Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 12, No. 4, pp. 795–799, October 2016

NEW CHARACTERIZATIONS OF EXACT REGULARIZATION OF NON-CONVEX PROGRAMS

S. Deng

Abstract: Optimization and regularization play an important role in solving inverse problems. For a non-convex program along with a regularization function, we show in [4] that the regularization is exact if and only if the Lagrangian function of a certain selection problem has a saddle point, and the regularization parameter threshold is inversely related to the Lagrange multiplier associated with the saddle point. These results completely generalize the main results in [6] on a characterization of exact regularization of a convex program to that of a nonlinear (not necessarily convex) program. In this short note, we give additional new characterizations of exact regularization including exact penalization and stability for associated perturbed problems. We also discuss implications of these new characterizations.

Key words: *saddle point, Lagrangian function, exact regularization, exact penalization, stability, and calmness.*

Mathematics Subject Classification: *90C26, 90C30, 90C46.*

1 Introduction

For many practical inverse problems, very often there are multiple solutions or solutions are not well-behaved with respect to problem data. A commonly used optimization technique to deal with this challenge is called *regularization*. Generally speaking, the technique of regularization is to solve an ill-posed optimization problem with non-unique solutions by constructing a related problem whose solution is well behaved and deviated only slightly from a solution of the original problem. Usually deviations from solutions of the original problem are generally accepted as a trade-off for obtaining solutions with other desirable properties. Let us use the following inverse problem to illustrate the ideas involved.

Example 1.1. The forward noiseless-observation model in many data acquisition scenarios can be formulated as follows

$$y_0 = Ax_s,$$

where $y_0 \in R^m$ is an observation vector, $x_s \in R^n$ is the unknown signal to recover, and A is a linear operator from the signal domain R^n into the observation domain R^m with $m \leq n$. Even for the case $m = n$, A is in general ill-conditioned or is not invertible. This makes the problem of minimizing $||y_0 - Ax||^2$ over R^n difficult to solve. A practical way to deal with this challenge is to solve the following regularized problem:

$$\min_{x\in R^n} \quad 1/2||y_0 - Ax||_2^2 + \delta f(x),$$

where f is an appropriate regularization term through which some regularity is enforced on the recovered signal, and $\delta > 0$.

Clearly it is desirable if solutions of the regularized problem are also solutions of the original problem. In Example 1.1, this amounts to saying that solution(s) of the regularized problem are solutions for $Ax = y_0$. Friedlander and Tseng [6] presented a systematic study for exact regularization of a convex program. According to [6] the regularization is exact if the solutions of perturbed problems are also solutions of the original problem for all values of penalty parameters below some positive threshold value. In [4], using a variational approach to saddle points of the Lagrangian function associated with a given problem, we demonstrate that the main results of [6] can be completely generalized to non-convex programs; thereby potential applications of the exact regularization technique have been significantly expanded. Further investigations on the notion of strongly exact regularization introduced in [5] and its connections to normal cone identity, and to weak sharp minima [2, 3] are carried out in [5]. In this note, we provide new characterizations (see Theorem 3.1) for exact regularization including exact penalization and stability for perturbed problems (a property closely related to the calmness property). Along the way, we discuss and illustrate, by examples, implications of newly derived results.

The notation used in this note is standard. See e.g. [8]. All vectors are column vectors and the symbol "T" denotes the transpose of a column vector.

2 Problem Statement and Review of Existing Results

To facilitate our discussion in a rigorous mathematical framework, let us consider the following nonlinear program

$$(\mathcal{P}) \qquad \min\ g(x) \quad \text{s.t. } x \in C,$$

where $g : R^n \to R$ is a continuous function, and C is a non-empty closed set in R^n. Let S be the set of all optimal solutions of $(\mathcal{P})$. Throughout the note we suppose that the solution set S is nonempty, and denote the optimal value of $(\mathcal{P})$ by p^*. When $(\mathcal{P})$ has multiple solutions or is very sensitive to data perturbations, a popular way to regularize the problem is to modify the objective function by adding a new function f (which is called a regularization function). This leads to the following regularized problem

$$(\mathcal{P}(\delta)) \qquad \min\ g(x)+\delta f(x) \quad \text{s.t. } x \in C,$$

where $f : R^n \to R$ is a continuous function and δ is a nonnegative regularization parameter. Let S_δ be the set of optimal solutions of $(\mathcal{P}(\delta))$. In this note, we do not assume f is *convex*; that is, the regularization function f may be nonlinear, non-convex or non-differentiable. A popular choice, commonly known as Tikhonov regularization, of f is $||x||_2^2$, which is used to select a least two-norm solution. Another popular choice is l_1 regularization with $f(x) = ||x||_1$, which usually find an optimal solution with the property of the least number of nonzero elements among all solutions giving the least-f value.

The notion of exact regularization follows.

Definition 2.1 ([6]). For $(\mathcal{P})$ with a given regularization function f, we say that *regularization is exact* if the solutions of $(\mathcal{P}(\delta))$ are also solutions of $(\mathcal{P})$ for all values of δ below some positive threshold value $\bar{\delta}$; that is, $S_\delta \subset S$ for all $\delta \leq \bar{\delta}$.

As done in [6, 4], a crucial approach to understanding exact regularization is to study a related nonlinear program that selects solutions of $(\mathcal{P})$ with the least value measured by f:

$$(\mathcal{Q}) \qquad \min\ f(x) \quad \text{s.t } x \in C, \ \ g(x) \leq p^*,$$

where p^* denotes the optimal value of $(\mathcal{P})$. Let S_Q be the set of optimal solutions of $(\mathcal{Q})$, and suppose that $S_Q \neq \emptyset$. The Lagrangian function for problem $(\mathcal{Q})$ is the function over $C \times R_+ \subset R^n \times R$ defined by

$$L(x, y) = f(x) + y(g(x) - p^*).$$

We say that a pair of vector $(\bar{x}, \bar{y}) \in C \times R_+$ gives a saddle point of the Lagrangian L on $C \times R_+$ if

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) \qquad \forall x \in C, \forall y \in R_+.$$

It is well known that problem $(\mathcal{Q})$ does not have to have a Lagrange multiplier even for the convex case as illustrated by the following example.

Example 2.2. Let $f(x) = x$, $g(x) = (x-1)^2$, and $C = R$. Then $S_Q = S = \{1\}$ and the optimal value of $(\mathcal{Q})$ is 1. For the Lagrangian $L(x, y) = x + y(x-1)^2$, there are no saddle points for L over $R \times R_+$ since for $y \geq 0$, $\inf_{x \in C} L(x, y) = -\infty$ if $y = 0$ and $\inf_{x \in C} L(x, y) = -\frac{1}{4y} + 1$ if $y > 0$.

For the convex case, by which we mean that f, g and C are convex, Friedlander and Tseng [6] were able to to characterize exact regularization by way of the existence of Lagrangian- multipliers for $(\mathcal{Q})$. For the general case, we presented a saddle-point approach to characterize exact regularization of a non-convex program in [4]. We list below as Theorems 2.3 and 2.4 some key results from [4], which will be used in this note. Note that Theorem 2.3 is not true for general non-convex programs.

Theorem 2.3 (Theorem 2.3 of [4])**.** *For problem $(\mathcal{Q})$, a pair $(\bar{x}, \bar{y}) \in C \times R_+$ is a saddle point of the Lagrangian L if and only if the pair satisfies the conditions:*

(I) *$\bar{x} \in S$;*

(II) *$\bar{x}$ is a minimizer of $L(\cdot, \bar{y})$ over C.*

In particular, $\bar{x}$ is an optimal solution of $(\mathcal{Q})$.

Theorem 2.4 (Corollary 2.7 of [4])**.** *If there is some $\bar{y} > 0$ such that $(\bar{x}, \bar{y})$ is a saddle point of L, then $\bar{x} \in S_\delta$ where $\delta = 1/\bar{y}$. Conversely if $S_\delta \cap S \neq \emptyset$, then for any $\bar{x} \in S_\delta \cap S$, $(\bar{x}, \bar{y})$ is a saddle point of L where $\bar{y} = 1/\delta$.*

3 New Characterizations

To better understand the role of Lagrange multiplier y, we define the perturbation function as follows:

$$\rho(u) = \inf\{f(x) \mid x \in C(u)\}, \quad \text{where } C(u) = \{x \in C \mid g(x) \leq p^* + u\},$$

and we use the convention that $\rho(u) = +\infty$ if $C(u) = \emptyset$. We see that $C(0) = C \cap S$.

Recall [7] that $(\mathcal{Q})$ is said to be *stable* if $\rho(0)$ is finite and there is a scalar $M > 0$ such that

$$\frac{\rho(u) - \rho(0)}{|u|} \geq -M \qquad \text{for all } |u| \neq 0.$$

Our new characterizations of exact regularization can now be stated as follows.

Theorem 3.1. *Let $\bar{x}$ be an optimal solution of $(\mathcal{Q})$. Then the following are equivalent:*

(a) *there is some* $\bar{y} \geq 0$ *such that* $(\bar{x}, \bar{y})$ *is a saddle point of the Lagrangian* L*;*

(b) *there is some* $\bar{y} \geq 0$ *such that* $\rho(u) + \bar{y}u \geq \rho(0) \quad \forall u \in R$*;*

(c) *there is some* $M > 0$ *such that* $\frac{\rho(u)-\rho(0)}{u} \geq -M$ *for all* $u > 0$*;*

(d) *there is some* $r > 0$ *such that* $\bar{x}$ *is an optimal solution to the problem of minimizing*

$$f(x) + r \max\{g(x) - p^*, 0\} \quad \text{over } C.$$

$(a) \Rightarrow (b)$. : For any given real number u, let $L_u(x, y) = f(x) + y(g(x) - p^* - u)$. From $\sup_{y \in R_+} L_u(x, y) \geq f(x) + \bar{y}(g(x) - p^* - u)$, we get

$$\rho(u) = \inf_{x \in C}\{\sup_{y \in R_+} L_u(x, y)\} \geq \inf_{x \in C}\{f(x) + \bar{y}(g(x) - p^* - u)\}$$

$$= \inf_{x \in C}\{f(x) + \bar{y}(g(x) - p^*)\} - \bar{y}u = f(\bar{x}) - \bar{y}u = \rho(0) - \bar{y}u.$$

$[(b) \Rightarrow (c)]$: This is evident.

$[(c) \Rightarrow (d)]$: We prove this statement by negation. Suppose that there are sequences $\{x_k\} \subset C$ and $\{r_k\} \subset R_+$ with $r_k \to +\infty$ such that

$$f(x_k) + r_k \max\{g(x_k) - p^*, 0\} < f(\bar{x}). \tag{3.1}$$

If $g(x_k) \leq p^*$, then $f(x_k) = f(x_k) + r_k \max\{g(x_k) - p^*, 0\} < f(\bar{x})$, a contradiction. Hence, $g(x_k) > p^*$ for all k. Set $u_k = g(x_k) - p^*$ for each k. Then $u_k > 0$ and $\rho(u_k) \leq f(x_k)$ and by (3.1)

$$\frac{\rho(u_k) - \rho(0)}{u_k} \leq \frac{f(x_k) - f(\bar{x})}{\max\{g(x_k) - p^*, 0\}} \leq -r_k \to -\infty,$$

which is a contradiction to Statement (c). This shows that (3.1) is impossible and we are done.

$[(d) \Rightarrow (a)]$: Suppose that there is some $r > 0$ such that $\bar{x}$ solves the problem of minimizing $f(x) + r \max\{g(x) - p^*, 0\}$ over C. As $g(x) \geq p^*$ for all $x \in C$, we have that $\bar{x}$ solves the problem of minimizing the Lagrangian $L(\cdot, r) = f(\cdot) + r(g(\cdot) - p^*)$ over C. By Theorem 2.3, $(\bar{x}, r)$ is a saddle point for L. This completes the proof. □

Remark 3.2. Statement (b) says that $\bar{y}$ is a 'subgradient ' of the perturbation function ρ, which may not be convex in general unless f, g, and C are convex. Statement (c) means that $(\mathcal{Q})$ is stable, which is closely related to the notion of calmness [1, 8]; and Statement (d) says $\bar{x}$ is a minimizer of an exact penalty function. Theorem 3.1 states that they all are equivalent to the existence of a saddle point for the Lagrangian L.

To illustrate Theorem 3.1, let us compute ρ for a non-convex program with $f(x) = x^3$, $g(x) = (x-1)^2$, and $C = R$, for $u \geq 0$. Then $(x-1)^2 \leq u$ is equivalent to $-\sqrt{u} \leq x-1 \leq \sqrt{u}$. This gives $\rho(u) = (-\sqrt{u} + 1)^3 = 1 - 3\sqrt{u} + 3u - u^{3/2}$, and $\frac{\rho(u)-\rho(0)}{u} \to -\infty$ as $u \downarrow 0$. By Theorem 3.1, there is no saddle point(s) for the Lagrangian function for $(\mathcal{Q})$.

We now consider a special case where g is convex and differentiable, and C is a closed convex set. In this case, as well-known $\nabla g(x)$ is a constant vector on S. So we just denote it by g', and we have the following consequence based on Theorem 3.1.

Corollary 3.3. *Let* $\bar{x} \in S_Q$*. Suppose that there is some* $\bar{y} \geq 0$ *such that* $\bar{x}$ *solves the problem of minimizing*

$$f(x) + \bar{y}(g')^T(x - \bar{x}) \qquad \text{over } C.$$

Then $(\bar{x}, \bar{y})$ *is a saddle point of the Lagrangian function* L *for* $(\mathcal{Q})$*.*

Proof. Since $\bar{x} \in S_Q \subset S$, $\nabla g(\bar{x}) = g'$ and $g(\bar{x}) = p^*$. By convexity of g, for any $x \in C$, we have

$$f(x) + \bar{y}(g(x) - p^*) = f(x) + \bar{y}(g(x) - g(\bar{x})) \geq f(x) + \bar{y}(g')^T(x - \bar{x}).$$

So $(\bar{x}, \bar{y})$ is a saddle point of the Lagrangian function L by Theorem 2.3. This completes the proof. □

Remark 3.4. An attractive feature of Corollary 3.3 is that the assumptions of Corollary 3.3 can be interpreted as $(\bar{x}, \bar{y})$ being a saddle point of the Lagrangian function for the following "linearized" problem of minimizing $f(x)$ subject to $x \in C$ and $(g')^T(x - \bar{x}) \leq 0$.

Acknowledgment

We thank one referee for the suggestions which helped us to improve the presentation of the paper.

References

[1] J. V. Burke, Calmness and exact penalization. *SIAM J. Control Optim.* 29 (1991) 493–497.

[2] J. V. Burke and M. Ferris, Weak sharp minima in mathematical programming. *SIAM J. Control Optim.* 31 (1993) 1340–1359.

[3] J. V. Burke and S. Deng, Weak sharp minima revisited, Part I: Basic theory. *Control and Cybernetics* 31 (2002) 439–469.

[4] S. Deng, A saddle point characterization of exact regularization of non-convex programs, *Pacific Journal of Optimization* 8 (2012) 27–32.

[5] S. Deng, Exact regularization, and its connections to normal cone identity and weak sharp minima in nonlinear programming, in *Optimization Methods, Theory and Applications*, H. Xu, S. Wang, and S-Y Wu (eds.), Springer-Verlag Berlin Heidelberg, 2015, pp. 175–182.

[6] M. Friedlander and P. Tseng, Exact regularization of convex programs, *SIAM J. Optim.* 18 (2007) 1326–1350.

[7] A. M. Geoffron, Duality in nonlinear programing: a simplified applications-oriented development, *SIAM Review* 13 (1971) 1–37.

[8] R. T. Rockafellar and R. Wets, *Variational Analysis,* Springer-Verlag Berlin Heidelberg, 1998.

Manuscript received 14 July 2014
revised 25 January 2015
accepted for publication 29 January 2016

S. Deng
Department of Mathematical Sciences, Northern Illinois University, DeKalb, Illinois, USA
E-mail address: sdeng@niu.edu