Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 8, No. 1, pp. 27–32, January 2012

A SADDLE POINT CHARACTERIZATION OF EXACT REGULARIZATION OF NON-CONVEX PROGRAMS

S. Deng

In memory of Paul Tseng, who made numerous important contributions to optimization including the topic of exact regularization of nonlinear optimization problems studied in this paper.

Abstract: The regularization of a nonlinear program is exact if all solutions of the regularized problem are also solutions of the original problem for all values of the regularization parameter below some positive threshold. In this note, we show that the regularization is exact if and only if the Lagrangian function of a certain selection problem has a saddle point. Moreover, the regularization parameter threshold is inversely related to the Lagrange multiplier associated with the saddle point. Our results not only provide a fresh perspective on exact regularization but also extend the main results of Friedlander and Tseng [2] on a characterization of exact regularization of a convex program to that of a nonlinear (not necessarily convex) program.

Key words: *saddle point, Lagrangian function, exact regularization*

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

---

1 Introduction

The technique of regularization is a common approach used to solve an ill-posed nonlinear 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. However, it would be more desirable if solutions of the regularized problem were also solutions of the original problem. In a recent paper by Friedlander and Tseng [2], the authors presented a systematic study for exact regularization of a convex program. The term *exact regularization* was coined in [2]; according to [2] 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 this note, we demonstrate that the main results of Friedlander and Tseng [2] can be extended to non-convex programs thereby the application domain of this exact regularization technique could be significantly expanded. It is our hope that theoretical tools developed in this paper will advance the research on the technique of exact regularization for non-convex programs.

Specifically, consider the general nonlinear program

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

ISSN 1348-9151

where $g : R^n \to R$ is a continuous function, and C is a closed set in R^n. 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. 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. 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 can be used to select a least two-norm solution. Another popular choice is l_1 regularization with $f(x) = ||x||_1$. See [2] for more examples of f, applications of exact regularization of convex programs, and connections between exact regularization and exact penalization which is commonly used for solving constrained nonlinear programs.

We make the following basic assumptions throughout the paper:

A 1: The functions f, g are locally Lipschitz continuous.

A 2: The feasible set C is a nonempty closed set in R^n and is Clarke regular [1] at every point of C, and the solution set S of $(\mathcal{P})$ is nonempty. The Clarke regularity assumption on C holds whenever C is convex.

A 3: The level set $\{x \in S \mid f(x) \le \beta\}$ is bounded for each $\beta \in R$, and $\inf_{x\in C} f(x) > -\infty$. This assumption holds whenever f is coercive.

Similar to the approaches used in [2], central to our analysis is a related nonlinear program that selects solutions of $(\mathcal{P})$ of the least $f-$value:

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

where p^* denotes the optimal value of $(\mathcal{P})$. Note that $S = \{x \in C \mid g(x) \le p^*\}$. By A 3, $(\mathcal{Q})$ has at least one optimal solution. Clearly, any solution of $(\mathcal{Q})$ is also a solution of $(\mathcal{P})$, i.e., $S_Q \subset S$ where S_Q is the set of solutions of $(\mathcal{Q})$. As already known in the convex case (by which we mean both the functions g and f, and the set C are convex), solutions of $(\mathcal{P}(\delta))$ need not be solutions of $(\mathcal{P})$, here we denote the set of solutions of $(\mathcal{P}(\delta))$ by S_δ. Following [2], we say that the 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 $\bar{\delta}$. The purpose of this note is to give necessary and sufficient conditions for the aforementioned exact regularization to hold. Specifically, in Theorem 2.6, we show that the regularization $(\mathcal{P}(\delta))$ is exact if and only if the Lagrangian function of the selection problem $(\mathcal{Q})$ has a saddle point. Moreover, the solution set of $(\mathcal{P}(\delta))$ coincides with the solution set of $(\mathcal{Q})$ for all $\delta < 1/\bar{y}$, where $\bar{y}$ is the Lagrange multiplier associated with the saddle point. We will report applications of the main results elsewhere.

The notation used in this note is standard. See e.g. [4].

2 Main Results

Let the Lagrangian function of $(\mathcal{Q})$ be

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

for $x \in C$ and $y \ge 0$. Let $\bar{x} \in C$ and suppose that $\bar{x}$ is a local minimizer of $(\mathcal{Q})$. We say that $\bar{y} \ge 0$ is a Lagrange multiplier at $\bar{x}$ if

$$0 \in \partial_x L(\bar{x}, \bar{y}) + N_C(\bar{x}),$$

where $\partial_x L(\bar{x}, \bar{y})$ and $N_C(\bar{x})$ denote the Clarke sub-differential [1] of $L(\cdot, \bar{y})$ at $\bar{x}$ and the Clarke normal cone [1] at $\bar{x}$ respectively. 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

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) \quad \forall x \in C \subset R^n, \forall y \in R_+.$$

Note that problem $(\mathcal{Q})$ may not have a Lagrange multiplier even for the convex case as illustrated by the following example.

Example 2.1. Let $g(x) = x^2$, $C = R$, and $f(x) = x$. Then $S = argmin_{x \in C} g = \{0\}$. For $L(x, y) = x + yx^2$, there are no saddle points for L over $R \times R_+$. For $y \geq 0$, $\inf_{x \in C} L(x, y) = -\infty$ if $y = 0$, $= -\frac{1}{4y}$ if $y > 0$.

To better understand the role of Lagrange multiplier λ, we define $C(u) = \{x \in C |\, g(x) \leq p^* + u\}$, and the perturbation function $\rho(u) = \inf\{f(x) \mid x \in C(u)\}$. We use the convention that $\rho(u) = +\infty$ if $C(u) = \emptyset$.

We begin with a basic characterization concerning saddle points of the Lagrangian L, which holds for general non-convex programs. More detailed discussion about this characterization along with a proof can be found in [3].

Theorem 2.2 (Theorem 6.1 [3])**.** *For problem $(\mathcal{Q})$, $(\bar{x}, \bar{y})$ is a saddle point of the Lagrangian L for $(\mathcal{Q})$ if and only if*

(a) *$\bar{x}$ solves $(\mathcal{Q})$;*

(b) *$\rho(u) + \bar{y}u \geq \rho(0) \quad \forall u \in R$.*

We now present another characterization of saddle point conditions for $(\mathcal{Q})$, which has direct consequences on the exact regularization of non-convex programs. Moreover this characterization is not true for general non-convex programs.

Theorem 2.3. *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:*

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

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

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

Proof. [$\Rightarrow$] Let $(\bar{x}, \bar{y})$ be a saddle point of L. Then (2) holds. So we only need to show that (1) holds. If $g(\bar{x}) > p^*$, then by $L(\bar{x}, y) \leq L(\bar{x}, \bar{y})$ for all $y \in R_+$, $0 \leq (\bar{y} - y)(g(\bar{x}) - p^*)$ for all $y \in R_+$. This implies that $\bar{y}$ cannot be zero. If $\bar{y} \neq 0$, we get a contradiction again by choosing $y = 2\bar{y}$. We conclude that $g(\bar{x}) \leq p^*$. So $\bar{x} \in S$, i.e., (1) holds.

[$\Leftarrow$] We observe the following.

$$\begin{aligned} f(\bar{x}) &= L(\bar{x}, \bar{y}) && \text{since } g(\bar{x}) = p^* \\ &= \inf_{x \in C} L(x, \bar{y}) && \text{by (2)} \\ &\leq \inf\{L(x, \bar{y}) \mid g(x) \leq p^* \ , x \in C\} \\ &= \inf\{f(x) \mid x \in S\} \end{aligned}$$

The above inequality implies that $\bar{x}$ is an optimal solution of $(\mathcal{Q})$. Since $g(\bar{x}) = p^*$, $(y - \bar{y})(g(\bar{x}) - p^*) = 0 \leq 0$ for all $y \in R_+$. This along with the equality $\inf_{x \in C} L(x, \bar{y}) = f(\bar{x})$ shows that $(\bar{x}, \bar{y})$ is a saddle point of L on $C \times R_+$. $\square$

When $(\mathcal{Q})$ is a convex program, it is well-known that the existence of Lagrange multipliers for $(\mathcal{Q})$ is equivalent to the existence of saddle point for L, and the set of Lagrange multipliers is the same for any solutions of $(\mathcal{Q})$. For the non-convex case, Lagrange multipliers are not necessary the same for different optimal solutions of $(\mathcal{Q})$. However, due to the special structure of $(\mathcal{Q})$, we do have the following two lemmas on saddle points of the Lagrangian L, which do not hold for general non-convex programs. The lemmas form the basis for our research on exact regularization of non-convex programs via saddle point conditions.

Lemma 2.4. *Let x_1 and x_2 be optimal solutions of $(\mathcal{Q})$. Suppose that $(x_1, \bar{y})$ with $\bar{y} \geq 0$ is a saddle point of L. Then $(x_2, \bar{y})$ is a saddle point of L too.*

Proof. Since x_1 and x_2 solve $(\mathcal{Q})$, $g(x_1) = g(x_2)$ and $f(x_1) = f(x_2)$. The saddle point $(x_1, \bar{y})$ of L implies that

$$f(x_1) + \bar{y}(g(x_1) - p^*) \leq f(x) + \bar{y}(g(x) - p^*) \quad \forall x \in C.$$

But $f(x_2) = f(x_1)$ and $g(x_2) = p^* = g(x_1)$. So

$$f(x_2) + \bar{y}(g(x_2) - p^*) \leq f(x) + \bar{y}(g(x) - p^*) \quad \forall x \in C.$$

Since $g(x_2) = p^*$, it is easy to verify that $L(x_2, y) \leq L(x_2, \bar{y})$ for all $y \geq 0$. This shows that $(x_2, \bar{y})$ is a saddle point of L. □

Lemma 2.5. *For any given optimal solution $\bar{x}$ of $(\mathcal{Q})$, let $Y \subset R_+$ be the set such that $(\bar{x}, \bar{y})$ is a saddle point of L with $\bar{y} \in Y$. If Y is non-empty, then Y is a closed convex set.*

Proof. Let $y_1, y_2 \in Y$. Without loss of generality suppose that $y_1 < y_2$. Let $y \in (y_1, y_2)$. Then there is some $\lambda \in (0,1)$ such that $y = \lambda y_1 + (1-\lambda)y_2$. Then for any $x \in C$,

$$\begin{aligned} f(\bar{x}) &= f(\bar{x}) + y(g(\bar{x}) - p^*) \qquad (\text{since } g(\bar{x}) = p^*) \\ &= \lambda(f(\bar{x}) + y_1[g(\bar{x}) - p^*)]) + (1-\lambda)(f(\bar{x}) + y_2[g(\bar{x}) - p^*)]) \\ &\leq \lambda(f(x) + y_1[g(x) - p^*)]) + (1-\lambda)(f(x) + y_2[g(x) - p^*)]) \\ &= f(x) + y(g(x) - p^*). \end{aligned}$$

This proves the convexity of Y. Let $y(i) \to \bar{y}$ with $y(i) \in Y$ as $i \to +\infty$. Then $f(\bar{x}) \leq f(x) + y(i)(g(x) - p^*)$ for each fixed $x \in C$. Letting $i \to +\infty$, we get $\bar{y} \in R_+$ and

$$f(\bar{x}) \leq f(x) + \bar{y}(g(x) - p^*)$$

for each $x \in C$. So Y is closed and convex. □

We summarize some important consequences of the proceeding lemmas and theorems in the following theorem which generalizes the main results (Theorem 2.1) of [2] on exact regularization to the non-convex case. The proof techniques are similar to these used in [2].

Theorem 2.6. *Consider problems $(\mathcal{P})$, $(\mathcal{Q})$, and $(\mathcal{P}(\delta))$. Then the following statements are true.*

(a) *For any $\delta > 0$, $S \cap S_\delta \subset S_Q$.*

(b) *If there exists a saddle point $(\bar{x}, \bar{y})$ of L for $(\mathcal{Q})$ with $\bar{x} \in S_Q$, then $S \cap S_\delta = S_Q$ for all $\delta \in (0, 1/\bar{y}]$. Here we use the convention $1/\bar{y} = +\infty$ when $\bar{y} = 0$.*

(c) *If there exists* $\bar{\delta} > 0$ *such that* $S \cap S_{\bar{\delta}} \neq \emptyset$, *then* $(\bar{x}, 1/\bar{\delta})$ *is a saddle point of* L *for* $(\mathcal{Q})$ *with any* $\bar{x} \in S \cap S_\delta = S_Q$ *for all* $\delta \in (0, \bar{\delta}]$.

(d) *If there exists* $\bar{\delta} > 0$ *such that* $S \cap S_{\bar{\delta}} \neq \emptyset$, *then* $S_\delta \subset S$ *for all* $\delta \in (0, \bar{\delta})$.

Proof. (a). Let $\bar{x} \in S$ and $\bar{x} \in S_\delta$ with $\delta > 0$. Then $\bar{x}$ is a minimizer of $f(x) + (1/\delta)g(x)$ over C since $\bar{x} \in S_\delta$. So Theorem 2.3 informs us that $\bar{x} \in S_Q$. that is, $S \cap S_\delta \subset S_Q$.

(b). We consider two cases $\bar{y} = 0$ and $\bar{y} > 0$. Case 1: $\bar{y} = 0$. Then for any $\hat{x} \in S_Q$, by Lemma 2.4, $(\hat{x}, 0)$ is a saddle point of L. So $f(\hat{x}) \leq f(x)$ for all $x \in C$. Since $\hat{x} \in S$, $g(\hat{x}) \leq g(x)$ for all $x \in C$. Then for any $\delta \geq 0$,

$$g(\hat{x}) + \delta f(\hat{x}) \leq g(x) + \delta f(x), \quad \forall x \in C.$$

So $\hat{x} \in S_\delta$.

Case 2: $\bar{y} > 0$. For any $\hat{x} \in S_Q$, by Lemma 2.4, $(\hat{x}, \bar{y})$ is a saddle point of L. Then

$$\hat{x} \in argmin_{x \in C}[(1/\bar{y})f(x) + (g(x) - p^*)] = argmin_{x \in C}[(1/\bar{y})f(x) + g(x)].$$

Also since $\hat{x} \in S$, $g(\hat{x}) \leq g(x)$ for all $x \in C$. The above two relations yield that

$$g(\hat{x}) + (\lambda/\bar{y})f(\hat{x}) \leq g(x) + (\lambda/\bar{y})f(x), \quad \forall x \in C, \quad \forall \lambda \in [0, 1].$$

So $\hat{x} \in S_\delta$ for all $\delta \in [0, 1/\bar{y}]$.

(c). For any $\bar{x} \in S \cap S_{\bar{\delta}}$, $\bar{x}$ is a minimizer of $f(x) + (1/\bar{\delta})g(x)$ over C. By Theorem 2.3, $\bar{x} \in S_Q$ and $(\bar{x}, 1/\bar{\delta})$ is a saddle point of L. By (b), $S \cap S_\delta = S_Q$ for all $\delta \in (0, \bar{\delta}]$.

(d). Let $g_\delta = g + \delta f$. Assume that there exists a $\bar{\delta} > 0$ such that $S \cap S_{\bar{\delta}} \neq \emptyset$. Let $\bar{x} \in S \cap S_{\bar{\delta}}$ be given. For any $\delta \in (0, \bar{\delta})$ and any $x \in C \backslash S$, we have

$$g_{\bar{\delta}}(\bar{x}) \leq g_{\bar{\delta}}(x), \quad g(\bar{x}) < g(x).$$

Since $0 < \frac{\delta}{\bar{\delta}} < 1$, we have

$$g_\delta(\bar{x}) = \frac{\delta}{\bar{\delta}} g_{\bar{\delta}}(\bar{x}) + (1 - \frac{\delta}{\bar{\delta}})g(\bar{x}) < \frac{\delta}{\bar{\delta}} g_{\bar{\delta}}(x) + (1 - \frac{\delta}{\bar{\delta}})g(x) = g_\delta(x).$$

This shows that $x \in C \backslash S$ cannot be a solution of $(\mathcal{P}(\delta))$. So $S_\delta \subset S$. □

Corollary 2.7. *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* $\bar{y} = 1/\delta$.

Proof. Suppose that $(\bar{x}, \bar{y}) \in C \times R_+$ with $\bar{y} > 0$ is a saddle point of L. Then $\bar{x}$ is a minimizer of $L(\cdot, \bar{y})$ over C. So $\bar{x}$ is an optimal solution of $(\mathcal{P}(\delta))$ with $\delta = 1/\bar{y}$.

Conversely if $S_\delta \cap S \neq \emptyset$, then for any $\bar{x} \in S_\delta \cap S$, then $\bar{x}$ is an optimal solution of minimizing $f(x) + \frac{1}{\delta}g(x)$ over C. Also $\bar{x} \in S$. By Theorem 2.3, $(\bar{x}, 1/\delta)$ is a saddle point of L. This completes the proof. □

To illuminate our achieved results further, we provide below a non-convex program example.

Example 2.8. Let $g(x_1, x_2) = \max\{-x_1 + x_2, 0\}$, $C = \{(x_1, x_2) \mid x_1 \geq 0,\ x_2 \geq 0,\ x_1 + x_2 \leq 4\}$, and $f(x_1, x_2) = -(x_1 - 4)^2 - (x_2 - 4)^2$. Then the solution set S (i.e., $argmin_{x \in C} g(x)$) of $(\mathcal{P})$ is the convex hull of the points $(0, 0)$, $(4, 0)$ and $(2, 2)$ and $p^* = 0$. Since f is a concave function, $(\mathcal{Q})$ is not a convex program. But the optimal value of $(\mathcal{Q})$ is achieved at an extreme point of S due to the concavity of f and the convexity of S. An easy computation shows that $S_Q = \{(0,0)\}$. We can easily verify that, for any $\bar{y} \geq 0$, $(0,0)$ is a minimizer of $L(\cdot, \bar{y})$ over C where $L(x, \bar{y}) = f(x) + \bar{y}(g(x) - p^*)$. By Theorem 2.3, we conclude that $((0,0), \bar{y})$ is a saddle point of L.

Example 2.8 can also be used to verify the conclusions of Theorem 2.6 and Corollary 2.7. The details are omitted here.

Acknowledgment

We thank one referee for the suggestions.

References

[1] F.H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983.

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

[3] R.T. Rockafellar,Lagrange multipliers and optimality, *SIAM Review* 35 (1993) 183–238.

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

Manuscript received 5 July 2010
revised 17 November 2010, 16 May 2011
accepted for publication 27 May 2011

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