A Unified Approach to the Global Exactness of Penalty and Augmented Lagrangian Functions I: Parametric Exactness

M. V. Dolgopolik

Abstract In this two-part study, we develop a unified approach to the analysis of the global exactness of various penalty and augmented Lagrangian functions for constrained optimization problems in finite-dimensional spaces. This approach allows one to verify in a simple and straightforward manner whether a given penalty/augmented Lagrangian function is exact, i.e., whether the problem of unconstrained minimization of this function is equivalent (in some sense) to the original constrained problem, provided the penalty parameter is sufficiently large. Our approach is based on the so-called localization principle that reduces the study of global exactness to a local analysis of a chosen merit function near globally optimal solutions. In turn, such local analysis can be performed with the use of optimality conditions and constraint qualifications. In the first paper, we introduce the concept of global parametric exactness and derive the localization principle in the parametric form. With the use of this version of the localization principle, we recover existing simple, necessary, and sufficient conditions for the global exactness of linear penalty functions and for the existence of augmented Lagrange multipliers of Rockafellar–Wets’ augmented Lagrangian. We also present completely new necessary and sufficient conditions for the global exactness of general nonlinear penalty functions and for the global exactness of a continuously differentiable penalty function for nonlinear second-order cone programming problems. We briefly discuss how one can construct a continuously differentiable exact penalty function for nonlinear semidefinite programming problems as well.

Keywords Penalty function · Augmented Lagrangian function · Exactness · Localization principle · Semidefinite programming
1 Introduction

One of the main approaches to the solution of a constrained optimization problem consists in the reduction of this problem to an unconstrained one (or a sequence of unconstrained problems) with the use of *merit* (or *auxiliary*) functions. Such merit functions are usually defined as a certain convolution of the objective function and constraints, and they almost always include the penalty parameter that must be properly chosen for the reduction to work. This approach led to the development of various penalty and barrier methods [1,2], primal–dual methods based on the use of augmented Lagrangians [3], and many other methods of constrained optimization.

There exist numerous results on the duality theory for various merit functions, such as penalty and augmented Lagrangian functions. A modern general formulation of the augmented Lagrangian duality for nonconvex problems based on a geometric interpretation of the augmented Lagrangian in terms of subgradients of the optimal value function was proposed by Rockafellar and Wets in [4] and further developed in [5–7]. Let us also mention several extensions [8,9] of this augmented Lagrangian duality theory aiming at including some other augmented Lagrangian and penalty functions into the unified framework proposed in [4]. A general duality theory for *nonlinear* Lagrangian and penalty functions was developed in [10,11]. Another general approach to the study of duality based on the image space analysis was systematically studied in [12–15].

In contrast to duality theory, few attempts [16–18] have been made to develop a general theory of *global exactness* of merit functions, despite the abundance of particular results on the exactness of various penalty/augmented Lagrangian functions. Furthermore, the existing general results on global exactness are unsatisfactory, since they are very restrictive and cannot be applied to many particular cases.

Recall that a penalty function is called exact iff its points of global minimum coincide with globally optimal solutions of the constrained optimization problem under consideration. The concept of exactness of a *linear penalty function* was introduced by Eremin [19] and Zangwill [20] in the mid-1960s and was further investigated by many researchers (see [16,21–25] and the references therein). A class of *continuously differentiable* exact penalty functions was introduced by Fletcher [26] in 1970. Fletcher’s penalty function was modified and investigated in detail in [26–30]. Di Pillo and Grippo proposed to consider an *exact augmented Lagrangian function* [31] in 1979. This class of augmented Lagrangian functions was studied and applied to various optimization problems in [32–34], while a general theory of globally exact augmented Lagrangian functions was developed by the author in [35]. The theory of *nonlinear exact penalty functions* was developed by Rubinov and his colleagues [10,36–39] in the late 1990s and the early 2000s. A new class of exact penalty functions was introduced by Huyer and Neumaier [40] in 2003. Later on, this class of penalty functions was modified and thoroughly investigated in [41,42]. Finally, there is also a problem of the existence of *augmented Lagrange multipliers*, which can be viewed as the study of the global exactness of Rockafellar–Wets’ augmented Lagrangian func-
Various results on the existence of augmented Lagrange multipliers were obtained in [43–45].

The analysis of the proofs of the main results of the aforementioned papers indicates that the underlying ideas of these papers largely overlap. Our main goal is to unveil the core idea behind these results and present a general theory of the global exactness of penalty and augmented Lagrangian functions for finite-dimensional constrained optimization problems that can be applied to all existing penalty and augmented Lagrangian functions. The central result of our theory is the so-called localization principle. This principle allows one to reduce the study of the global exactness of a given merit function to a local analysis of the behavior of this function near globally optimal solutions of the original constrained problem. In turn, such local analysis can be usually performed with the use of sufficient optimality conditions and/or constraint qualifications. Thus, the localization principle furnish one with a simple technique for proving the global exactness of almost any merit function with the use of the standard tools of constrained optimization (namely, constraint qualifications and optimality conditions). The localization principle was first derived by the author for linear penalty functions in [25] and was further extended to other penalty and augmented Lagrangian functions in [35,42,45].

In order to include almost all imaginable penalty and augmented Lagrangian functions into the general theory, we introduce and study the concept of global exactness for an arbitrary function depending on the primal variables, the penalty parameter, and some additional parameters and do not impose any assumptions on the structure of this function. Instead, natural assumptions on the behavior of this function arise within the localization principle as necessary and sufficient conditions for the global exactness. Being inspired by the ideas of the image space analysis [12–15], we call this function a separating function for a constrained optimization problem.

Finally, let us note that there are several possible ways to introduce the concept of the global exactness of a merit function. Each part of this two-part study is devoted to the analysis of one of the possible approaches to the definition of global exactness. In this paper, we study the so-called global parametric exactness, which naturally arises during the study of various exact penalty functions and augmented Lagrangian multipliers.

The paper is organized as follows. In Sect. 2, we introduce the definition of global parametric exactness and derive the localization principle in the parametric form. This version of localization principle is applied to the study of the global exactness of several penalty and augmented Lagrangian functions in Sect. 3. Due to the page limitations, we omit many details and proofs of some results. For a much more detailed exposition of the subject containing the proofs of all main results as well as a significantly larger list of references, see the preprint of this paper [46].

## 2 A General Theory of Parametric Exactness

Let $X$ be a finite-dimensional normed space, and $M, A \subset X$ be nonempty sets. Throughout this article, we study the following optimization problem
\[
\min f(x) \quad \text{subject to} \quad x \in M, \quad x \in A,
\]

where \( f : X \to \mathbb{R} \cup \{+\infty\} \) is a given function. Denote by \( \Omega = M \cap A \) the set of feasible points of this problem. From this point onward, we suppose that there exists \( x \in \Omega \) such that \( f(x) < +\infty \), and that there exists a globally optimal solution of \((P)\).

Our aim is to somehow “get rid” of the constraint \( x \in M \) in the problem \((P)\) with the use of an auxiliary function \( F(\cdot) \). Namely, we want to construct an auxiliary function \( F(\cdot) \) such that globally optimal solutions of the problem \((P)\) can be easily recovered from points of global minimum of \( F(\cdot) \) on the set \( A \). To be more precise, our aim is to develop a general theory of such auxiliary functions. It should be underlined that only the constraint \( x \in M \) is incorporated into an auxiliary function \( F(\cdot) \), while the constraint \( x \in A \) must be taken into account explicitly. Often, the set \( A \) represents “simple” constrains such as bound or linear ones.

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

The function \( F \) can be, for instance, a penalty function with \( \Lambda \) being the empty set or an augmented Lagrangian function with \( \lambda \) being a Lagrange multiplier. However, in order not to restrict ourselves to any specific case, we call \( F(x, \lambda, c) \) a separating function for the problem \((P)\).

In the first part of our study, we consider the simplest case when one minimizes the function \( F(x, \lambda, c) \) with respect to \( x \) and views \( \lambda \) as a tuning parameter. Let us introduce the formal definition of exactness of the function \( F(x, \lambda, c) \) in this case. Denote by \( \Omega^* \) the set of globally optimal solutions of problem \((P)\).

**Definition 2.1** The separating function \( F(x, \lambda, c) \) is said to be globally parametrically exact iff there exist \( \lambda^* \in \Lambda \) and \( c^* > 0 \) such that for any \( c \geq c^* \), one has \( \arg \min_{x \in A} F(x, \lambda^*, c) = \Omega^* \). The greatest lower bound of all such \( c^* > 0 \) is called the least exact penalty parameter of the function \( F(x, \lambda^*, c) \) and is denoted by \( c^*(\lambda^*) \), while \( \lambda^* \) is called an exact tuning parameter.

Thus, if \( F(x, \lambda, c) \) is globally parametrically exact and an exact tuning parameter \( \lambda^* \) is known, then one can choose sufficiently large \( c > 0 \) and minimize the function \( F(\cdot, \lambda^*, c) \) over the set \( A \) in order to find globally optimal solutions of the problem \((P)\). In other words, if the function \( F(x, \lambda, c) \) is globally exact, then one can remove the constraint \( x \in M \) with the use of the function \( F(x, \lambda, c) \) without loosing any information about globally optimal solutions of the problem \((P)\).

Our main goal is to demonstrate that the study of the global parametric exactness of the separating function \( F(x, \lambda, c) \) can be easily reduced to the study of a local behavior of \( F(x, \lambda, c) \) near globally optimal solutions of the problem \((P)\). This reduction procedure is called the localization principle.

At first, let us describe a desired local behavior of the function \( F(x, \lambda, c) \) near optimal solutions.

**Definition 2.2** Let \( x^* \) be a locally optimal solution of the problem \((P)\). The separating function \( F(x, \lambda, c) \) is called locally parametrically exact at \( x^* \) iff there exist \( \lambda^* \in \Lambda \),
c^* > 0 and a neighborhood U of x^* such that for any c ≥ c^* and x ∈ U ∩ A, one has \(F(x, \lambda^*, c) ≥ F(x^*, \lambda^*, c)\). The greatest lower bound of all such c^* > 0 is called the least exact penalty parameter of the function \(F(x, \lambda^*, c)\) at \(x^*\) and is denoted by \(c^*(x^*, \lambda^*)\), while \(\lambda^*\) is called an exact tuning parameter at \(x^*\).

Thus, \(F(x, \lambda, c)\) is locally parametrically exact at a point \(x^*\) with an exact tuning parameter \(\lambda^*\) iff there exists \(c^* > 0\) such that \(x^*\) is a local (uniformly with respect to \(c \in [c^*, +∞)\)) minimizer of the function \(F(\cdot, \lambda^*, c)\) on the set \(A\). Observe also that if the function \(F(x, \lambda, c)\) is nondecreasing in \(c\), then \(F(x, \lambda, c)\) is locally parametrically exact at \(x^*\) with an exact tuning parameter \(\lambda^*\) iff there exists \(c^*\) such that \(x^*\) is a local minimizer of \(F(\cdot, \lambda^*, c)\) on \(A\).

Recall that \(c > 0\) in \(F(x, \lambda, c)\) is called the penalty parameter; however, a connection of the parameter \(c\) with penalization is unclear from the definition of the function \(F(x, \lambda, c)\). We need the following definition in order to clarify this connection.

**Definition 2.3** Let \(\lambda^* \in \Lambda\) be fixed. One says that \(F(x, \lambda, c)\) is a penalty-type separating function for \(\lambda = \lambda^*\) iff there exists \(c_0 > 0\) such that if
1. \(\{c_n\} \subset [c_0, +∞)\) is an increasing and unbounded sequence;
2. \(x_n \in \text{arg min}_{x \in A} F(x, \lambda^*, c_n), n \in \mathbb{N};\)
3. \(x^*\) is a cluster point of the sequence \(\{x_n\},\)

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

Roughly speaking, \(F(x, \lambda, c)\) is a penalty-type separating function for \(\lambda = \lambda^*\) iff global minimizers of \(F(\cdot, \lambda^*, c)\) on the set \(A\) tend to globally optimal solutions of the problem \((\mathcal{P})\) as \(c \to +∞\). Thus, if the separating function \(F(x, \lambda, c)\) is of penalty type, then \(c\) plays the role of the penalty parameter, since the increase in \(c\) forces global minimizers of \(F(\cdot, \lambda^*, c)\) to get closer to the feasible set of the problem \((\mathcal{P})\).

Note that if the function \(F(\cdot, \lambda^*, c)\) does not attain a global minimum on the set \(A\) for any \(c\) greater than some \(c_0 > 0\), then, formally, \(F(x, \lambda, c)\) is a penalty-type separating function for \(\lambda = \lambda^*\). Similarly, if all sequences \(\{x_n\}\), such that \(x_n \in \text{arg min}_{x \in A} F(x, \lambda^*, c_n), n \in \mathbb{N}\), and \(c_n \to +∞\) as \(n \to ∞\), do not have cluster points, then \(F(x, \lambda, c)\) is a penalty-type separating function for \(\lambda = \lambda^*\), as well. Therefore, we need an additional definition that allows one to exclude such pathological behavior of the function \(F(x, \lambda, c)\) as \(c \to ∞\) (see [25], Sects. 3.2–3.4, for the motivation behind this definition).

Recall that \(A\) is a subset of a finite-dimensional normed space \(X\).

**Definition 2.4** Let \(\lambda^* \in \Lambda\) be fixed. The separating function \(F(x, \lambda, c)\) is said to be nondegenerate for \(\lambda = \lambda^*\) iff there exist \(c_0 > 0\) and \(R > 0\) such that for any \(c ≥ c_0\), the function \(F(\cdot, \lambda^*, c)\) attains a global minimum on the set \(A\), and there exists \(x(c) \in \text{arg min}_{x \in A} F(x, \lambda^*, c)\) such that \(\|x(c)\| ≤ R\).

Roughly speaking, the nondegeneracy condition does not allow global minimizers of \(F(\cdot, \lambda^*, c)\) on the set \(A\) to escape to infinity as \(c \to ∞\). Note that if the set \(A\) is bounded, then \(F(x, \lambda, c)\) is nondegenerate for \(\lambda = \lambda^*\) iff the function \(F(\cdot, \lambda^*, c)\) attains a global minimum on the set \(A\) for any \(c\) large enough.

Now, we are ready to formulate and prove the localization principle. Recall that \(Ω\) is the feasible set of the problem \((\mathcal{P})\).
Theorem 2.1 (Localization Principle in the Parametric Form I) Suppose that the validity of the condition

\[ \Omega^* \cap \arg\min_{x \in A} F(x, \lambda^*, c) \neq \emptyset \]  

(1)

for some \( \lambda^* \in \Lambda \) and \( c > 0 \) implies that \( F(x, \lambda, c) \) is globally parametrically exact with the exact tuning parameter \( \lambda^* \). Let also \( \Omega \) be closed, and \( f \) be l.s.c. on \( \Omega \). Then, the separating function \( F(x, \lambda, c) \) is globally parametrically exact if and only if there exists \( \lambda^* \in \Lambda \) such that

1. \( F(x, \lambda, c) \) is of penalty type and nondegenerate for \( \lambda = \lambda^* \);
2. \( F(x, \lambda, c) \) is locally parametrically exact with the exact tuning parameter \( \lambda^* \) at every globally optimal solution of the problem \((P)\).

Proof Suppose that \( F(x, \lambda, c) \) is globally parametrically exact with an exact tuning parameter \( \lambda^* \). Then, for any \( c > c^*(\lambda^*) \), one has \( \arg\min_{x \in A} F(x, \lambda^*, c) = \Omega^* \). In other words, for any \( c > c^*(\lambda^*) \), every globally optimal solution \( x^* \) of the problem \((P)\) is a global (and hence local uniformly with respect to \( c \in (c^*(\lambda^*), +\infty) \)) minimizer of \( F(\cdot, \lambda^*, c) \) on the set \( A \). Thus, \( F(x, \lambda, c) \) is locally parametrically exact with the exact tuning parameter \( \lambda^* \) at every globally optimal solution of the problem \((P)\).

Fix arbitrary \( x^* \in \Omega^* \). Then, for any \( c > c(\lambda^*) \), the point \( x^* \) is a global minimizer of \( F(\cdot, \lambda^*, c) \), which implies that \( F(x, \lambda, c) \) is nondegenerate for \( \lambda = \lambda^* \). Furthermore, if a sequence \( \{x_n\} \subset A \) is such that \( x_n \in \arg\min_{x \in A} F(x, \lambda^*, c_n) \) for all \( n \in \mathbb{N} \), where \( c_n \to +\infty \) as \( n \to \infty \), then due to the global exactness of \( F \), one has that for all \( n \) large enough the point \( x_n \) coincides with one of the globally optimal solutions of \((P)\), which implies that \( x_n \in \Omega \), and \( f(x_n) = \min_{x \in \Omega} f(x) \). Hence, applying the facts that \( \Omega \) is closed and \( f \) is l.s.c. on \( \Omega \), one can easily verify that a cluster point of the sequence \( \{x_n\} \), if exists, is a globally optimal solution of \((P)\). Thus, \( F(x, \lambda, c) \) is a penalty-type separating function for \( \lambda = \lambda^* \).

Let us prove the converse statement. Our aim is to verify that there exist \( c > 0 \) and \( x^* \in \Omega^* \) such that

\[ \inf_{x \in A} F(x, \lambda^*, c) = F(x^*, \lambda^*, c). \]  

(2)

Then, taking into account the fact that (2) is obviously equivalent to (1), one obtains that the separating function \( F(x, \lambda, c) \) is globally parametrically exact. Arguing by reductio ad absurdum, suppose that (2) is not valid. Then, in particular, for any \( n \in \mathbb{N} \), one has

\[ \inf_{x \in A} F(x, \lambda^*, n) < F(x^*, \lambda^*, n) \quad \forall x^* \in \Omega^*. \]  

(3)

By condition 1, the function \( F(x, \lambda, c) \) is nondegenerate for \( \lambda = \lambda^* \), which implies that there exist \( n_0 \in \mathbb{N} \) and \( R > 0 \) such that for any \( n \geq n_0 \), there exists a global minimizer \( x_n \in \arg\min_{x \in A} F(x, \lambda^*, n) \) with \( \|x_n\| \leq R \).

Recall that \( X \) is a finite-dimensional normed space. Therefore, there exists a subsequence \( \{x_{n_k}\} \) converging to some \( x^* \). Consequently, applying the fact that \( F(x, \lambda, c) \) is a penalty-type separating function for \( \lambda = \lambda^* \), one obtains that \( x^* \in \Omega^* \). By condition 2, \( F(x, \lambda, c) \) is locally parametrically exact at \( x^* \) with the exact tuning parameter.
\( \lambda^* \). Therefore, there exist \( c_0 > 0 \) and a neighborhood \( U \) of \( x^* \) such that for any \( c \geq c_0 \), one has
\[
F(x, \lambda^*, c) \geq F(x^*, \lambda^*, c) \quad \forall x \in U \cap A.
\] (4)

Since the subsequence \( \{x_{n_k}\} \) converges to \( x^* \), there exists \( k_0 \) such that for any \( k \geq k_0 \), one has \( x_{n_k} \in U \). Moreover, one can suppose that \( n_k \geq c_0 \) for all \( k \geq k_0 \).

Hence, with the use of (4), one obtains that \( F(x_{n_k}, \lambda^*, n_k) \geq F(x^*, \lambda^*, n_k) \), which contradicts (3) and the fact that \( x_{n_k} \in \arg\min_{x \in A} F(x, \lambda^*, n_k) \). Thus, \( F(x, \lambda, c) \) is globally parametrically exact.

Remark 2.1 (i) Condition (1) simply means that in order to prove the global parametric exactness of \( F(x, \lambda, c) \), it is sufficient to check that at least one globally optimal solution of the problem \( (P) \) is a point of global minimum of the function \( F(\cdot, \lambda^*, c) \) instead of verifying that the sets \( \arg\min_{x \in A} F(x, \lambda^*, c) \) and \( \Omega^* \) actually coincide. It should be pointed out that in most particular cases, the validity of condition (1) is equivalent to global parametric exactness. In fact, the equivalence between (1) and global parametric exactness automatically, i.e., without any additional assumptions, holds true in all but one example (see Sect. 3.4 below) presented in this article. Note, finally, that condition (1) is needed only to prove the “if” part of the theorem.

(ii) The theorem above describes how to construct a globally exact separating function \( F(x, \lambda, c) \). Namely, one has to ensure that a chosen function \( F(x, \lambda, c) \) is of penalty type (which can be guaranteed by adding a penalty term to the function \( F(x, \lambda, c) \)), nondegenerate (which can usually be guaranteed by the introduction of a barrier term into the function \( F(x, \lambda, c) \)), and is locally exact near all globally optimal solutions of the problem \( (P) \), which is typically done with the use of constraint qualifications (metric (sub-)regularity assumptions) and/or sufficient optimality conditions. Below, we present several particular examples illustrating the usage of the previous theorem.

(iii) Note that the previous theorem can be reformulated as a theorem describing necessary and sufficient conditions for a tuning parameter \( \lambda^* \in \Lambda \) to be exact. It should also be mentioned that the theorem above can be utilized in order to obtain necessary and/or sufficient conditions for the uniqueness of an exact tuning parameter. In particular, it is easy to see that a globally exact tuning parameter \( \lambda^* \) is unique, if there exists \( x^* \in \Omega^* \) such that a locally exact tuning parameter at \( x^* \) is unique.

The theorem above can be vaguely formulated as follows. The separating function \( F(x, \lambda, c) \) is globally parametrically exact iff it is of penalty type, nondegenerate, and locally exact at every globally optimal solution of the problem \( (P) \). Thus, under natural assumptions, the function \( F(x, \lambda, c) \) is globally exact iff it is exact near globally optimal solutions of the original problem. That is why Theorem 2.1 is called the localization principle.

Let us reformulate the localization principle in the form that is slightly more convenient for applications.

Theorem 2.2 (Localization Principle in the Parametric Form II) Suppose that the validity of condition (1) for some \( \lambda^* \in \Lambda \) and \( c > 0 \) implies that \( F(x, \lambda, c) \) is globally parametrically exact with the exact tuning parameter \( \lambda^* \). Let also the sets \( A \) and \( \Omega \) be closed, the objective function \( f \) be l.s.c. on \( \Omega \), and the function \( F(\cdot, \lambda, c) \) be l.s.c.
on $A$ for all $\lambda \in \Lambda$ and $c > 0$. Then, the separating function $F(x, \lambda, c)$ is globally parametrically exact if and only if there exists $\lambda^* \in \Lambda$ such that

1. $F(x, \lambda, c)$ is of penalty type for $\lambda = \lambda^*$;
2. there exist $c_0 > 0$, $x^* \in \Omega^*$, and a bounded set $K \subset A$ such that

$$S(c, x^*) := \left\{ x \in A \mid F(x, \lambda^*, c) < F(x^*, \lambda^*, c) \right\} \subset K \quad \forall c \geq c_0; \quad (5)$$

3. $F(x, \lambda, c)$ is locally parametrically exact at every globally optimal solution of the problem $(P)$ with the exact tuning parameter $\lambda^*$.

Proof Suppose that $F(x, \lambda, c)$ is globally parametrically exact with the exact tuning parameter $\lambda^*$. Then, as it was proved in Theorem 2.1, $F(x, \lambda, c)$ is a penalty-type separating function for $\lambda = \lambda^*$, and $F(x, \lambda, c)$ is locally parametrically exact with the exact tuning parameter $\lambda^*$ at every $x^* \in \Omega^*$. Furthermore, from the definition of global exactness, it follows that $S(c, x^*) = \emptyset$ for all $c > c^*(\lambda^*)$ and $x^* \in \Omega^*$, which implies that (5) is satisfied for all $c_0 > c^*(\lambda^*)$, $x^* \in \Omega^*$ and any bounded set $K$.

Let us prove the converse statement. By our assumption, there exist $c_0 > 0$ and $x^* \in \Omega^*$ such that for all $c \geq c_0$, the sublevel set $S(c, x^*)$ is contained in a bounded set $K$ and, thus, is bounded. Therefore, taking into account the facts that the function $F(\cdot, \lambda^*, c)$ is l.s.c. on $A$ and the set $A$ is closed, one obtains that $F(\cdot, \lambda^*, c)$ attains a global minimum on the set $A$ at a point $x(c) \in K$ (if $S(c, x^*) = \emptyset$ for some $c \geq c_0$, then $x(c) = x^*$). From the boundedness of $K$, it follows that that there exists $R > 0$ such that $\| x(c) \| \leq R$ for all $c \geq c_0$, which implies that $F(x, \lambda, c)$ is nondegenerate for $\lambda = \lambda^*$. Consequently, applying Theorem 2.1, one obtains the desired result. \hfill \Box

Note that the definition of global parametric exactness does not specify how the optimal value of the problem $(P)$ and the infimum of the function $F(\cdot, \lambda^*, c)$ over the set $A$ are connected. In some particular cases (see Sect. 3.4 below), this fact might significantly complicate the application of the localization principle. Therefore, let us show how one can incorporate the assumption on the value of $\inf_{x \in A} F(x, \lambda^*, c)$ into the localization principle.

Definition 2.5 The separating function $F(x, \lambda, c)$ is said to be strictly globally parametrically exact, if $F(x, \lambda, c)$ is globally parametrically exact, and there exists $c_0 > 0$ such that $\inf_{x \in A} F(x, \lambda^*, c) = f^*$ for all $c \geq c_0$, where $\lambda^*$ is an exact tuning parameter, and $f^* = \inf_{x \in \Omega} f(x)$ is the optimal value of the problem $(P)$. Any such exact tuning parameter is called strictly exact.

Arguing in a similar way to the proofs of Theorems 2.1 and 2.2, one can easily extend the localization principle in the parametric form to the case of strict global exactness. Here, we present only an extension of Theorem 2.1.

Theorem 2.3 (Strengthened Localization Principle in the Parametric Form) Suppose that the validity of the conditions

$$\Omega^* \cap \arg \min_{x \in A} F(x, \lambda^*, c) \neq \emptyset, \quad \min_{x \in A} F(x, \lambda^*, c) = f^* \quad (6)$$

\hfill \Box
for some \( \lambda^* \in \Lambda \) and \( c > 0 \) implies that \( F(x, \lambda, c) \) is strictly globally parametrically exact with \( \lambda^* \) being a strictly exact tuning parameter. Let also \( \Omega \) be closed, and \( f \) be l.s.c. on \( \Omega \). Then, the separating function \( F(x, \lambda, c) \) is strictly globally parametrically exact if and only if there exists \( \lambda^* \in \Lambda \) such that

1. \( F(x, \lambda, c) \) is of penalty type and nondegenerate for \( \lambda = \lambda^* \);
2. \( F(x, \lambda, c) \) is locally parametrically exact at every globally optimal solution of the problem \((P)\) with the exact tuning parameter \( \lambda^* \);
3. there exists \( c_0 > 0 \) such that \( F(x^*, \lambda^*, c) = f^* \) for all \( x^* \in \Omega^* \) and \( c \geq c_0 \).

3 Applications of the Localization Principle

Below, we provide several examples demonstrating how one can apply the localization principle in the parametric form to the study of the global exactness of various penalty and augmented Lagrangian functions. Due to the page limitations, below we omit the proofs. For the proofs of the main results of this section as well as some additional results on penalty and augmented Lagrangian functions, see [46].

3.1 Example I: Linear Penalty Functions

We start with the simplest case when the function \( F(x, \lambda, c) \) is affine with respect to the penalty parameter \( c \) and does not depend on any additional parameters. Let a function \( \varphi: X \to [0, +\infty] \) be such that \( \varphi(x) = 0 \) iff \( x \in M \). Define \( F(x, c) = f(x) + c\varphi(x) \). The function \( F(x, c) \) is called a linear penalty function for the problem \((P)\).

In order to rigorously include linear penalty functions (as well as nonlinear penalty functions from the following two examples) into the theory of parametrically exact separating functions, one has to define \( \Lambda \) to be a one-point set, say \( \Lambda = \{-1\} \), introduce a new separating function \( \hat{F}(x, -1, c) \equiv F(x, c) \), and consider the separating function \( \hat{F}(x, \lambda, c) \) instead of the penalty function \( F(x, c) \). However, since this transformation is purely formal, we omit it for the sake of shortness. Moreover, since in the case of penalty functions the parameter \( \lambda \) is absent, it is natural to omit the term “parametric,” and say that \( F(x, c) \) is globally/locally exact.

With the use of the localization principle (Theorems 2.1 and 2.2), one can easily obtain a simple characterization of the global exactness of the linear penalty function \( F(x, c) \). This characterization was first obtained by the author in ([25], Theorems 3.10 and 3.17).

**Theorem 3.1** (Localization Principle for Linear Penalty Functions) Let \( \Lambda \) and \( \Omega \) be closed, and let \( f \) and \( \varphi \) be l.s.c. on \( \Lambda \). Then, the linear penalty function \( F(x, c) \) is globally exact if and only if \( F(x, c) \) is locally exact at every globally optimal solution of the problem \((P)\), and either \( F \) is nondegenerate or there exists \( c_0 > 0 \) such that the set \( \{x \in A \mid F(x, c_0) < f^*\} \) is bounded.
3.2 Example II: Nonlinear Penalty Functions

Let the function $\phi : X \rightarrow [0, +\infty]$ be as above. For the sake of convenience, suppose that the objective function $f$ is nonnegative on $X$. From the theoretical point of view, this assumption is not restrictive, since one can always replace the function $f$ with the function $e^{f(x)}$. Furthermore, it should be noted that the nonnegativity assumption on the objective function $f$ is standard in the theory of nonlinear penalty functions (cf. [10,36–39]).

Let a function $Q : [0, +\infty) \rightarrow (-\infty, +\infty]$ be fixed. Suppose that the restriction of $Q$ to the set $[0, +\infty)^2$ is strictly monotone, i.e., $Q(t_1, s_1) < Q(t_2, s_2)$ for any $(t_1, s_1), (t_2, s_2) \in [0, +\infty)^2$ such that $t_1 \leq t_2, s_1 \leq s_2$ and $(t_1, s_1) \neq (t_2, s_2)$. Suppose also that $Q(+\infty, s) = Q(t, +\infty) = +\infty$ for all $t, s \in [0, +\infty]$.

Define $F(x, c) = Q(f(x), c\phi(x))$. Then, $F(x, c)$ is a nonlinear penalty function for the problem $(P)$. This type of nonlinear penalty functions was studied in [10,36–39]. The simplest particular example of a nonlinear penalty function is the function $F_q(x, c) = (f(x)^q + (c\phi(x))^q)^{1/q}$ with $q > 0$. Here, $Q(t, s) = (t^q + s^q)^{1/q}$. Clearly, this function is strictly monotone. Let us note that the least exact penalty parameter of the nonlinear penalty function $F_q(x, c)$ is often smaller than the least exact penalty parameter of the linear penalty function $f(x) + c\phi(x)$ (see [10,37] for more details).

With the use of the localization principle, one can easily obtain a new simple characterization of the global exactness of the nonlinear penalty function $F(x, c)$, which does not rely on any assumptions on the perturbation function for the problem $(P)$ (cf. [10,37]). Furthermore, to the best of author’s knowledge, exact nonlinear penalty functions have only been considered for mathematical programming problems, while our result is applicable in the general case.

**Theorem 3.2** (Localization Principle for Nonlinear Penalty Functions) Let the set $A$ be closed, and the functions $f$, $\phi$ and $F(x, c)$ be l.s.c. on the set $A$. Suppose also that $Q(0, s) \rightarrow +\infty$ as $s \rightarrow +\infty$. Then, the nonlinear penalty function $F(x, c)$ is globally exact if and only if it is locally exact at every globally optimal solution of the problem $(P)$ and one of the two following assumptions is satisfied:

1. the function $F(x, c)$ is nondegenerate;
2. there exists $c_0 > 0$ such that the set $\{x \in A \mid Q(f(x), c_0\phi(x)) < Q(f^*, 0)\}$ is bounded.

3.3 Example III: Continuously Differentiable Exact Penalty Functions

In this section, we utilize the localization principle in order to improve existing results on the global exactness of continuously differentiable exact penalty functions. A continuously differentiable exact penalty function for mathematical programming problems was introduced by Fletcher in [26,27]. Later on, Fletcher’s penalty function was modified and thoroughly investigated in [28–30]. Here, we study a modification of the continuously differentiable penalty function for nonlinear second-order cone programming problems proposed by Fukuda, Silva, and Fukushima in [30]. However, it should be pointed out that the results of this subsection can be easily extended to the case of any existing modification of Fletcher’s penalty function.
Let $X = A = \mathbb{R}^d$ and $M = \{ x \in \mathbb{R}^d | g_i(x) \in Q_{l_i+1}, i \in I, h(x) = 0 \}$, where $g_i: X \to \mathbb{R}^{l_i+1}$, $I = \{1, \ldots, r\}$, and $h: X \to \mathbb{R}^s$ are given functions, and $Q_{l_i+1} = \{ y = (y^0, y^1) \in \mathbb{R} \times \mathbb{R}^{l_i} | y^0 \geq \|y^1\| \}$ is the second-order (Lorentz) cone of dimension $l_i + 1$ (here $\| \cdot \|$ is the Euclidean norm). In this case, the problem $(P)$ is a nonlinear second-order cone programming problem.

Following the ideas of [30], let us introduce a continuously differentiable penalty function for the problem under consideration. Suppose that the functions $f$, $g_i$, $i \in I$ and $h$ are twice continuously differentiable. For any $\lambda = (\lambda_1, \ldots, \lambda_r) \in \mathbb{R}^{l_1+1} \times \ldots \times \mathbb{R}^{l_r+1}$ and $\mu \in \mathbb{R}^s$, denote by $L(x, \lambda, \mu) = f(x) + \sum_{i=1}^r (\lambda_i, g_i(x)) + \langle \mu, h(x) \rangle$ the standard Lagrangian function for the nonlinear second-order cone programming problem. Here $\langle \cdot, \cdot \rangle$ is the inner product in $\mathbb{R}^k$. For a chosen $x \in \mathbb{R}^n$, consider the following unconstrained minimization problem, which allows one to obtain an estimate of Lagrange multipliers:

$$\min_{\lambda, \mu} \| \nabla_x L(x, \lambda, \mu) \|^2 + \sum_{i=1}^r (\lambda_i, g_i(x))^2 + \sum_{i=1}^r \| \lambda_i \|_2^2 + \| \mu \|^2,$$

where $\xi_1$ and $\xi_2$ are some positive constants, $\lambda_i = ((\lambda_i)_0, \lambda_i) \in \mathbb{R} \times \mathbb{R}^{l_i}$, and the same notation is used for $g_i(x)$. Observe that if $(x^*, \lambda^*, \mu^*)$ is a KKT-point of the problem $(P)$, then $(\lambda^*, \mu^*)$ is a globally optimal solution of problem (7) (see [30]). Moreover, it is easily seen that for any $x \in \mathbb{R}^d$, there exists a globally optimal solution of this problem, which we denote by $(\lambda(x), \mu(x))$. In order to ensure that an optimal solution is unique, one has to utilize a proper constraint qualification.

Suppose that every feasible point of the problem $(P)$ is nondegenerate (see [47], Def. 4.70, [30], Lemma 3.1). Then, one can verify that a globally optimal solution $(\lambda(x), \mu(x))$ of problem (7) is unique for all $x \in \mathbb{R}^d$, and the functions $\lambda(\cdot)$ and $\mu(\cdot)$ are continuously differentiable ([30], Proposition 3.3).

Now we can introduce a new continuously differentiable exact penalty function for nonlinear second-order cone programming problems, which is a simple modification of the penalty function from [30]. Namely, choose $\alpha > 0$ and $\kappa \geq 2$, and define

$$p(x) = \frac{\alpha - \sum_{i=1}^r \text{dist}^\kappa (g_i(x), Q_{l_i+1})}{1 + \sum_{i=1}^r \| \lambda_i(x) \|^2}, \quad q(x) = \frac{\alpha - \| h(x) \|^2}{1 + \| \mu(x) \|^2}.$$ 

Denote $\Omega_\alpha = \{ x \in \mathbb{R}^d | p(x) > 0, q(x) > 0 \}$, and define

$$F(x, c) = f(x) + \frac{c}{2 p(x)} \sum_{i=1}^r \text{dist}^\kappa (g_i(x) + \frac{p(x)}{c} \lambda_i(x), Q_{l_i+1}) - \frac{p(x)}{c^2} \| \lambda_i(x) \|^2,$$

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

(8)
if \( x \in \Omega_\alpha \), and \( F(x, c) = +\infty \) otherwise. Let us point out that \( F(x, c) \) is, in essence, a straightforward modification of the Hestenes–Powell–Rockafellar augmented Lagrangian function to the case of nonlinear second-order cone programming problems [48] with Lagrange multipliers \( \lambda \) and \( \mu \) replaced by their estimates \( \hat{\lambda}(x) \) and \( \hat{\mu}(x) \). One can easily verify that the function \( F(\cdot, c) \) is l.s.c. on \( \mathbb{R}^d \) and continuously differentiable on its effective domain (see [30]).

Applying the localization principle, one can easily obtain first simple necessary and sufficient conditions for the global exactness of continuously differentiable penalty functions.

**Theorem 3.3** (Localization Principle for \( C^1 \) Penalty Functions) Let the functions \( f \), \( g_i \), \( i \in I \), and \( h \) be twice continuously differentiable, and suppose that every feasible point of the problem (\( P \)) is nondegenerate. Then, the continuously differentiable penalty function \( F(x, c) \) is globally exact if and only if it is locally exact at every globally optimal solution of the problem (\( P \)) and one of the two following assumptions is satisfied:

1. the function \( F(x, c) \) is nondegenerate;
2. there exists \( c_0 > 0 \) such that the set \{ \( x \in \mathbb{R} \mid F(x, c_0) < f^* \) \} is bounded.

**Remark 3.1** Let us note that utilizing some specific properties ([30], Proposition 4.9) of penalty function (8), one can significantly strengthen the theorem above (see [46], Theorem 4.5).

The results of this subsection can be easily extended to the case of nonlinear semidefinite programming problems (cf. [35], Sects. 8.3 and 8.4). Namely, suppose that \( A = \mathbb{R}^d \), and let \( M = \{ x \in \mathbb{R}^d \mid G(x) \leq 0, h(x) = 0 \} \), where \( G \colon X \to \mathbb{S}/ \) and \( h \colon X \to \mathbb{R}_+ \) are given functions, \( \mathbb{S}/ \) is the set of all \( l \times l \) real symmetric matrices, and the relation \( G(x) \leq 0 \) means that the matrix \( G(x) \) is negative semidefinite. We suppose that the space \( \mathbb{S}/ \) is equipped with the Frobenius norm \( \| A \|_F = \sqrt{\text{Tr}(A^2)} \). In this case, the problem (\( P \)) is a nonlinear semidefinite programming problem.

Suppose that the functions \( f \), \( G \), and \( h \) are twice continuously differentiable. For any \( \lambda \in \mathbb{S}/ \) and \( \mu \in \mathbb{R}^4 \), denote by \( L(x, \lambda, \mu) = f(x) + \text{Tr}(\lambda G(x)) + \langle \mu, h(x) \rangle \) the standard Lagrangian function for the nonlinear semidefinite programming problem. For a chosen \( x \in \mathbb{R}^n \), consider the following unconstrained minimization problem, which allows one to compute an estimate of Lagrange multipliers:

\[
\begin{aligned}
\min_{\lambda, \mu} \left\| \nabla_x L(x, \lambda, \mu) \right\|^2 + \xi_1 \text{Tr}(\lambda^2 G(x)^2) \\
+ \frac{\xi_2}{2} \left( \| h(x) \|^2 + \sum_{i=1}^r \text{dist}^2 \left( G(x), \mathbb{S}/_i \right) \right) \cdot (\| \lambda \|^2_F + \| \mu \|^2),
\end{aligned}
\]

(9)

where \( \xi_1 \) and \( \xi_2 \) are some positive constants, and \( \mathbb{S}/_i \) is the cone of \( l \times l \) real negative semidefinite matrices. One can verify (cf. [35], Lemma 4) that for any \( x \in \mathbb{R}^d \), there exists a unique globally optimal solution \( (\lambda(x), \mu(x)) \) of this problem, provided every feasible point of the problem (\( P \)) is nondegenerate (see [47], Def. 4.70 and Proposition 5.71).
Now we can introduce first continuously differentiable exact penalty function for nonlinear semidefinite programming problems. Namely, choose $\alpha > 0$ and $\kappa \geq 1$, and define
\[
p(x) = \frac{\alpha - \text{Tr} \left( [G(x)]_+^2 \right)}{1 + \text{Tr}(\lambda(x))^2}, \quad q(x) = \frac{\alpha - \|h(x)\|^2}{1 + \|\mu(x)\|^2},
\]
where $\cdot_+$ denotes the projection of a matrix onto the cone of $l \times l$ positive semidefinite matrices. Denote $\Omega_\alpha = \{ x \in \mathbb{R}^d \mid p(x) > 0, \ q(x) > 0 \}$, and define
\[
F(x, c) = f(x) + \frac{1}{2cp(x)} \left( \text{Tr} \left( [cG(x) + p(x)\lambda(x)]_+^2 \right) - p(x)^2 \text{Tr}(\lambda(x)^2) \right)
\]
\[
+ \langle \mu(x), h(x) \rangle + \frac{c}{2q(x)} \|h(x)\|^2,
\]
(10)
if $x \in \Omega_\alpha$, and $F(x, c) = +\infty$ otherwise. Let us point out that $F(x, c)$ is, in essence, a direct modification of the Hestenes–Powell–Rockafellar augmented Lagrangian function to the case of nonlinear semidefinite programming problems [49,50] with Lagrange multipliers $\lambda$ and $\mu$ replaced by their estimates $\hat{\lambda}(x)$ and $\hat{\mu}(x)$. One can verify that the function $F(\cdot, c)$ is l.s.c. on $\mathbb{R}^d$ and continuously differentiable on its effective domain. Furthermore, it is possible to extend Theorem 3.3 to the case of continuously differentiable penalty function (10), thus obtaining first necessary and sufficient conditions for the global exactness of $C^1$ penalty functions for nonlinear semidefinite programming problems. However, we do not present an exact formulation of this result here and leave it to the interested reader.

3.4 Example IV: Rockafellar–Wets’ Augmented Lagrangian Function

The separating functions studied in the previous examples do not depend on any additional parameters apart from the penalty parameter $c$. This fact does not allow one to fully understand the concept of parametric exactness. In order to illuminate the main features of parametric exactness, in this example we consider a separating function that depends on additional parameters, namely Lagrange multipliers. Below, we apply the general theory of parametrically exact separating functions to the augmented Lagrangian function introduced by Rockafellar and Wets in [4] (see also [5,7,43–45]).

Let $P$ be a topological vector space of parameters. Recall that a function $\Phi : X \times P \to \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ is called a dualizing parameterization function for $f$ iff $\Phi(x, 0) = f(x)$ for any feasible point of the problem ($P$). A function $\sigma : P \to [0, +\infty]$ such that $\sigma(0) = 0$ and $\sigma(p) > 0$ for all $p \neq 0$ is called an augmenting function. Let, finally, $\Lambda$ be a vector space of multipliers, and let the pair $(\Lambda, P)$ be equipped with a bilinear coupling function $(\cdot, \cdot) : \Lambda \times P \to \mathbb{R}$.

Following the ideas of Rockafellar and Wets [4], define the augmented Lagrangian function
\[
\mathcal{L}(x, \lambda, c) = \inf_{p \in P} \left( \Phi(x, p) - (\lambda, p) + c\sigma(p) \right).
\]
(11)
We suppose that $\mathcal{L}(x, \lambda, c) > -\infty$ for all $x \in X, \lambda \in \Lambda$, and $c > 0$. Let us obtain simple, necessary, and sufficient conditions for the strict global parametric exactness.
of the augmented Lagrangian function $\mathcal{L}(x, \lambda, c)$ with the use of the localization principle. These conditions were first obtained by the author in [45].

**Remark 3.2** It is worth mentioning that in the context of the theory of augmented Lagrangian functions, a vector $\lambda^* \in \Lambda$ is a strictly exact tuning parameter of the function $\mathcal{L}(x, \lambda, c)$ iff $\lambda^*$ supports an exact penalty representation of the problem $(P)$ (see [4], Definition 11.60). Furthermore, if the infimum in (11) is attained for all $x, \lambda,$ and $c$, then the strict global parametric exactness of the augmented Lagrangian function $\mathcal{L}(x, \lambda, c)$ is equivalent to the existence of an augmented Lagrange multiplier (see [4], Theorem 11.61, and [45], Proposition 4 and Corollary 1). Furthermore, in this case, $\lambda^*$ is a strictly exact tuning parameter iff it is an augmented Lagrange multiplier.

Recall that the augmenting function $\sigma$ is said to have a valley at zero iff for any neighborhood $U \subset P$ of zero there exists $\delta > 0$ such that $\sigma(p) \geq \delta$ for any $p \in P \setminus U$.

**Theorem 3.4** (Localization Principle for Augmented Lagrangians) Let $A$ and $\Omega$ be closed, $f$ and $\mathcal{L}(\cdot, \lambda, c)$ for all $\lambda \in \Lambda$ and $c > 0$ be l.s.c. on $A$, $\Phi$ be l.s.c. on $A \times \{0\}$, and $\sigma$ have a valley at zero. Suppose also that there exists $r > 0$ such that $\arg \min_{p \in P} (\Phi(x, p) - (\lambda, p) + c \sigma(p)) \neq \emptyset$ (i.e., the infimum in (11) is attained) for any $c \geq r$, $x \in A$, and $\lambda \in \Lambda$. Then, the augmented Lagrangian function $\mathcal{L}(x, \lambda, c)$ is strictly globally parametrically exact if and only if there exist $\lambda^*$ and $c_0 > 0$ such that $\mathcal{L}(x, \lambda, c)$ is locally parametrically exact at every globally optimal solution of the problem $(P)$ with the exact tuning parameter $\lambda^*$, $\mathcal{L}(x^*, \lambda^*, c) = f^*$ for all $x^* \in \Omega^*$ and $c \geq c_0$, and one of the following two conditions is valid:

1. the function $\mathcal{L}(x, \lambda, c)$ is nondegenerate for $\lambda = \lambda^*$;  
2. the set $\{x \in A \mid \mathcal{L}(x, \lambda^*, c_0) < f^*\}$ is bounded.

Note that from the localization principle, it follows that for the strict global parametric exactness of the augmented Lagrangian $\mathcal{L}(x, \lambda, c)$, it is necessary that there exists a tuning parameter $\lambda^* \in \Lambda$ such that $\lambda^*$ is a locally exact tuning parameter at every globally optimal solution of the problem $(P)$. One can give a simple interpretation of this condition in the case when $\mathcal{L}(x, \lambda, c)$ is a proximal Lagrangian. Namely, let $\mathcal{L}(x, \lambda, c)$ be the proximal Lagrangian (see [4], Example 11.57), and suppose that it is strictly globally parametrically exact with a strictly exact tuning parameter $\lambda^* \in \Lambda$. By the definition of strict global exactness, any globally optimal solution $x^*$ of the problem $(P)$ is a global minimizer of the function $\mathcal{L}(\cdot, \lambda^*, c)$ for all sufficiently large $c$. Applying the first-order necessary optimality condition to the function $\mathcal{L}(\cdot, \lambda^*, c)$, one can check that under some natural assumptions for any $x^* \in \Omega^*$, the pair $(x^*, \lambda^*)$ is a KKT-point of the problem $(P)$ (see [43], Proposition 3.1). Consequently, one gets that for the strict global parametric exactness of the augmented Lagrangian function $\mathcal{L}(x, \lambda, c)$, it is necessary that there exists a Lagrange multiplier $\lambda^*$ such that the pair $(x^*, \lambda^*)$ is a KKT-point of the problem $(P)$ for any globally optimal solution $x^*$ of this problem. In particular, if there exist two globally optimal solutions of the problem $(P)$ with disjoint sets of Lagrange multipliers, then the proximal Lagrangian cannot be strictly globally parametrically exact.
4 Conclusions

In this paper, we developed a general theory of global parametric exactness of separating functions for constrained optimization problems in finite-dimensional spaces. This theory allows one to reduce a constrained optimization problem to an unconstrained one, provided an exact tuning parameter is known. With the use of the general results obtained in this article, we recovered existing results on the global exactness of linear penalty functions and Rockafellar–Wets’ augmented Lagrangian function. We also obtained new simple, necessary, and sufficient conditions for the global exactness of nonlinear and continuously differentiable penalty functions.

Acknowledgements The author is grateful to Professor Franco Giannessi for pointing out to the author the importance of development of a general theory of the exactness of penalty and augmented Lagrangian functions several years ago. The author also wishes to express his thanks to Professor X. Yang for thoughtful and stimulating comments that helped to improve the quality of the article.

References

1. Fiacco, A.V., McCormick, G.P.: Nonlinear Programming: Sequential Unconstrained Minimization Techniques. SIAM, Philadelphia (1990)
2. Auslender, A.: Penalty and barrier methods: a unified framework. SIAM J. Optim. 10, 211–230 (1999)
3. Birgin, E.G., Martinez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization. SIAM, Philadelphia (2014)
4. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
5. Huang, X.X., Yang, X.Q.: A unified augmented Lagrangian approach to duality and exact penalization. Math. Oper. Res. 28, 533–552 (2003)
6. Zhou, Y.Y., Yang, X.Q.: Some results about duality and exact penalization. J. Glob. Optim. 29, 497–509 (2004)
7. Huang, X.X., Yang, X.Q.: Further study on augmented Lagrangian duality theory. J. Glob. Optim. 31, 193–210 (2005)
8. Zhou, Y.Y., Yang, X.Q.: Augmented Lagrangian functions for constrained optimization problems. J. Glob. Optim. 52, 95–108 (2012)
9. Wang, C.Y., Yang, X.Q., Yang, X.M.: Nonlinear augmented Lagrangian and duality theory. Math. Oper. Res. 38, 740–760 (2013)
10. Rubino, A.M., Yang, X.: Lagrange-Type Functions in Constrained Non-convex Optimization. Kluwer Academic Publishers, Dordrecht (2003)
11. Wang, C.Y., Yang, X.Q., Yang, X.M.: Unified nonlinear Lagrangian approach to duality and optimal paths. J. Optim. Theory Appl. 135, 85–100 (2007)
12. Giannessi, F.: Constrained Optimization and Image Space Analysis. Separation of Sets and Optimality Conditions, vol. 1. Springer, New York (2005)
13. Li, J., Feng, S., Zhang, Z.: A unified approach for constrained extremum problems: image space analysis. J. Optim. Theory Appl. 159, 69–92 (2013)
14. Zhu, S.K., Li, S.J.: Unified duality theory for constrained extremum problems. Part I: image space analysis. J. Optim. Theory Appl. 161, 738–762 (2014)
15. Zhu, S.K., Li, S.J.: Unified duality theory for constrained extremum problems. Part II: special duality schemes. J. Optim. Theory Appl. 161, 763–782 (2014)
16. Di Pillo, G., Grippo, L.: Exact penalty functions in constrained optimization. SIAM J. Control. Optim. 27, 1333–1360 (1989)
17. Yevtushenko, YuG, Zhadan, V.G.: Exact auxiliary functions in optimization problems. U.S.S.R. Comput. Math. Math. Phys. 30, 31–42 (1990)
18. Di Pillo, G.: Exact penalty methods. In: Spedicato, E. (ed.) Algorithms for Continuous Optimization: The State of the Art, pp. 1–45. Kluwer Academic Press, Boston (1994)
19. Eremin, I.I.: Penalty method in convex programming. Soviet Math. Dokl. 8, 459–462 (1966)
20. Zangwill, W.I.: Nonlinear programming via penalty functions. Manag. Sci. 13, 344–358 (1967)
21. Han, S.P., Mangasarian, O.L.: Exact penalty functions in nonlinear programming. Math. Program. 17, 251–269 (1979)
22. Di Pillo, G., Grippo, L.: On the exactness of a class of nondifferentiable penalty functions. J. Optim. Theory Appl. 57, 399–410 (1988)
23. Burke, J.V.: An exact penalization viewpoint on constrained optimization. SIAM J. Control. Optim. 29, 968–998 (1991)
24. Dem'yanov, V.F.: Nonsmooth optimization. In: Di Pillo, G., Schoen, F. (eds.) Nonlinear Optimization. Lecture Notes in Mathematics, vol. 1989, pp. 55–164. Springer, Berlin (2010)
25. Dolgopolik, M.V.: A unifying theory of exactness of linear penalty functions. Optimization 65, 1167–1202 (2016)
26. Fletcher, R.: A class of methods for nonlinear programming with termination and convergence properties. In: Abadie, J. (ed.) Integer and Nonlinear Programming, pp. 157–173. North-Holland, Amsterdam (1970)
27. Fletcher, R.: An exact penalty function for nonlinear programming with inequalities. Math. Program. 5, 129–150 (1973)
28. Di Pillo, G., Grippo, L.: A continuously differentiable exact penalty function for nonlinear programming problems with inequality constraints. SIAM J. Control Optim. 23, 72–84 (1985)
29. Lucidi, S.: New results on a continuously differentiable exact penalty function. SIAM J. Optim. 2, 558–574 (1992)
30. Fukuda, E.H., Silva, P.J.S., Fukushima, M.: Differentiable exact penalty functions for nonlinear second-order cone programs. SIAM J. Optim. 22, 1607–1633 (2012)
31. Di Pillo, G., Grippo, L.: A new class of augmented Lagrangians in nonlinear programming. SIAM J. Control Optim. 17, 618–628 (1979)
32. Di Pillo, G., Grippo, L.: A new augmented Lagrangian function for inequality constraints in nonlinear programming problems. J. Optim. Theory Appl. 36, 495–519 (1982)
33. Di Pillo, G., Lucidi, S.: An augmented Lagrangian function with improved exactness properties. SIAM J. Optim. 12, 376–406 (2001)
34. Fukuda, E.H., Lourenco, B.F.: Exact augmented Lagrangian functions for nonlinear semidefinite programming. arXiv:1705.06551 (2017)
35. Dolgopolik, M.V.: Augmented Lagrangian functions for cone constrained optimization: the existence of global saddle points and exact penalty property. J. Glob. Optim. (2018). https://doi.org/10.1007/s10898-017-0603-0
36. Rubinov, A.M., Glover, B.M., Yang, X.Q.: Decreasing functions with applications to penalization. SIAM J. Optim. 10, 289–313 (1999)
37. Rubinov, A.M., Yang, X.Q., Bagirov, A.M.: Penalty functions with a small penalty parameter. Optim. Methods Softw. 17, 931–964 (2002)
38. Rubinov, A.M., Gasimov, R.N.: Strictly increasing positively homogeneous functions with applications to exact penalization. Optimization 52, 1–28 (2003)
39. Yang, X.Q., Huang, X.X.: Partially strictly monotone and nonlinear penalty functions for constrained mathematical programs. Comput. Optim. Appl. 25, 293–311 (2003)
40. Huyer, W., Neumaier, A.: A new exact penalty function. SIAM J. Optim. 13, 1141–1158 (2003)
41. Wang, C., Ma, C., Zhou, J.: A new class of exact penalty functions and penalty algorithms. J. Glob. Optim. 58, 51–73 (2014)
42. Dolgopolik, M.V.: A unifying theory of exactness of linear penalty functions II: parametric penalty functions. Optimization 66, 1577–1622 (2017)
43. Shapiro, A., Sun, J.: Some properties of the augmented Lagrangian in cone constrained optimization. Math. Oper. Res. 29, 479–491 (2004)
44. Zhou, Y.Y., Zhou, J.C., Yang, X.Q.: Existence of augmented Lagrange multipliers for cone constrained optimization problems. J. Glob. Optim. 58, 243–260 (2014)
45. Dolgopolik, M.V.: Existence of augmented Lagrange multipliers: reduction to exact penalty functions and localization principle. Math. Program. 166, 297–326 (2017)
46. Dolgopolik, M.V.: A unified approach to the global exactness of penalty and augmented Lagrangian functions I: parametric exactness. arXiv:1709.07073 (2017)
47. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2000)
48. Liu, Y.J., Zhang, L.W.: Convergence of the augmented Lagrangian method for nonlinear optimization problems over second-order cones. J. Optim. Theory Appl. 139, 557–575 (2008)
49. Sun, D., Sun, J., Zhang, L.: The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. Math. Program. 114, 349–391 (2008)
50. Wu, H., Luo, H., Ding, X., Chen, G.: Global convergence of modified augmented Lagrangian methods for nonlinear semidefinite programming. Comput. Optim. Appl. 56, 531–558 (2013)