ such that $\|\lambda_n\| \geq n$. Note that by the choice of c_1 one has $x_n \in \Omega(\gamma, \alpha)$ for all $n \in \mathbb{N}$, which implies that without loss of generality one can suppose that the sequence $\{x_n\}$ converges to a point x_*.

Let us check that x_* is a globally optimal solution of problem (61). Indeed, denote

$$u_n^i = \frac{c_n}{2} \Big[\operatorname{dist}^2 \Big(g_i(x_n) + \frac{p(x_n, \lambda_n)}{c_n}(\lambda_n)_i, Q_{l_i+1} \Big) - \frac{p(x_n, \lambda_n)^2}{c_n^2} \|(\lambda_n)_i\|^2 \Big],$$

$$w_n = q(x_n, \lambda_n)\langle \mu_n, h(x_n)\rangle + \frac{c_n}{2}\|h(x_n)\|^2,$$

where $\lambda_n = ((\lambda_n)_1, \ldots, (\lambda_n)_r, \mu_n)$, and $i \in I$. Note that $u_n^i/p(x_n, \lambda_n) \geq -\alpha/2c_n$ and $w_n/q(x_n, \lambda_n) \geq -\alpha/2c_n$ for all $i \in I$ and $n \in \mathbb{N}$. In other words, the sequences $\{u_n^i/p(x_n, \lambda_n)\}$, $i \in I$, and $\{w_n/q(x_n, \lambda_n)\}$ are bounded below.

Observe that from the fact that $(x_n, \lambda_n) \in S_e(c_n)$ and $c_n \geq c_1 > \alpha/\gamma$ due to our choice it follows that $\tau \leq \mathscr{L}_e(x_n, \lambda_n, c_n) < f_*$ for all $n \in \mathbb{N}$. Hence applying (64) and the fact that $\eta(x, \lambda)$ is nonnegative one obtains that the sequences $\{f(x_n)\}$ and $\{\eta(x_n, \lambda_n)\}$ are bounded. Therefore the sequence $\{\sum_{i=1}^r u_n^i/p(x_n, \lambda_n) + w_n/q(x_n, \lambda_n)\}$ is bounded as well, which implies that the sequences $\{u_n^i/p(x_n, \lambda_n)\}$, $i \in I$, and $\{w_n/q(x_n, \lambda_n)\}$ are bounded due to the fact that they are bounded from below. Since by definition one has $0 < p(x_n, \lambda_n) \leq \alpha$ and $0 < q(x_n, \lambda_n) \leq \alpha$ for all $n \in \mathbb{N}$, the sequences $\{u_n^i\}$, $i \in I$, and $\{w_n\}$ are bounded as well. Consequently, applying the fact that $c_n \to +\infty$ as $n \to \infty$ one can easily check that $\operatorname{dist}(g_i(x_n), Q_{l_i+1}) \to 0$ and $h(x_n) \to 0$ as $n \to \infty$, which yields that x_* is a feasible point of problem (61).

From (64) and the fact that $(x_n, \lambda_n) \in S_e(c_n)$ it follows that $f(x_n) - \alpha/c_n < f_*$. Passing to the limit as $n \to \infty$ and taking into account the fact that x_* is feasible one obtains that x_* is a globally optimal solution of problem (61).

Note that the function $\eta(x_*, \cdot)$ is quadratic. Furthermore, arguing in the same way as in the proof of Proposition 3.3 in [34] one can check that the Hessian of $\eta(x_*, \cdot)$ at λ_* is positive definite, which implies that there exist a neighbourhood U of x_*, $\delta_1 > 0$ and $\delta_2, \delta_3 \in \mathbb{R}$ such that $\eta(x, \lambda) \geq \delta_1\|\lambda\|^2 + \delta_2\|\lambda\| + \delta_3$ for all $x \in U$ and $\lambda \in Y^*$. Therefore $\eta(x_n, \lambda_n) \to +\infty$ as $n \to \infty$, since $\{x_n\}$ converges to x_* and $\|\lambda\| \geq n$. Consequently, with the use of (64) one obtains that $\mathscr{L}_e(x_n, \lambda_n, c_n) \to +\infty$ as $n \to \infty$, which contradicts the fact that $(x_n, \lambda_n) \in S_e(c_n)$. Thus, the set $S_e(c)$ is bounded for some $c > 0$. $\qquad\square$

Remark 27. To the best of author's knowledge, penalized augmented Lagrangian functions for nonlinear second-order cone programming problems have never been studied before. Thus, augmented Lagrangian function (63) is the first globally exact augmented Lagrangian function for problem (61).

8.4 Nonlinear Semidefinite Programming

Let the problem $(\mathcal{P})$ be the nonlinear semidefinite programming problem of the form

$$\min f(x) \quad \text{subject to} \quad G_0(x) \preceq 0, \quad h(x) = 0, \tag{65}$$

where the functions $f\colon X \to \mathbb{R}$, $G_0\colon X \to \mathbb{S}^l$ and $h\colon X \to \mathbb{R}^s$ are twice continuously differentiable.

Let x_* be a locally optimal solution of problem (65). Recall that the point x_* is called *nondegenerate* (see [4], Def. 4.70), if

$$\begin{bmatrix} DG_0(x_*) \\ Jh(x_*) \end{bmatrix} \mathbb{R}^d + \begin{bmatrix} \lim T_{\mathbb{S}_-^l}\big(G_0(x_*)\big) \\ \{0\} \end{bmatrix} = \begin{bmatrix} \mathbb{S}^l \\ \mathbb{R}^s \end{bmatrix}.$$

As in the case of second order cone programming problems, the above nondegeneracy condition can be rewritten as a "linear independence-type" condition. Namely, let $\operatorname{rank} G_0(x_*) = r$. Then the point x_* is nondegenerate iff the d-dimensional vectors

$$v_{ij} = \left(e_i^T \frac{\partial G_0(x_*)}{\partial x_1} e_j, \ldots, e_i^T \frac{\partial G_0(x_*)}{\partial x_d} e_j \right)^T, \quad \nabla h_k(x_*) \tag{66}$$

are linearly independent, where $1 \le i \le j \le l - r$, $e_1, \ldots e_{l-r}$ is a basis of the null space of the matrix $G_0(x_*)$, $1 \le k \le s$, and $h(x) = (h_1(x), \ldots, h_s(x))$ (see [4], Proposition 5.71).

For any $\lambda = (\lambda_0, \mu) \in Y^* = \mathbb{S}^l \times \mathbb{R}^s$ define

$$\eta(x, \lambda) = \left\| D_x L(x, \lambda) \right\|^2 + \operatorname{Tr}(\lambda_0^2 G_0(x)^2).$$

Let us demonstrate that the nondegeneracy condition ensures that the function $\eta(x, \lambda)$ has desired properties.

Lemma 4. *Let a locally optimal solution x_* of problem (65) be nondegenerate. Then there exists a unique Lagrange multiplier λ_* at x_*, and $\eta(x_*, \lambda) = 0$ if and only if $\lambda = \lambda_*$. Furthermore, the matrix $D^2_{\lambda\lambda} \eta(x_*, \lambda_*)$ is positive definite.*

Proof. The existence of a unique Lagrange multiplier λ_* at x_* follows directly from [4], Proposition 4.75. Furthermore, note that the function $\eta(x_*, \cdot)$ is quadratic. Therefore it remains to check that the matrix $D^2_{\lambda\lambda} \eta(x_*, \lambda_*)$ is positive definite.

Suppose that $\lambda = (\lambda_0, \mu) \in Y^*$ is such that $D^2_{\lambda\lambda} \eta(x_*, \lambda_*)(\lambda, \lambda) = 0$. Then $\operatorname{Tr}(\lambda_0^2 G_0(x_*)^2) = 0$. Let $G_0(x_*) = E \operatorname{diag}(\rho_1(x_*), \ldots, \rho_l(x_*)) E^T$ be a spectral decomposition of $G_0(x_*)$ such that the eigenvalues $\rho_i(x)$ are listed in the decreasing order. Then $G_0(x_*)^2 = E \operatorname{diag}(\rho_1(x_*)^2, \ldots, \rho_l(x_*)^2) E^T$. Applying the fact that the trace operator is invariant under cyclic permutations one obtains that

$$\operatorname{Tr}\left(\lambda_0^2 G_0(x_*)^2\right) = \operatorname{Tr}\left(E^T \lambda_0^2 E \operatorname{diag}(\rho_1(x_*)^2, \ldots, \rho_l(x_*)^2)\right)$$

$$= \sum_{i=1}^{l} \rho_i(x_*)^2 e_i^T \lambda_0^2 e_i = 0,$$

where e_i is an eigenvector of $G_0(x_*)$ corresponding to the eigenvalue $\rho_i(x_*)$. Let $\operatorname{rank} G_0(x_*) = r$. Then the above equalities imply that $\lambda_0 e_i = 0$ for any $i \in \{l - r + 1, \ldots, l\}$. Therefore there exists a $(l - r) \times (l - r)$ symmetric matrix Γ such that

$$E^T \lambda_0 E = \begin{pmatrix} \Gamma & 0 \\ 0 & 0 \end{pmatrix}, \quad \lambda_0 = E \begin{pmatrix} \Gamma & 0 \\ 0 & 0 \end{pmatrix} E^T = E_0 \Gamma E_0^T,$$

where E_0 is a $l \times (l - r)$ matrix whose columns are a basis of the null space of the matrix $G_0(x_*)$ (note that in the case $r = l$ one has $\lambda_0 = 0$).

Since $D^2_{\lambda\lambda} \eta(x_*, \lambda_*)(\lambda, \lambda) = 0$, for the function $\omega(\cdot) = \|D_x L(x_*, \cdot)\|^2$ one has

$D^2\omega(\lambda_*)(\lambda,\lambda) = 0$ or, equivalently,

$$\sum_{i=1}^{d}\mathrm{Tr}\left(\lambda_0\frac{\partial G_0(x_*)}{\partial x_i}\right)\cdot\mathrm{Tr}\left(\lambda_0\frac{\partial G_0(x_*)}{\partial x_i}\right)+2\sum_{j=1}^{s}\sum_{i=1}^{d}\mathrm{Tr}\left(\lambda_0\frac{\partial G_0(x_*)}{\partial x_i}\right)\cdot\mu_j\frac{\partial h_j(x_*)}{\partial x_i}$$

$$+\sum_{j,k=1}^{s}\sum_{i=1}^{d}\mu_j\mu_k\frac{\partial h_j(x_*)}{\partial x_i}\frac{\partial h_k(x_*)}{\partial x_i}=0. \quad (67)$$

Applying the equality $\lambda_0 = E_0\Gamma E_0^T$ one obtains that

$$\mathrm{Tr}\left(\lambda_0\frac{\partial G_0(x_*)}{\partial x_i}\right) = \mathrm{Tr}\left(\Gamma E_0^T\frac{\partial G_0(x_*)}{\partial x_i}E_0\right) = \sum_{j,k=1}^{l-r}\Gamma_{jk}e_j^T\frac{\partial G_0(x_*)}{\partial x_i}e_k.$$

Hence and from (67) one gets that

$$\sum_{i,j=1}^{l-r}\sum_{p,q=1}^{l-r}\Gamma_{ij}\Gamma_{pq}\langle v_{ij},v_{pq}\rangle + 2\sum_{i,j=1}^{l-r}\sum_{k=1}^{s}\Gamma_{ij}\mu_k\langle v_{ij},\nabla h_k(x_*)\rangle$$

$$+\sum_{i,j=1}^{s}\mu_i\mu_j\langle\nabla h_i(x_*),\nabla h_j(x_*)\rangle = 0,$$

where v_{jk} are defined as in (66). Denote $z = \sum_{i,j=1}^{l-r}\Gamma_{ij}v_{ij} + \sum_{k=1}^{s}\mu_k\nabla h_k(x_*)$. The equality above implies that $\|z\|^2 = 0$, i.e. $z = 0$ or, equivalently,

$$\sum_{i=1}^{l-r}\Gamma_{ii}v_{ii} + \sum_{1\le i<j\le l-r}2\Gamma_{ij}v_{ij} + \sum_{k=1}^{s}\mu_k\nabla h_k(x_*) = 0.$$

Here we used the facts that the matrix Γ is symmetric and $v_{ij} = v_{ji}$. With the use of the nondegeneracy condition one obtains that $\Gamma = 0$ and $\mu = 0$, i.e. $\lambda = 0$, which implies that the matrix $D_{\lambda\lambda}^2\eta(x_*,\lambda_*)$ is positive definite. $\quad\square$

Now, we can introduce the penalized augmented Lagrangian function for problem (65). Choose $\alpha > 0$ and $\varkappa > 1$, and for any $\lambda = (\lambda_0,\mu) \in Y^*$ define

$$p(x,\lambda) = \frac{a(x)}{1+\mathrm{Tr}(\lambda_0^2)}, \quad q(x,\lambda) = \frac{b(x)}{1+\|\mu\|^2},$$

where

$$a(x) = \alpha - \mathrm{Tr}\left([G_0(x)]_+^2\right)^\varkappa, \quad b(x) = \alpha - \|h(x)\|^2.$$

Denote $\Omega_\alpha = \{x \in \mathbb{R}^d \mid a(x) > 0,\ b(x) > 0\}$. Finally, for any $\lambda = (\lambda_0,\mu) \in Y^*$ define

$$\mathscr{L}_e(x,\lambda,c) = f(x) + \frac{1}{2cp(x,\lambda)}\left(\mathrm{Tr}\left([cG_0(x)+p(x,\lambda)\lambda_0]_+^2\right) - p(x,\lambda)^2\,\mathrm{Tr}(\lambda_0^2)\right)$$

$$+\langle\mu,h(x)\rangle + \frac{c}{2q(x,\lambda)}\|h(x)\|^2 + \eta(x,\lambda), \quad (68)$$

if $x \in \Omega_\alpha$, and $\mathscr{L}_e(x,\lambda,c) = +\infty$, otherwise (see Example 12). Note that the function $\mathscr{L}_e(x,\lambda,c)$ is l.s.c. jointly in (x,λ) on $\mathbb{R}^d \times Y^*$, and continuously

differentiable in (x, λ) on its effective domain $\Omega_\alpha \times Y^*$ by [4], Theorem 4.13. One can also check that the function $\Phi(y, \lambda, c)$ corresponding to penalized augmented Lagrangian function (68) satisfies assumptions $(A2)$, $(A4)_s$, $(A6)_s$, $(A11)$ with $\Phi_0(\lambda) \equiv \lambda$, and $(A12)$. Furthermore, $\Phi(G(x_*), \lambda_*, c) = 0$, if (x_*, λ_*) is a KKT-pair, and arguing in the same way as in ([72], pp. 487–488) one can check that $\Phi(G(x), \lambda, c)$ admits the second order expansion in (x, λ) at every KKT-pair of problem (65).

Applying Lemma 4, Theorem 8 and the localization principle, and arguing in the same way as in the proofs of Theorems 9 and 12 one can easily verify that the following result holds true (see [29] for the detailed proof of this result).

Theorem 13. *Let the functions f, G_0, and h be twice continuously differentiable. Suppose also that every globally optimal solution of problem (65) is nondegenerate, and for any $x_* \in \Omega_*$ a unique KKT-pair (x_*, λ_*) satisfies the second order sufficient optimality condition. Then penalized augmented Lagrangian function (68) is globally exact if and only if there exists $c_0 > 0$ such that the set $\{(x, \lambda) \in \mathbb{R}^d \times Y^* \mid \mathscr{L}_e(x, \lambda, c_0) < f_*\}$ is either bounded or empty. In particular, if the set $\{x \in \mathbb{R}^d \mid f(x) < f_* + \gamma, \, a(x) > 0, \, b(x) > 0\}$ is bounded for some $\gamma > 0$, then penalized augmented Lagrangian function (68) is globally exact in the sense that for any sufficiently large $c > 0$ its points of global minimum in (x, λ) are exactly KKT-pairs of problem (65) corresponding to globally optimal solutions of this problem.*

Remark 28. (i) One can easily construct a globally exact penalized augmented Lagrangian function for problem (65) from augmented Lagrangian from Example 14. Namely, for any $\lambda = (\lambda_0, \mu) \in Y^*$ set

$$\eta(x, \lambda) = \left\| D_x L(x, \lambda_0^2, \mu) \right\|^2 + \mathrm{Tr}(\lambda_0^2 G_0(x)^2),$$

and define

$$\mathscr{L}_e(x, \lambda, c) = f(x) + \langle \mu, h(x) \rangle + \frac{c}{2q(x, \lambda)} \|h(x)\|^2 + \eta(x, \lambda)$$

$$+ \frac{p(x, \lambda)}{c} \left\langle \lambda_0^2, \Psi\left(\frac{c}{p(x, \lambda)} G_0(x) \right) \right\rangle + \frac{p(x, \lambda)}{c} \mathrm{Tr}\left[\Xi\left(\frac{c}{p(x, \lambda)} G_0(x) \right) \right], \quad (69)$$

if $x \notin \Omega_\alpha$, and $\mathscr{L}_e(x, \lambda, c) = +\infty$ otherwise (see Remark 23), where Ψ and Ξ are the same as in Example 14. Then penalized augmented Lagrangian (69) is globally exact provided every globally optimal solution of problem (65) is nondegenerate, satisfies the second order sufficient optimality and strict complementarity conditions, the set $\{x \in \mathbb{R}^d \mid f(x) < f_* + \gamma, \, a(x) > 0, \, b(x) > 0\}$ is bounded, and the function $\psi(t)$ is bounded from below. We do not present the proof of this result here, and leave it to the interested reader.

(ii) Let us note that penalized augmented Lagrangian functions (68) and (69) for problem (65) are completely new. A different exact augmented Lagrangian function for nonlinear semidefinite programming problems was recently introduced in [33]. It should be pointed out that our augmented Lagrangian function is defined via the problem data directly, while the augmented Lagrangian function from [33] depends on a solution of a certain system of linear equations, which makes the computation of the value of this augmented Lagrangian function and its derivatives more expensive. Furthermore, in order to correctly define the

augmented Lagrangian function from [33] one must suppose that *every* feasible point of problem (65) is nondegenerate, which might be a too restrictive assumption for many applications. In contrast, we assume that only globally optimal solutions of problem (65) are nondegerate.

8.5 A Different Approach to Global Exactness

As the examples above show, one must suppose that *both* sufficient optimality conditions and a constraint qualification hold true at globally optimal solutions of the problem $(\mathcal{P})$ in order to guarantee the global exactness of penalized augmented Lagrangian functions with the use of the localization principle and Theorem 8.

Being inspired by the ideas of [22] and the localization principle, we present different necessary and sufficient conditions for the global exactness of $\mathscr{L}_e(x, \lambda, c)$ that are based on the use of constraint qualifications *only*. However, let us point out that this conditions are applicable only in the case when $\mathscr{L}(x, \lambda, c)$ is the Hestenes-Powell-Rockafellar augmented Lagrangian function.

Theorem 14. *Let Y be finite dimensional, $A = X$, $\Lambda = Y^*$, G be continuous, and $\mathscr{L}_e(\cdot, \cdot, c)$ be l.s.c. for all $c > 0$. Suppose also that assumptions $(A2)$, $(A4)_s$, $(A6)_s$ and $(A12)$ are satisfied, and*

$$\Phi(G(x_*), \lambda_*, c) = 0 \quad \forall c > 0 \quad \forall (x_*, \lambda_*) \in \Omega_* \times Y^*\colon \eta(x_*, \lambda_*) = 0. \tag{70}$$

Suppose, finally, that for any $(x_*, \lambda_*) \in \Omega_* \times Y^*$ such that $\eta(x_*, \lambda_*) = 0$ there exist a neighbourhood U of (x_*, λ_*) and $\overline{c} > 0$ such that the function $\mathscr{L}_e(\cdot, \cdot, c)$ is Gâteaux differentiable in U for all $c \geq \overline{c}$, and

$$\Big((x, \lambda) \in U \wedge D_{(x,\lambda)} \mathscr{L}_e(x, \lambda, c) = 0 \Big) \implies \Big(x \in \Omega \wedge \Phi(G(x), \lambda, c) = 0 \Big). \tag{71}$$

Then $\mathscr{L}_e(x, \lambda, c)$ is globally exact (with respect to the function η) if and only if there exists $(x_0, \lambda_0) \in \Omega_* \times Y^*$ such that $\eta(x_0, \lambda_0) = 0$, and there exists $c_0 > 0$ such that the set $S_e(c_0)$ is either bounded or empty.

Proof. The "only if" part of the theorem is proved in the same way as the "only if" part of Theorem 6. Therefore, let us prove the "if" part.

Choose an increasing unbounded sequence $\{c_n\} \subset [c_0, +\infty)$. Assumption $(A4)_s$ implies that $S_e(c_n) \subseteq S_e(c_0)$ for all $n \in \mathbb{N}$. If $S_e(c_n) = \emptyset$ for some $n \in \mathbb{N}$, then with the use of (70) one can easily verify that $\mathscr{L}_e(x, \lambda, c)$ is globally exact. Therefore, let us suppose that $S_e(c_n) \neq \emptyset$ for all $n \in \mathbb{N}$.

Taking into account the facts that $S_e(c_0)$ is either bounded or empty, $S_e(c_n) \subseteq S_e(c_0)$, and $\mathscr{L}_e(\cdot, \cdot, c)$ is l.s.c. one obtains that for any $n \in \mathbb{N}$ the function $\mathscr{L}_e(\cdot, \cdot, c_n)$ attains a global minimum at a point (x_n, λ_n), and the sequence $\{(x_n, \lambda_n)\}$ is bounded. Since both X and Y are finite dimensional, without loss of generality one can suppose that $\{(x_n, \lambda_n)\}$ converges to a point (x_*, λ_*). By Lemma 2 one has $x_* \in \Omega_*$ and $\eta(x_*, \lambda_*) = 0$. Therefore there exist a neighbourhood U of (x_*, λ_*) and $\overline{c} > 0$ such that (71) holds true.

From the facts that (x_*, λ_*) is a limit point of $\{(x_n, \lambda_n)\}$, and $\{c_n\}$ is an increasing unbounded sequence it follows that there exists $n \in \mathbb{N}$ such that $(x_n, \lambda_n) \in U$ and $c_n \geq \overline{c}$. Applying the first order necessary optimality condition one obtains that $D_{(x,\lambda)} \mathscr{L}_e(x_n, \lambda_n, c_n) = 0$, which with the use of (71) implies

that x_n is a feasible point of $(\mathcal{P})$, and $\mathscr{L}_e(x_n, \lambda_n, c_n) = f(x_n) + \eta(x_n, \lambda_n) \geq f(x_n) \geq f_*$. Thus, $S_e(c_n) = \emptyset$, which contradicts our assumption. $\qquad\square$

Note that condition (71) is satisfied for augmented Lagrangian (56) (in the case when there are no equality constraints) by [22], Proposition 4.3, provided LICQ holds at every globally optimal solution. Therefore with the use of the theorem above one can obtain the following result that improves Theorem 5.4 from [22], since we do note assume that LICQ holds true at every feasible point of problem (55), and obtain *necessary and sufficient* conditions for the global exactness of augmented Lagrangian (56) for problem (55).

Theorem 15. *Suppose that $J = \emptyset$. Let f and g_i, $i \in I$, be twice continuously differentiable, and let LICQ hold true at every globally optimal solution of problem (55). Then penalized augmented Lagrangian function (56) is globally exact if and only if there exists $c_0 > 0$ such that the set $S_e(c_0)$ is either bounded or empty. In particular, if there exists $\gamma > 0$ such that the set $\Omega(\gamma, \alpha)$ is bounded, then augmented Lagrangian (56) is globally exact.*

Remark 29. (i) It is natural to assume that Proposition 4.3 from [22] (and, thus, the theorem above) can be extended to the case of other penalized augmented Lagrangian functions (in particular, to the case of augmented Lagrangians (63) and (68)). However, we do not discuss the possibility of such an extension here, and pose it as an interesting open problem.
(ii) Let us note that with the use of Theorem 14 one can easily obtain *necessary and sufficient* conditions for the global exactness of the augmented Lagrangian function for nonlinear semidefinite programming problems from [33], which strengthen Theorem 4.8 from this paper.

9   Conclusions

In this article we developed a general theory of augmented Lagrangian functions for cone constrained optimization problems. Let us briefly discuss some conclusions that can be drawn from this theory.

Note that there are two main classes of augmented Lagrangian functions for cone constrained optimization problem. The first class consists of the well-known Hestenes-Powell-Rockafellar augmented Lagrangian for mathematical programming problems and its various modifications and extensions to the case of other cone constrained optimization problem (Examples 1–4, 10, 12 and 15). The second class of augmented Lagrangians consists of the exponential penalty functions and its numerous modifications (Examples 5–8, 11, 13, 14 and 16). It is natural to refer to the augmented Lagrangians from this class as *nonlinear rescaling augmented Lagrangians*, since, in essence, all these augmented Lagrangians are constructed as the standard Lagrangian function for the problem with rescaled constraints (see [65]). Finally, there is also He-Wu-Meng's augmented Lagrangian (Example 9) that does not fall into either of those classes.

Nonlinear rescaling augmented Lagrangians are smoother than Hestenes-Power-Rockafellar-type augmented Lagrangians. However, in order to guarantee the existence of global saddle points (or local/global exactness) of nonlinear rescaling augmented Lagrangians one must impose the strict complementarity condition, which is not necessary in the case of the Hestenes-Power-Rockafellar-type augmented Lagrangians and He-Wu-Meng's augmented Lagrangian (see

Remarks 15 and 23) with the only exclusion being the cubic augmented Lagrangian (Example 3). Therefore, it seems that in order to avoid the strict complementarity condition one must consider augmented Lagrangians that are not twice continuously differentiable. Furthermore, as it was noted above (see Remark 19), all existing augmented Lagrangian functions are not suitable for handling semi-infinite programming problems, and a new approach to the construction of augmented Lagrangians for these problem is needed.

From the theoretical point of view, the augmented Lagrangian functions that are least suitable for the study of the existence of global saddle points (and global exactness) are those augmented Lagrangians that do not contain penalty terms (namely, the exponential penalty function and the p-th power augmented Lagrangian; see Example 18 and Remark 14). However, numerical methods based on these augmented Lagrangians sometimes work well for convex (even infinite dimensional, see [61]) problems.

Finally, let us note that we did not discuss augmented Lagrangian methods for cone constrained optimization problems as well as numerical methods based on the use of globally exact augmented Lagrangian functions for these problems. It seems that under some additional assumptions one might extend the existing augmented Lagrangian methods to the case of the augmented Lagrangian $\mathscr{L}(x, \lambda, c)$ for problem $(\mathcal{P})$, thus developing a general theory of augmented Lagrangian methods for cone constrained optimization problems. We leave the development of this theory as an open problem for future research.

References

[1] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York, 1982.

[2] E. G. Birgin and J. M. Martinez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. SIAM, Philadelphia, 2014.

[3] J. F. Bonnans and C. H. Ramírez. Perturbation analysis of second-order cone programming problems. *Math. Program.*, 104:205–227, 2005.

[4] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Science+Business Media, New York, 2000.

[5] R. S. Burachik, A. N. Iusem, and J. G. Melo. Duality and exact penalization for general augmented Lagrangians. *J. Optim. Theory Appl.*, 147:125–140, 2010.

[6] R. S. Burachik, X. Q. Yang, and Y. Y. Zhou. Existence of augmented Lagrange multipliers for semi-infinite programming problems. *J. Optim. Theory Appl.*, 173:471–503, 2017.

[7] J.-X. Chen, X. Chen, and P. Tseng. Analysis of nonsmooth vector-valued functions associated with second-order cones. *Math. Program., Ser. B.*, 101:95–117, 2004.

[8] X. Chen, H. Qi, and P. Tseng. Analysis of nonsmooth symmetric-matrix-valued functions with applications to semidefinite complementarity problems. *SIAM J. Optim.*, 13:960–985, 2003.

[9] V. A. Daugavet. Alternance properties of the solutions of non-linear min-imax problems with non-linear constraints. *USSR Comput. Math. Math. Phys.*, 16:236–241, 1976.

[10] V. A. Daugavet and V. N. Malozemov. Alternance properties of solutions of nonlinear minimax problems with nonconvex constraints. *Soviet Math. Dokl.*, 16:1474–1476, 1975.

[11] V. A. Daugavet and V. N. Malozemov. Quadratic rate of convergence of a linearization method for solving discrete minimax problems. *USSR Comput. Math. Math. Phys.*, 21:19–28, 1981.

[12] V. F. Demyanov and V. N. Malozemov. Alternance form of optimality conditions in the finite-dimensional space. In V. F. Demyanov, P. M. Pardalos, and M. Batsyn, editors, *Constructive Nonsmooth Analysis and Related Topics*, pages 185–205. Springer, New York, 2014.

[13] V. F. Demyanov and V. N. Malozemov. Optimality conditions in terms of alternance: two approaches. *J. Optim. Theory Appl.*, 162:805–820, 2014.

[14] G. Di Pillo and L. Grippo. A new class of augmented Lagrangians in nonlinear programming. *SIAM J. Control Optim.*, 17:618–628, 1979.

[15] G. Di Pillo and L. Grippo. A new augmented Lagrangian function for in-equality constraints in nonlinear programming problems. *J. Optim. Theory Appl.*, 36:495–519, 1982.

[16] G. Di Pillo, L. Grippo, and F. Lampariello. A method for solving equal-ity constrained optimization problems by unconstrained minimization. In K. Iracki, K. Malanowski, and S. Walukiewicz, editors, *Optimization tech-niques: proceedings of the 9th IFIP Conference on Optimization Tech-niques*, pages 96–105. Springer-Verlag, Berlin, Heidelberg, 1980.

[17] G. Di Pillo, L. Grippo, and S. Lucidi. A smooth method for the finite minimax problem. *Math. Program.*, 60:187–214, 1993.

[18] G. Di Pillo, G. Liuzzi, S. Lucidi, and L. Palagi. An exact augmented La-grangian function for nonlinear programming with two-sided constraints. *Comput. Optim. Appl.*, 25:57–83, 2003.

[19] G. Di Pillo, G. Liuzzi, S. Lucidi, and L. Palagi. Fruitful uses of smooth exact merit functions in constrained optimization. In G. Di Pillo and A. Murli, editors, *High Performance Algorithms and Software for Nonlin-ear Optimization*, pages 201–225. Kluwer Academic Publishers, Dordrecht, 2003.

[20] G. Di Pillo, G. Liuzzi, S. Lucidi, and L. Palagi. A truncated Newton method in an augmented Lagrangian framework for nonlinear program-ming. *Comput. Optim. Appl.*, 45:311–352, 2010.

[21] G. Di Pillo and S. Lucidi. On exact augmented Lagrangian functions in nonlinear programming. In G. Di Pillo and F. Giannessi, editors, *Non-linear Optimization and Applications*, pages 85–100. Plenum Press, New York, 1996.

[22] G. Di Pillo and S. Lucidi. An augmented Lagrangian function with improved exactness properties. *SIAM J. Optim.*, 12:376–406, 2001.

[23] G. Di Pillo, S. Lucidi, and L. Palagi. An exact penalty-Lagrangian approach for a class of constrained optimization problems with bounded variables. *Optim.*, 28:129–148, 1993.

[24] G. Di Pillo, S. Lucidi, and L. Palagi. A truncated Newton method for constrained optimization. In G. Di Pillo and F. Giannessi, editors, *Nonlinear Optimization and Related Topics*, pages 79–103. Kluwer Academic Publishers, Dordrecht, 2000.

[25] G. Di Pillo, S. Lucidi, and L. Palagi. Convergence to second-order stationary points of a primal-dual algorithm model for nonlinear programming. *Math. Oper. Res.*, 30:897–915, 2005.

[26] G. Di Pillo, G. Luizzi, and S. Lucidi. An exact penalty-Lagrangian approach for large-scale nonlinear programming. *Optim.*, 60:223–252, 2011.

[27] M. V. Dolgopolik. A unifying theory of exactness of linear penalty functions. *Optim.*, 65:1167–1202, 2016.

[28] M. V. Dolgopolik. Existence of augmented Lagrange multipliers: reduction to exact penalty functions and localization principle. *Math. Program.*, 166:297–326, 2017.

[29] M. V. Dolgopolik. A unified approach to the global exactness of penalty and augmented Lagrangian functions II: extended exactness. *arXiv preprint: 1710.01961*, 2017.

[30] M. V. Dolgopolik. A unifying theory of exactness of linear penalty functions II: parametric penalty functions. *Optim.*, 66:1577–1622, 2017.

[31] X. Du, Y. Liang, and L. Zhang. Further study on a class of augmented Lagrangians of Di pillo and Grippo in nonlinear programming. *J. Shanghai Univ. (Engl. Ed.)*, 10:293–298, 2006.

[32] X. Du, L. Zhang, and Y. Gao. A class of augmented Lagrangians for equality constraints in nonlinear programming problems. *Appl. Math. Comput.*, 172:644–663, 2006.

[33] E. H. Fukuda and B. F. Lourenco. Exact augmented Lagrangian functions for nonlinear semidefinite programming. *Comput. Optim. Appl.*, 71:457–482, 2017.

[34] E. H. Fukuda, P. J. S. Silva, and M. Fukushima. Differentiable exact penalty functions for nonlinear second-order cone programs. *SIAM J. Optim.*, 22:1607–1633, 2012.

[35] F. Giannessi. On the theory of Lagrangian duality. *Optim. Lett.*, 1:9–20, 2007.

[36] S. He, L. Wu, and H. Meng. A nonlinear Lagrangian for constrained optimization problems. *J. Appl. Math. Comput.*, 38:669–685, 2012.

[37] R. A. Horn and C. R. Johnson. *Topics in matrix analysis.* Cambridge University Press, Cambridge, 1991.

[38] X. X. Huang and X. Q. Yang. A unified augmented Lagrangian approach to duality and exact penalization. *Math. Oper. Res.*, 28:533–552, 2003.

[39] X. X. Huang and X. Q. Yang. Further study on augmented Lagrangian duality theory. *J. Glob. Optim.*, 31:193–210, 2005.

[40] X. X. Huang, X. Q. Yang, and K. L. Teo. Augmented Lagrangian and nonlinear semidefinite programs. In F. Giannessi and A. Maugeri, editors, *Variational Analysis and Applications*, pages 513–529. Springer, Boston, 2005.

[41] N. Q. Huy and D. S. Kim. Stability and augmented Lagrangian duality in nonconvex semi-infinite programming. *Nonlinear Anal.*, 75:163–176, 2012.

[42] K. C. Kiwiel. On the twice differentiable cubic augmented Lagrangian. *J. Optim. Theory Appl.*, 88:233–236, 1996.

[43] D. Li. Zero duality gap for a class of nonconvex optimization problems. *J. Optim. Theory Appl.*, 85:309–324, 1995.

[44] D. Li. Saddle-point generation in nonlinear nonconvex optimization. *Nonlinear Anal.*, 30:4339–4344, 1997.

[45] D. Li and X. L. Sun. Convexification and existence of saddle point in a pth-power reformulation for nonconvex constrained optimization. *Nonlinear Anal.*, 47:5611–5622, 2001.

[46] D. Li and X. L. Sun. Existence of a saddle point in nonconvex constrained optimization. *J. Glob. Optim.*, 21:39–50, 2001.

[47] J. Li, S. Q. Feng, and Z. Zhang. A unified approach for constrained extremum problems: image space analysis. *J. Optim. Theory Appl.*, 159:69–92, 2013.

[48] Y. Li and L. Zhang. A new nonlinear Lagrangian method for nonconvex semidefinite programming. *J. Appl. Anal.*, 15:149–172, 2009.

[49] Q. Liu, W. M. Tang, and X. M. Yang. Properties of saddle points for generalized augmented Lagrangian. *Math. Meth. Oper. Res.*, 69:111–124, 2009.

[50] Q. Liu and X. Yang. Zero duality and saddle points of a class of augmented Lagrangian functions in constrained non-convex optimization. *Optim.*, 57:655–667, 2008.

[51] Y. J. Liu and L. W. Zhang. Convergence analysis of the augmented Lagrangian method for nonlinear second-order cone optimization problems. *Nonlinear Anal.: Theory, Methods, Appl.*, 67:1359–1373, 2007.

[52] Y. J. Liu and L. W. Zhang. Convergence of the augmented Lagrangian method for nonlinear optimization problems over second-order cones. *J. Optim. Theory Appl.*, 139:557–575, 2008.

[53] S. Lucidi. New results on a class of exact augmented Lagrangians. *J. Optim. Theory Appl.*, 58:259–282, 1988.

[54] H. Luo, H. Wu, and J. Liu. Some results on augmented Lagrangians in constrained global optimization via image space analysis. *J. Optim. Theory Appl.*, 159:360–385, 2013.

[55] H. Luo, H. Wu, and J. Liu. On saddle points in semidefinite optimization via separation scheme. *J. Optim. Theory Appl.*, 165:113–150, 2015.

[56] H. Z. Luo, G. Mastroeni, and H. X. Wu. Separation approach for augmented Lagrangians in constrained nonconvex optimization. *J. Optim. Theory Appl.*, 144:275–290, 2010.

[57] H. Z. Luo, H. X. Wu, and G. T. Chen. On the convergence of augmented Lagrangian methods for nonlinear semidefinite programming. *J. Glob. Optim.*, 54:599–618, 2012.

[58] V. N. Malozemov and A. B. Pevnyi. Alternation properties of solutions of nonlinear minimax problems. *Soviet Math. Dokl.*, 14:1303–1306, 1973.

[59] O. L. Mangasarian. Unconstrained Lagrangians in nonlinear programming. *SIAM J. Control*, 12:772–791, 1975.

[60] D. Noll. Local convergence of an augmented Lagrangian method for matrix inequality constrained programming. *Optim. Methods Softw.*, 22:777–802, 2007.

[61] P. Pedregal. Direct numerical algorithm for constrained variational problems. *Numer. Funct. Anal. Optim.*, 38:486–506, 2017.

[62] E. Polak and J. O. Royset. On the use of augmented Lagrangians in the solution of generalized semi-infinite min-max problems. *Comput. Optim. Appl.*, 31:173–192, 2005.

[63] R. Polyak. Modified barrier functions: theory and methods. *Math. Program.*, 54:177–222, 1992.

[64] R. A. Polyak. Log-sigmoid multipliers method in constrained optimization. *Ann. Oper. Res.*, 101:427–460, 2001.

[65] R. A. Polyak. Nonlinear rescaling vs. smoothing technique in convex optimization. *Math. Program.*, 92:197–235, 2002.

[66] R. T. Rockafellar. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM J. Control Optim.*, 12:268–285, 1974.

[67] R. T. Rockafellar. Lagrange multipliers and optimality. *SIAM Review*, 35:183–238, 1993.

[68] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis.* Springer-Verlar, Berlin, 1998.

[69] A. M. Rubinov and X. Q. Yang. *Lagrange-Type Functions in Constrained Non-Convex Optimization.* Kluwer Academic Publishers, Dordrecht, 2003.

[70] J.-J. Rückmann and A. Shapiro. Augmented Lagrangians in semi-infinite programming. *Math. Program., Ser. B.*, 116:499–512, 2009.

[71] 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.

[72] A. Shapiro and J. Sun. Some properties of the augmented Lagrangian in cone constrained optimization. *Math. Oper. Res.*, 29:479–491, 2004.

[73] T. Q. Son, D. D. S. Kim, and N. N. Tam. Weak stability and strong duality of a class of nonconvex infinite programs via augmented Lagrangian. *J. Glob. Optim.*, 53:165–184, 2012.

[74] M. Stingl. *On the solution of nonlinear semidefinite programs by augmented Lagrangian methods.* PhD thesis, Institute of Applied Mathematics II, Friedrech-Alexander University of Erlangen-Nuremberg, Erlangen, Germany, 2006.

[75] D. Sun and J. Sun. Löwner's operator and spectral functions in euclidean jordan algebras. *Math. Oper. Res.*, 33:421–445, 2008.

[76] D. Sun, J. Sun, and L. Zhang. The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. *Math. Program.*, 114:349–391, 2008.

[77] J. Sun. On methods for solving nonlinear semidefinite optimization problems. *Numer. Algebra Control Optim.*, 1:1–14, 2011.

[78] J. Sun, L. W. Zhang, and Y. Wu. Properties of the augmented Lagrangian in nonlinear semidefinite optimization. *J. Optim. Theory Appl.*, 129:437–456, 2006.

[79] X. L. Sun, D. Li, and K. I. M. McKinnon. On saddle points of augmented Lagrangians for constrained nonconvex optimization. *SIAM J. Optim.*, 15:1128–1146, 2005.

[80] P. Tseng and D. P. Bertsekas. On the convergence of the exponential multiplier method for convex programming. *Math. Program.*, 60:1–19, 1993.

[81] C. Wang, Q. Liu, and B. Qu. Global saddle points of nonlinear augmented Lagrangian functions. *J. Glob. Optim.*, 68:125–146, 2017.

[82] C. Wang, J. Zhou, and X. Xu. Saddle points theory of two classes of augmented Lagrangians and its applications to generalized semi-infinite programming. *Appl. Math. Optim.*, 59:413–434, 2009.

[83] C. Y. Wang and D. Li. Unified theory of augmented Lagrangian methods for constrained global optimization. *J. Glob. Optim.*, 44:433–458, 2009.

[84] C. Y. Wang, X. Yang, and X. M. Yang. Unified nonlinear Lagrangian approach to duality and optimal paths. *J. Optim. Theory Appl.*, 135:85–100, 2007.

[85] C. Y. Wang, X. Q. Yang, and X. M. Yang. Nonlinear augmented Lagrangian and duality theory. *Math. Oper. Res.*, 38:740–760, 2012.

[86] Z. Wen, D. Goldfarb, and W. Yin. Alternating direction augmented Lagrangian methods for semidefinite programming. *Math. Program. Comput.*, 2:203–230, 2010.

[87] H. Wu, H. Luo, X. Ding, and G. Chen. Global convergence of modified augmented Lagrangian methods for nonlinear semidefintie programming. *Comput. Optim. Appl.*, 56:531–558, 2013.

[88] H. X. Wu and H. Z. Luo. A note on the existence of saddle points of p-th power Lagrangian for constrained nonconvex optimization. *Optim.*, 61:1231–1245, 2012.

[89] H. X. Wu and H. Z. Luo. Saddle points of general augmented Lagrangians for constrained nonconvex optimization. *J. Glob. Optim.*, 53:683–697, 2012.

[90] H. X. Wu, H. Z. Luo, and J. F. Yang. Nonlinear separation approach for the augmented Lagrangian in nonlinear semidefinite programming. *J. Glob. Optim.*, 59:695–727, 2014.

[91] Z. K. Xu. Local saddle points and convexification for nonconvex optimization problems. *J. Optim. Theory Appl.*, 94:739–746, 1997.

[92] H. Yamashita and H. Yabe. A survey of numerical methods for nonlinear semidefinite programming. *J. Oper. Res. Soc. Jpn.*, 58:24–60, 2015.

[93] L. Zhang, J. Gu, and X. Xiao. A class of nonlinear Lagrangians for nonconvex second order cone programming. *Comput. Optim. Appl.*, 49:61–99, 2011.

[94] L. Zhang, Y. Li, and J. Wu. Nonlinear rescaling Lagrangians for nonconvex semidefinite programming. *Optim.*, 63:899–920, 2014.

[95] W. Zhao, J. Zhang, and J. Zhou. Existence of local saddle points for a new augmented Lagrangian function. *Math. Probl. Eng.*, 2010:1–13, 2010.

[96] X. Y. Zhao, D. Sun, and K.-C. Toh. A Newton-CG augmented Lagrangian method for semidefinite programming. *SIAM J. Optim.*, 20:1737–1765, 2010.

[97] J. Zhou and J. S. Chen. On the existence of saddle points for nonlinear second-order cone programming problems. *J. Glob. Optim.*, 62:459–480, 2015.

[98] J. Zhou, N. Xiu, and C. Wang. Saddle point and exact penalty representation for generalized proximal Lagrangians. *J. Glob. Optim.*, 56:669–687, 2012.

[99] Y. Y. Zhou, J. C. Zhou, and X. Q. Yang. Existence of augmented Lagrange multipliers for cone constrained optimization problems. *J. Glob. Optim.*, 58:243–260, 2014.

[100] S. K. Zhu and S. J. Li. Unified duality theory for constrained extremum problems. Part I: image space analysis. *J. Optim. Theory Appl.*, 161:738–762, 2014.

[101] S. K. Zhu and S. J. Li. Unified duality theory for constrained extremum problems. Part II: special duality schemes. *J. Optim. Theory Appl.*, 161:763–782, 2014.