On the Existence of Saddle Points for l_1-Minimization Problems

Yuefang Lian, Jinchuan Zhou*, Jingyong Tang, Xia Liu

***Abstract*—The sparse optimization problem has a wide range of applications in image processing, compressed sensing, and machine learning, etc. It is well known that l_1-minimization problem plays an important role in studying sparse optimization problem from theoretical and algorithm aspects. In this paper, we mainly study the existence theory on saddle points for l_1-minimization problem. Firstly, to overcome the nonsmoothness of l_1-norm, we translate l_1-minimization problem to an optimization programming with linear cost function by introducing new variable. Secondly, based on a new augmented Lagrangian function, the relationship on saddle points between the primal problem and the translated problems, associated with their duality problems, is established. It allows us to establish local saddle points by taking into account of second-order sufficient conditions. Finally, global saddle points is established by using two different approaches. One is requiring that the optimal solution is unique. This assumption can be further removed in our another approach by using the perturbation analysis of primal problem.**

***Index Terms*—Saddle points, augmented Lagrangian functions, l_1-minimization problems, dual problem, perturbation analysis.**

I. Introduction

CONSIDER The following l_1-minimization problem

$$
\begin{aligned}
(P) \qquad \min \quad & \|x\|_1 \\
s.t. \quad & g_i(x) \le 0, i = 1, 2, ..., m, \\
& h_j(x) = 0, j = 1, 2, ..., l, \\
& x \in X,
\end{aligned}
$$

where $\|x\|_1 := \sum_{i=1}^{n} |x_i|$, $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \dots, m$, $h_j : \mathbb{R}^n \to \mathbb{R}$ for $j = 1, \dots, l$ are twice differentiable functions, and X is a nonempty closed set in $\mathbb{R}^n$.

The l_1-minimization problem has been attracted a lot of attentions after introducing by Chen, Donoho and Saunders [5] to tackle the NP-hard l_0-minimization arising from signal and imaging processing. How to seek a sparse solution has become a common request in many scientific areas. Hence, due to its capability for locating sparse solutions, l_1-minimization has found numerous applications in pattern recognition, machine learning, computer vision, etc. The relation between l_0- and l_1- minimization, stability of solution sets, reweighted l_1-methods, dual-density-based l_1-methods, and other related theory, algorithm, and applications can be found in [1], [2], [9], [28], [29], [30], [31] and references therein.

The Lagrangian function of (P) is

$$\mathcal{L}(x, \lambda, \mu) := \|x\|_1 + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{l} \mu_j h_j(x),$$

where $\lambda \in \mathbb{R}^m_+$ and $\mu \in \mathbb{R}^l$. The dual problem (D) is

$$
\begin{aligned}
(D) \qquad \max \quad & \theta(\lambda, \mu) := \inf_{x \in X} \mathcal{L}(x, \lambda, \mu) \\
s.t. \quad & \lambda \ge 0.
\end{aligned}
$$

The duality theory provide a theoretical foundation for developing various algorithms which are widely used in practical applications, e.g., [8], [14], [22]. The strong duality theorem (i.e., the zero-duality gap property between the primal and dual problems) can be obtained under convexity assumptions. Unfortunately, a nonzero-duality gap maybe arise for nonconvex programming as using the above Lagrangian. This drawback has been solved by adding an augmented term to the classical Lagrangian function, referred to augmented Lagrangian functions. For example, in [15], the augmented function is requiring to be convex. This assumption imposed on augmented function was further weakened to be nonconvex, level-boundedness, or even valley-at-zero property; see [3], [34] for more information.

In recent years, by introducing different augmented Lagrangian functions, saddle points theory has been established for various types of optimization problems, such as nonlinear programming [6], [16], [23], [24], [33], second-order cone programming [12], [26], [32], semi-definite programming [10], [13], [19], [20], [25], [27], cone programming [7], [18], [35], semi-infinite programming [4], [11], [17]. Compared with the above existing results, it should be pointed out that l_1-minimization belongs to non-smooth optimization problems, due to the non-smoothness of l_1-norm. Hence, the existing results cannot be applied to l_1-minimization directly. The main aim of this research is to fill up this gap, i.e., studying the existence theory on saddle points of l_1-minimization problem (P). Our main contributions are listed as follows.

i) To overcome the non-smoothness caused by l_1-norm, we translate the primal problem (P) to a new problem (P') by introducing a new variable. The main advantage of this transform (P') is that the local saddle points can be established by second-order sufficient conditions.

Manuscript received June 8, 2020; revised October 15, 2020. This work is supported by National Natural Science Foundation of China (11771255, 11801325), Young Innovation Teams of Shandong Province (2019KJI013), Program of Science and Technology Activities for Overseas Students in Henan Province in 2020, and Nanhu Scholars Program for Young Scholars of Xinyang Normal University.

Yuefang Lian is with School of Mathematics and Statistics, Shandong University of Technology, Zibo 255000, P.R. China (e-mail: lyf199603@163.com.)

Jinchuan Zhou (corresponding author) is with School of Mathematics and Statistics, Shandong University of Technology, Zibo 255000, P.R. China (e-mail: jinchuanzhou@163.com.)

Jingyong Tang is with School of Mathematics and Statistics, Xinyang Normal University, Xinyang 464000, Henan, P.R. China (e-mail:tangjy@xynu.edu.cn.)

Xia Liu is with School of Mathematics and Statistics, Shandong University of Technology, Zibo 255000, P.R. China (e-mail: lxia010@163.com.)

ii) Develop the relationship of saddle points between (P) and (P'). Our research shows an interesting fact: the saddle point of (P') can ensure that of (P), while the converse statement maybe false unless some assumptions are added.
iii) Establish the global saddle point by using two different approaches. One is requiring that the optimal solution is unique. This assumption can be removed in our another approach by using the perturbation analysis of primal problem.

The paper are organized as follows. Section II deals with saddle points for l_1-minimization problem with linear constraints. In Section III, we discuss the relationship of saddle point between $\mathcal{L}_1$ and $\mathcal{L}_2$. Section IV studies saddle points of l_1-minimization problem with nonlinear constraints. Conclusion is given in Section V.

II. Saddle points with linear constraints

We first study the following l_1-minimization problem with linear constraints.

Let $B \in \mathbb{R}^{m_1 \times n}, C \in \mathbb{R}^{m_2 \times n}$, $D \in \mathbb{R}^{m_3 \times n}$ be three given matrices with $m_1 + m_2 + m_3 < n$, and $b \in \mathbb{R}^{m_1}, c \in \mathbb{R}^{m_2}$, $d \in \mathbb{R}^{m_3}$ be three vectors, respectively. Consider

$$\min_x \{\|x\|_1 : Bx \geq b, Cx \leq c, Dx = d\}. \quad (1)$$

Denoted by $\mathcal{F}$ the feasible region, i.e.,

$$\mathcal{F} := \{x \in \mathbb{R}^n | Bx \geq b, Cx \leq c, Dx = d\}.$$

Any polyhedron can be represented by finite linear equality and inequality in this way.

At a reference point x^*, some inequalities among $Bx^* \geq b$ and $Cx^* \leq c$ might be binding. Let us use the index sets $\mathcal{A}_1(x^*)$ and $\overline{\mathcal{A}_1}(x^*)$, respectively, to record the binding and non-binding constraints in the first group of the inequalities $Bx \geq b$, i.e.,

$$\mathcal{A}_1(x^*) := \{i : (Bx^*)_i = b_i\},$$

$$\overline{\mathcal{A}_1}(x^*) := \{i : (Bx^*)_i > b_i\},$$

and the index sets for the second group of the inequality $Cx \leq c$, i.e.,

$$\mathcal{A}_2(x^*) := \{i : (Cx^*)_i = c_i\},$$

$$\overline{\mathcal{A}_2}(x^*) := \{i : (Cx^*)_i < c_i\}.$$

By introducing $\alpha \in \mathbb{R}^{m_1}_+$ and $\beta \in \mathbb{R}^{m_2}_+$, (1) takes the form

$$\min_x \left\{ \|x\|_1 : \begin{array}{l} Bx - \alpha = b, Cx + \beta = c, \\ Dx = d, \alpha \geq 0, \beta \geq 0 \end{array} \right\}. \quad (2)$$

The following result is based on complementarity theory of linear programming.

Lemma 1. *(see [28], Lemma 2.4.1). At an optimal solution x^* of the problem (1), there exists α^*, β^* such that*

$$\begin{cases} \alpha_i^* = 0 & \forall\, i \in \mathcal{A}_1(x^*) \\ \alpha_i^* = (Bx^*)_i - b_i > 0 & \forall\, i \in \overline{\mathcal{A}_1}(x^*) \\ \beta_i^* = 0 & \forall\, i \in \mathcal{A}_2(x^*) \\ \beta_i^* = c_i - (Cx^*)_i > 0 & \forall\, i \in \overline{\mathcal{A}_2}(x^*). \end{cases} \quad (3)$$

By introducing $u, v, t \in \mathbb{R}^n_+$, where t satisfies $|x| \leq t$, the problem (2) can be written equivalently as a linear programming

$$\begin{array}{ll} \min\limits_{(x,t,u,v,\alpha,\beta)} & e^T t \\ s.t. & Bx - \alpha = b,\ Cx + \beta = c,\ Dx = d, \\ & x + u - t = 0,\ x - v + t = 0, \\ & (t, u, v, \alpha, \beta) \geq 0. \end{array} \quad (4)$$

The Lagrangian dual problem (4) in terms of the variables $h^{(1)}, ..., h^{(5)}$ is given as follows
(DP)

$$\begin{array}{ll} \max\limits_{(h^1,\ldots,h^5)} & b^T h^3 + c^T h^4 + d^T h^5 \\ s.t. & h^1 + h^2 + B^T h^3 + C^T h^4 + D^T h^5 = 0, \\ & -h^1 + h^2 \leq e, \\ & (h^1, -h^2, -h^3, h^4) \geq 0. \end{array}$$

Here x is the key variable of this problem, because the remaining variables (t, u, v, α, β) can be determined by x. This point is illustrated by the following result.

Lemma 2. *(see [28], Lemma 2.4.4).*

(i) *If $(x^*, t^*, u^*, v^*, \alpha^*, \beta^*)$ is an optimal solution of the problem (4), then*

$$(t^*, u^*, v^*) = (|x^*|, |x^*| - x^*, |x^*| + x^*)$$

and α^*, β^* is in (3).

(ii) *x^* is a solution to the problem (1) if and only if $(x^*, |x^*|, |x^*| - x^*, |x^*| + x^*, \alpha^*, \beta^*)$ is a solution to (4), where (α^*, β^*) is in (3).*

The existence theory of saddle points with linear constraints is given below by using the duality theory of linear programming.

Theorem 1. *The saddle point for l_1-minimization problem with linear constraints (1) exists if and only if primal problem (4) is feasible.*

Proof: Note first that the problem (4) is linear programming and the objective function is bounded from below. Hence the optimal value of primal problem (4) is finite whenever the feasible region is nonempty. It ensures the zero-duality gap between LP problem (4) and dual problem. ■

III. Relation of Saddle Points between $\mathcal{L}_1$ and $\mathcal{L}_2$

In this section, let us consider the saddle point of (P) with nonlinear constraint. Define

$$u(x, t) := x - t \quad \text{and} \quad v(x, t) := -x - t.$$

The problem (P) takes the form

$$(P') \quad \begin{array}{lll} \min\limits_{(x,t)} & e^T t & \\ s.t. & g_i(x) \leq 0, & i = 1, 2, \ldots, m \\ & h_j(x) = 0, & j = 1, 2, \ldots, l \\ & u_k(x, t) \leq 0, & k = 1, 2, \ldots, n \\ & v_k(x, t) \leq 0, & k = 1, 2, \ldots, n \\ & x \in X. & \end{array}$$

Note that the objective function in (P') is linear. The relation between optimal solutions of (P) and (P') is given.

Lemma 3. (i) *If (x^*, t^*) is a solution of (P'), then t^* equals to $|x^*|$.*
(ii) *x^* is a solution of (P) if and only if $(x^*, |x^*|)$ is a solution of $(P^{'})$.*

To deal with non-convex optimization problems, it naturally needs to use an augmented Lagrangian function, instead of the classical Lagrangian. Here we propose two generalized essentially quadratic augmented Lagrangian functions for (P) and $(P^{'})$, respectively,

$$\begin{aligned}&\mathcal{L}_1(x,\lambda,\mu,c)\\ &:= \|x\|_1+\sum_{j=1}^{l}\mu_j h_j(x)+\frac{c}{2}\sum_{j=1}^{l}h_j^2(x)\\ &\quad+\frac{1}{2c}\sum_{i=1}^{m}\Big\{[\phi(cg_i(x),\lambda_i)]_+^2-\lambda_i^2\Big\},\end{aligned}$$

and

$$\begin{aligned}&\mathcal{L}_2(x,t,\lambda,\mu,\xi,\eta,c)\\ &:= e^T t+\sum_{j=1}^{l}\mu_j h_j(x)+\frac{c}{2}\sum_{j=1}^{l}h_j^2(x)\\ &\quad+\frac{1}{2c}\sum_{i=1}^{m}\Big\{[\phi(cg_i(x),\lambda_i)]_+-\lambda_i^2\Big\}\\ &\quad+\frac{1}{2c}\sum_{k=1}^{n}\Big\{[\phi(cu_k(x,t),\xi_k)]_+-\xi_k^2\Big\}\\ &\quad+\frac{1}{2c}\sum_{k=1}^{n}\Big\{[\phi(cv_k(x,t),\eta_k)]_+-\eta_k^2\Big\},\end{aligned}$$

where $(\lambda,\mu,\xi,\eta,c)\in\mathbb{R}_+^m\times\mathbb{R}^l\times\mathbb{R}_+^n\times\mathbb{R}_+^n\times\mathbb{R}_{++}$ and $\phi:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ is required to satisfy

(A_1) convex and twice continuously differentiable,
(A_2) $\phi^{'}(0,0)=(1,1)$, $\phi(0,y)=y,\ \forall y\in\mathbb{R}$,
(A_3) $\phi_x^{'}(x,y)>0,\ \forall y\in\mathbb{R}$.

Clearly, $\phi(x,y):=x+y$ satisfies the above assumptions. In this special case, $\mathcal{L}_i$ for $i=1,2$ reduces to the essentially quadratic augmented Lagrangian.

The Lagrangian dual problem of (P) and $(P^{'})$ are presented as below

$$(D)\qquad \max_{\lambda\geq 0}\theta(\lambda,\mu,c):=\inf_{x\in X}\mathcal{L}_1(x,\lambda,\mu,c),$$

and

$$(D')\qquad \max_{(\lambda,\xi,\eta)\geq 0}\theta(\lambda,\mu,\xi,\eta,c),$$

where

$$\theta(\lambda,\mu,\xi,\eta,c):=\inf_{(x,t)\in X\times\mathbb{R}_+^n}\mathcal{L}_2(x,t,\lambda,\mu,\xi,\eta,c).$$

Note that

$$\begin{aligned}&\mathcal{L}_2(x,t,\lambda,\mu,\xi,\eta,c)\\ =\ & e^T t-\|x\|_1+\mathcal{L}_1(x,\lambda,\mu,c)\\ &+\frac{1}{2c}\sum_{k=1}^{n}\Big[\Big(\phi(cu_k(x,t),\xi_k)\Big)_+^2-\xi_k^2\Big]\\ &+\frac{1}{2c}\sum_{k=1}^{n}\Big[\Big(\phi(cv_k(x,t),\eta_k)\Big)_+^2-\eta_k^2\Big].\end{aligned}\tag{5}$$

The assumption on ϕ ensures the monotonicity of $\phi(x,y)$ in x. Hence if

$$\phi\Big(cu_k(x,|x|),\xi_k\Big)\leq\xi_k,\quad \phi\Big(cv_k(x,|x|),\eta_k\Big)\leq\eta_k,$$

then

$$\Big[\phi\Big(cu_k(x,|x|),\xi_k\Big)\Big]_+^2-\xi_k^2\leq 0$$

and

$$\Big[\phi\Big(cv_k(x,|x|),\eta_k\Big)\Big]_+^2-\eta_k^2\leq 0.$$

Thus

$$\begin{aligned}\mathcal{L}_2(x,|x|,\lambda,\mu,\xi,\eta,c) &\leq \mathcal{L}_1(x,\lambda,\mu,c)\\ &= \mathcal{L}_2(x,|x|,\lambda,\mu,0,0,c).\end{aligned}\tag{6}$$

Definition 1. *A solution $(x^*,\lambda^*,\mu^*)\in X\times\mathbb{R}_+^m\times\mathbb{R}^l$ is a global saddle point of $\mathcal{L}_1$, if there exists some $c>0$ such that for all $(x,\lambda,\mu)\in X\times\mathbb{R}_+^m\times\mathbb{R}^l$,*

$$\mathcal{L}_1(x^*,\lambda,\mu,c)\leq\mathcal{L}_1(x^*,\lambda^*,\mu^*,c)\leq\mathcal{L}_1(x,\lambda^*,\mu^*,c),\tag{7}$$

If the above inequality holds by restricting $(x,\lambda,\mu)\in X\cap N(x^*,\delta)\times\mathbb{R}_+^m\times\mathbb{R}^l$, where $N(x^*,\delta):=\{x\in\mathbb{R}^n|\|x-x^*\|\leq\delta\}$, then (x^*,λ^*,μ^*) is said to be a local saddle point of $\mathcal{L}_1$.

Note that the saddle point is also dependent on the parameter c. But for simplification, we omit it in the following analysis. It does not cause any confusion from the context.

Lemma 4. *If (x^*,λ^*,μ^*) is a local (global) saddle point of $\mathcal{L}_1$, then x^* is a local (global) optimal solution of (P) and*

$$\mathcal{L}_1(x^*,\lambda^*,\mu^*,c)=\|x^*\|_1.$$

Proof: If (x^*,λ^*,μ^*) is a local saddle point of $\mathcal{L}_1$, by definition there exists $\delta>0$ such that for all $x\in x^*+\delta\mathbb{B}$ and $\lambda\in\mathbb{R}_+^m$,

$$\mathcal{L}_1(x^*,\lambda,\mu,c)\leq\mathcal{L}_1(x^*,\lambda^*,\mu^*,c)\leq\mathcal{L}_1(x,\lambda^*,\mu^*,c).\tag{8}$$

If x^* is infeasible, then we need to consider the following two cases.

Case (a): $\exists i\in\{1,\ldots,m\}$ such that $g_i(x^*)>0$. Taking into account of convexity of ϕ, we have $\phi(cg_i(x^*),\lambda_i)\geq cg_i(x^*)+\lambda_i>0$, and hence

$$\begin{aligned}&\frac{1}{2c}\Big[\Big(\phi(cg_i(x^*),\lambda_i)\Big)_+^2-\lambda_i^2\Big]\\ =\ &\frac{1}{2c}\Big[\Big(\phi(cg_i(x^*),\lambda_i)\Big)^2-\lambda_i^2\Big]\\ \geq\ &\frac{1}{2c}\Big[c^2g_i^2(x^*)+2c\lambda_i g_i(x^*))\Big]\\ \to\ &\infty \text{ as } \lambda_i\to\infty.\end{aligned}$$

Case (b): $\exists j\in\{1,2,\ldots,l\}$ such that $h_j(x^*)\neq 0$. Then

$$\frac{1}{2c}\Big[\Big(\phi(cg_i(x^*),\lambda_i)\Big)_+^2-\lambda_i^2\Big]=0$$

and

$$\mu_j h_j(x^*)+\frac{c}{2}h_j(x^*)\to\infty \text{ as } \mu_j\to\infty.$$

The above two cases both yield a contradict with the finiteness of $\mathcal{L}_1(x^*, \lambda^*, \mu^*, c)$. Hence x^* is feasible.
According to the first inequality in (8), we have

$$\mathcal{L}_1(x^*, \lambda^*, \mu^*, c) \geq \mathcal{L}_1(x^*, 0, \mu, c) = \|x^*\|_1,$$

where the equality is due to the feasibility of x^* as shown above. It then further implies

$$\Big(\phi(cg_i(x^*), \lambda^*)\Big)_+^2 \geq (\lambda_i^*)^2. \quad (9)$$

The feasibility of x^* means

$$\Big(\phi(cg_i(x^*), \lambda^*)\Big)_+^2 \leq (\lambda_i^*)^2. \quad (10)$$

Putting (9) and (10) together yields

$$\Big(\phi(cg_i(x^*), \lambda^*)\Big)_+^2 = (\lambda_i^*)^2.$$

Hence

$$\mathcal{L}_1(x^*, \lambda^*, \mu^*, c) = \|x^*\|_1. \quad (11)$$

Let us use the second inequality in (8) to show the local optimality of x^*. For any feasible point x satisfying $x \in x^* + \delta\mathbb{B}$, according to (11) and the second inequality in (8), we have

$$\|x^*\|_1 \leq \mathcal{L}_1(x, \lambda^*, \mu^*, c) \leq \|x\|_1.$$

So x^* is a local optimal solution.

By a similar argument, we can show that x^* is a global optimal solution, provided that (x^*, λ^*, μ^*) is a global saddle point of $\mathcal{L}_1$. ■

The similar argument is applicable to $\mathcal{L}_2$.

Corollary 1. *If $(x^*, t^*, \lambda^*, \mu^*, \xi^*, \eta^*)$ is a local (global) saddle point of $\mathcal{L}_2$, then (x^*, t^*) is a local (global) optimal solution of (P'), and*

$$t^* = |x^*| \;\; and \;\; \mathcal{L}_2(x^*, t^*, \lambda^*, \mu^*, \xi^*, \eta^*, c) = \|x^*\|_1.$$

Proof: Following the argument given in Lemma 4, it is readily obtaining that (x^*, t^*) is a local (global) optimal solution of (P'). Furthermore, according to the special structure of (P'), $t^* = |x^*|$ by Lemma 3. ■

We next turn attention to study the relationship of saddle point between $\mathcal{L}_1$ and $\mathcal{L}_2$.

Theorem 2. *If $(x^*, t^*, \lambda^*, \mu^*, \xi^*, \eta^*)$ is a local (global) saddle point of $\mathcal{L}_2$, then (x^*, λ^*, μ^*) is a local (global) saddle point of $\mathcal{L}_1$.*

Proof: If $(x^*, t^*, \lambda^*, \mu^*, \xi^*, \eta^*)$ is local (global) saddle point, then by Corollary 1 (x^*, t^*) with $t^* = |x^*|$ is a local (global) optimal solution of (P'). Hence, by Lemma 3 (ii) x^* is a local (global) optimal solution of (P). So

$$\begin{aligned} & \left[\Big(\phi(cg_i(x^*), \lambda_i^*)\Big)_+^2 - (\lambda_i^*)^2\right] \\ = & \left[\Big(\phi(cu_k(x^*, t^*), \xi_k^*)\Big)_+^2 - (\xi_k^*)^2\right] \\ = & \left[\Big(\phi(cv_k(x^*, t^*), \eta_k^*)\Big)_+^2 - (\eta_k^*)^2\right] \\ = & \;\; 0. \end{aligned}$$

This implies $\mathcal{L}_1(x^*, \lambda^*, \mu^*, c) = \|x^*\|_1$. Thus

$$\begin{aligned} \mathcal{L}_1(x^*, \lambda, \mu, c) \leq \|x^*\|_1 &= \mathcal{L}_2(x^*, t^*, \lambda^*, \mu^*, \xi^*, \eta^*, c) \\ &\leq \mathcal{L}_2(x, t, \lambda^*, \mu^*, \xi^*, \eta^*, c). \end{aligned}$$

In particular, letting $t = |x|$ we obtain

$$\begin{aligned} \mathcal{L}_1(x^*, \lambda, \mu, c) &\leq \mathcal{L}_2(x, |x|, \lambda^*, \mu^*, \xi^*, \eta^*, c) \\ &\leq \mathcal{L}_1(x, \lambda^*, \mu^*, c), \end{aligned}$$

where the last step is due to (6). ■

However, the converse statement of Theorem 2 needs to modify a little by restricting the region of (x, t).

Theorem 3. *If (x^*, λ^*) is a local (global) saddle point of $\mathcal{L}_1$, then $(x^*, t^*, \lambda^*, \xi^*, \eta^*)$ with $t^* := |x^*|$ and $\xi^* = \eta^* = 0$ is a restricted local (global) saddle point of $\mathcal{L}_2$ over $\Gamma := \{(x, t) | t \geq |x|\}$, i.e.,*

$$\begin{aligned} \mathcal{L}_2(x^*, t^*, \lambda, \mu, \xi, \eta, c) &\leq \mathcal{L}_2(x^*, t^*, \lambda^*, \mu^*, \xi^*, \eta^*, c) \\ &\leq \mathcal{L}_2(x, t, \lambda^*, \mu^*, \xi^*, \eta^*, c), \end{aligned}$$

whenever $(x, t) \in \Gamma$ and $(\lambda, \mu, \xi, \eta) \in \mathbb{R}_+^m \times \mathbb{R}^l \times \mathbb{R}_+^n \times \mathbb{R}_+^n$.

Proof: Since x^* is feasible and $t^* = |x^*|$, we have

$$\Big(\phi(cg_i(x^*), \lambda_i)\Big)_+^2 - \lambda_i^2 \leq 0, \Big(\phi(cu_k(x^*, t^*), \xi_k)\Big)_+^2 - \xi_k^2 \leq 0,$$

$$\Big(\phi(cv_k(x^*, t^*), \eta_k)\Big)_+^2 - \eta_k^2 \leq 0.$$

Note that

$$\Big(\phi(cu_k(x^*, t^*), 0)\Big)_+ = \Big(\phi(cv_k(x^*, t^*), 0)\Big)_+ = 0.$$

Hence

$$\begin{aligned} \mathcal{L}_2(x^*, t^*, \lambda, \mu, \xi, \eta, c) &\leq \mathcal{L}_1(x^*, \lambda^*, \mu^*, c) \\ &= \mathcal{L}_2(x^*, t^*, \lambda^*, \mu^*, 0, 0, c), \end{aligned}$$

On the other hand, since $\xi^* = \eta^* = 0$ and $(x, t) \in \Gamma$, (i.e., $t \geq |x|$), then

$$\Big(\phi(cu_k(x, t), \xi_k^*)\Big)_+^2 - (\xi_k^*)^2 = 0,$$

$$\Big(\phi(cv_k(x, t), \eta_k^*)\Big)_+^2 - (\eta_k^*)^2 = 0.$$

This implies

$$\begin{aligned} \mathcal{L}_2(x, t, \lambda^*, \mu^*, 0, 0, c) &= e^T t + \mathcal{L}_1(x, \lambda^*, \mu^*, c) - \|x\|_1 \\ &\geq \mathcal{L}_1(x^*, \lambda^*, \mu^*, c) \\ &= \mathcal{L}_2(x^*, t^*, \lambda^*, \mu^*, 0, 0, c). \end{aligned}$$

■

The existence of saddle point between $\mathcal{L}_1$ and $\mathcal{L}_2$ is not exactly equivalent to each other. However, we can obtain the saddle point of $\mathcal{L}_2$ by that of $\mathcal{L}_1$ if some added assumptions are improved.

Theorem 4. *If (x^*, λ^*, μ^*) is a local saddle point of $\mathcal{L}_1$ and $I(x^*) := \{k | x_k^* = 0\} = \emptyset$, then $(x^*, t^*, \lambda^*, \mu^*, \xi^*, \eta^*)$ is a local saddle point of $\mathcal{L}_2$ where $t^* := |x^*|$,*

$$\xi_k^* := \begin{cases} 1 & x_k^* > 0, \\ 0 & x_k^* < 0, \end{cases} \quad \text{and} \quad \eta_k^* := \begin{cases} 0 & x_k^* > 0, \\ 1 & x_k^* < 0. \end{cases}$$

Proof: According to the formula of ξ^*, η^*, $t^* = |x^*|$ and the property of ϕ, it is easy to see

$$\Big(\phi(cu_k(x^*, t^*), \xi_k^*)\Big)_+^2 = (\xi_k^*)^2,$$

$$\Big(\phi(cv_k(x^*,t^*),\eta_k^*)\Big)_+^2=(\eta_k^*)^2,\quad\forall\, k=1,\ldots,n.$$

Using (5), we obtain

$$\mathcal{L}_2(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*,c)=\mathcal{L}_1(x^*,\mu^*,\lambda^*,c)=\|x^*\|_1.$$

Furthermore, using (6) yields

$$\mathcal{L}_2(x^*,t^*,\lambda,\mu,\xi,\eta,c)\le\mathcal{L}_2(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*,c).\quad(12)$$

For $k\in\{1,\ldots,n\}$ be fixed and $x_k^*>0$, as (x_k,t_k) near (x_k^*,t_k^*) by using the convexity of ϕ, i.e., $\phi(cu_k(x,t),1)\ge cu_k(x,t)+1$, we have

$$\begin{aligned}
&\frac{1}{2c}\Big[\Big(\phi(cu_k(x,t),\xi_k^*)\Big)_+^2-(\xi_k^*)^2\Big]\\
&+\frac{1}{2c}\Big[\Big(\phi(cv_k(x,t),\eta_k^*)\Big)_+^2-(\eta_k^*)^2\Big]-|x_k|+t_k\\
=\;&\frac{1}{2c}\Big[\Big(\phi(cu_k(x,t),1)\Big)_+^2-1^2\Big]\\
&+\frac{1}{2c}\Big[\Big(\phi(cv_k(x,t),0)\Big)_+^2\Big]-x_k+t_k\\
=\;&\frac{1}{2c}\Big[\Big(\phi(cu_k(x,t),1)\Big)^2-1\Big]-(x_k-t_k)\\
\ge\;&\frac{1}{2c}\Big[\Big(cu_k(x,t)+1\Big)^2-1\Big]-(x_k-t_k)\\
\ge\;&0.
\end{aligned}\quad(13)$$

By symmetrical argument, if $x_k^*<0$, for all (x_k,t_k) near (x_k^*,t_k^*),

$$\begin{aligned}
&\frac{1}{2c}\Big[\Big(\phi(cu_k(x,t),\xi_k^*)\Big)_+^2-(\xi_k^*)^2\Big]+\\
&\frac{1}{2c}\Big[\Big(\phi(cv_k(x,t),\eta_k^*)\Big)_+^2-(\eta_k^*)^2\Big]-|x_k|+t_k\\
&\ge 0.
\end{aligned}\quad(14)$$

Hence

$$\begin{aligned}
\mathcal{L}_2(x,t,\lambda^*,\mu^*,\xi^*,\eta^*,c)&\ge\mathcal{L}_1(x^*,\lambda^*,\mu^*,c)\quad(15)\\
&=\mathcal{L}_2(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*,c),
\end{aligned}$$

in which the first inequality is due to (5), (7), (13), (14). Combining (12) and (15) together yields that $(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*)$ is a local saddle point of $\mathcal{L}_2$ where $t^*:=|x^*|$,

$$\xi_k^*:=\begin{cases}1 & x_k^*>0,\\ 0 & x_k^*<0,\end{cases}\quad\text{and}\quad\eta_k^*:=\begin{cases}0 & x_k^*>0,\\ 1 & x_k^*<0.\end{cases}$$

This completes the proof. ∎

IV. Saddle points with nonlinear constraints

A. Local saddle points

Assumption 1. *(Second-order sufficiency conditions) For $s^*:=(x^*,t^*)$, denote $I(x^*):=\{i|g_i(x^*)=0\}$, $U(x^*,t^*):=\{k|u_k(x^*,t^*)=0\}$ and $V(x^*,t^*):=\{k|v_k(x^*,t^*)=0\}$.*

(i) *$\exists\lambda^*\ge0$, μ^*, $\xi^*\ge0$, and $\eta^*\ge0$ such that*

$$\begin{cases}
0 &= \begin{pmatrix}\sum_{i=1}^m\lambda_i^*\nabla g_i(x^*)+\\ \sum_{j=1}^l\mu_j^*\nabla h_j(x^*)+\xi^*-\eta^*\\ e-\xi^*-\eta^*\end{pmatrix}\\
0 &= \lambda_i^*g_i(x^*),\quad\forall\, i=1,\ldots,m\\
0 &= \xi_k^*u_k(x^*,t^*),\quad\forall\, k=1,\ldots,n\\
0 &= \eta_k^*v_k(x^*,t^*),\quad\forall\, k=1,\ldots,n.
\end{cases}$$

(ii) *The Hessian matrix*

$$\nabla^2_{(x,t)}\mathcal{L}(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*,c)=\begin{pmatrix}(\lambda^*)^T\nabla^2g(x^*)+(\mu^*)^T\nabla^2h(x^*) & 0\\ 0 & 0\end{pmatrix}\quad(16)$$

is positive definite over the following set

$$\left\{d\;\middle|\;\begin{array}{ll}\nabla h_j(x^*)^\top d=0, & j=1,2,\ldots,l,\\ \nabla g_i(x^*)^\top d=0, & i\in J(x^*),\\ \nabla g_i(x^*)^\top d\le0, & i\in J'(x^*),\\ \nabla u_k(x^*,t^*)^\top d=0, & k\in K_1(x^*,t^*),\\ \nabla u_k(x^*,t^*)^\top d\le0, & k\in K_1'(x^*,t^*),\\ \nabla v_k(x^*,t^*)^\top d=0, & k\in K_2(x^*,t^*),\\ \nabla v_k(x^*,t^*)^\top d\le0, & k\in K_2'(x^*,t^*)\end{array}\right\},$$

where

$$\begin{cases}J(x^*):=\{i\in I(x^*)\,|\lambda_i^*>0\},\\ J'(x^*):=\{i\in I(x^*)\,|\lambda_i^*=0\},\\ K_1(x^*,t^*):=\{k\in U(x^*,t^*)\,|\xi_k^*>0\},\\ K_1'(x^*,t^*):=\{k\in U(x^*,t^*)\,|\xi_k^*=0\},\\ K_2(x^*,t^*):=\{k\in V(x^*,t^*)\,|\eta_k^*>0\},\\ K_2'(x^*,t^*):=\{k\in V(x^*,t^*)\,|\eta_k^*=0\}.\end{cases}$$

Similar to the proof in [21, Theorem 2.1] except for some technical details, we can obtain the following result.

Theorem 5. *If Assumption 1 holds at $s^*=(x^*,t^*)$, then $(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*)$ is a local saddle point of $\mathcal{L}_2$.*

B. Global saddle points

According to Theorem 2 the saddle point of $\mathcal{L}_1$ exists whenever the saddle point of $\mathcal{L}_2$ exists. Thus, it firstly needs to study sufficient conditions for existence of saddle point of $\mathcal{L}_2$.

For any constant $\alpha\ge0$, let

$$S(\alpha):=\left\{(x,t)\;\middle|\;\begin{array}{l}x\in X,|h_j(x)|\le\alpha,\ j=1,\ldots,l;\\ g_i(x)\le\alpha,\ i=1,\ldots,m;\\ u_k(x,t)\le\alpha,v_k(x,t)\le\alpha,\ k=1,\ldots,n\end{array}\right\}.$$

Clearly, $S(0)$ is the feasible set of problem (P'). Denote S^* as the set of the optimal solutions to problem (P'). The perturbation function is

$$\beta_f(\alpha):=\inf\{e^Tt|\,(x,t)\in S(\alpha)\}.$$

Clearly, $\beta_f(0)=val(P)$. Let

$$U(\alpha):=\{(x,t)|x\in X,e^Tt\le val(P)+\alpha\}.$$

Throughout the rest of this section, unless stated otherwise, we assume that $S(0)\neq\emptyset$.

In the following analysis, we study global saddle points under two different approaches.

1) ***Unique optimal solution****:* If the primal problem has a unique solution, then we can obtain the following result.

Theorem 6. *Suppose that the primal solution has a unique solution x^*. If X is compact and Assumption 1 holds at x^*, then $(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*)$ is a global saddle point of $\mathcal{L}_2$.*

Proof: According to Assumption 1 and Theorem 5, $\exists c_0>0,\delta>0$ such that $(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*)$ is local saddle point for all $c\ge c_0$, i.e.,

$$\mathcal{L}_2(x^*,t^*,\lambda,\mu,\xi,\eta,c)$$

$$\begin{aligned} &\le \mathcal{L}_2\left(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*,c\right) \\ &\le \mathcal{L}_2\left(x,t,\lambda^*,\mu^*,\xi^*,\eta^*,c\right), \end{aligned} \tag{17}$$

by requiring that (x,t) belongs to the neighborhood $N((x^*,t^*),\delta)$. In what follows we claim that by increasing c the second inequality in (17) holds even if $(x,t)\in X\times\Re^m$.

Since by Lemma 3 (x^*,t^*) is a unique optimal solution of (P'), i.e., $U(0)\bigcap S(0)=x^*$, then

$$\Big[U(0)\backslash N\left((x^*,t^*),\delta\right)\Big]\bigcap\Big[S(0)\backslash N\left((x^*,t^*),\delta\right)\Big]=\emptyset. \tag{18}$$

The compactness of X further ensures the existence of an $\epsilon_1>0$ such that

$$\Big[U(\epsilon_1)\backslash N\left((x^*,t^*),\delta\right)\Big]\bigcap\Big[S(\epsilon_1)\backslash N\left((x^*,t^*),\delta\right)\Big]=\emptyset. \tag{19}$$

In fact, if (19) is invalid, i.e., $\forall\epsilon>0$ one has

$$\Big[U(\epsilon)\backslash N\left((x^*,t^*),\delta\right)\Big]\bigcap\Big[S(\epsilon)\backslash N\left((x^*,t^*),\delta\right)\Big]\neq\emptyset.$$

For $\epsilon_w\to 0$, picking

$$(x^w,t^w)\in\Big[U(\epsilon_w)\backslash N\left((x^*,t^*),\delta\right)\Big]\cap\Big[S(\epsilon_w)\backslash N\left((x^*,t^*),\delta\right)\Big].$$

Since $(x^w,t^w)\in U(\epsilon_w)$ and $(x^w,t^w)\in S(\epsilon_w)$, then $e^Tt^w\le\beta_f(0)+\epsilon_w$, and $t^w_k\ge|x^w_k|$ for $k=1,2\cdots n$. Noting that X is compact, so $\{t^w\}$ is bounded as well. Thus, any accumulation point of $\{(x^w,t^w)\}$ belongs to $U(0)\cap S(0)$ as $\epsilon_w\to 0$, which yields a contraction to (18).

Pick $(x,t)\in\left(X\times\mathbb{R}^n_+\right)\backslash N\left((x^*,t^*),\delta\right)$.

Case (a): $(x,t)\in U(\epsilon_1)\backslash N\left((x^*,t^*),\delta\right)$. So $(x,t)\notin S(\epsilon_1)$ by (19). Hence there exists the following possibilities: $g_{i_0}(x)>\epsilon_1$, or $|h_{j_0}(x)|>\epsilon_1$, or $|u_{k_0}(x,t)|>\epsilon_1$, or $|v_{k_0}(x,t)|>\epsilon_1$. These subcases are further considered below.

Subcase (a)-1: $|h_j(x)|>\epsilon_1$ for some j. Denote $\Omega:=\{1,2,\ldots,l\}$ and $\Omega_k:=\{j\in\Omega\|h_j\left(x\right)|>\epsilon_1\}$. Hence, $\Omega_k\neq\emptyset$. Note that

$$\frac{|\mu^*_j|}{c}\le\frac{1}{4}\epsilon_1 \tag{20}$$

for all $j\in\Omega$ whenever c enough largely. Hence,

$$\begin{aligned} &\mathcal{L}_2(x,t,\lambda^*,\mu^*,\xi^*,\eta^*,c)\\ &\ge \sum_{j=1}^{l}\mu^*_jh_j\left(x\right)+\frac{c}{2}\sum_{j=1}^{l}h^2_j\left(x\right)\\ &\quad-\frac{1}{2c}\left\{\sum_{i=1}^{m}\lambda^{*2}_i+\sum_{k=1}^{n}(\xi^{*2}_k+\eta^{*2}_k)\right\}\\ &\ge -\sum_{j\in\Omega/\Omega_k}|\mu^*_j|\left|h_j\left(x\right)\right|\\ &\quad-\left(\sum_{j\in\Omega_k}(\mu^*_j)^2\right)^{\frac{1}{2}}\left(\sum_{j\in\Omega_k}h^2_j\left(x\right)\right)^{\frac{1}{2}}\\ &\quad+\frac{c}{2}\sum_{j=1}^{l}h^2_j\left(x\right)-\frac{1}{2c}\left\{\sum_{i=1}^{m}\lambda^{*2}_i+\sum_{k=1}^{n}(\xi^{*2}_k+\eta^{*2}_k)\right\}\\ &\ge -\epsilon_1\sum_{j=1}^{l}|\mu^*_j|+c\left(\sum_{j\in\Omega_k}h^2_j\left(x\right)\right)^{\frac{1}{2}}\\ &\quad\times\left[\frac{1}{2}\left(\sum_{j\in\Omega_k}h^2_j\left(x\right)\right)^{\frac{1}{2}}-\left(\sum_{j\in\Omega_k}\frac{\mu^*_j}{c^2}\right)^{\frac{1}{2}}\right]\\ &\quad-\frac{1}{2c}\left\{\sum_{i=1}^{m}\lambda^{*2}_i+\sum_{k=1}^{n}(\xi^{*2}_k+\eta^{*2}_k)\right\}\\ &\ge -\epsilon_1\sum_{j=1}^{l}|\mu^*_j|+\frac{1}{4}c\epsilon^2_1\left|\Omega_k\right| \qquad (21)\\ &\quad-\frac{1}{2c}\left\{\sum_{i=1}^{m}\lambda^{*2}_i+\sum_{k=1}^{n}(\xi^{*2}_k+\eta^{*2}_k)\right\}, \end{aligned}$$

where we have used the non-negativity of $(\phi(cg_i(x),\lambda^*_i))_+$, $(\phi(cu_k(x),\xi^*_k))_+$, $(\phi(cv_k(x),\eta^*_k))_+$, and the last inequality is due to $|h_j(x)|>\varepsilon_1$ for $j\in\Omega_k$ and (20).

Subcase (a)-2: $g_{i_0}(x)>\epsilon_1$ for some i_0 and $|h_j(x)|\le\epsilon_1$ for all j. Then by the convexity of ϕ, we have

$$\begin{aligned} &\mathcal{L}_2(x,t,\lambda^*,\mu^*,\xi^*,\eta^*,c)\\ &\ge -\sum_{j=1}^{l}\epsilon_1|\mu^*_j|+\frac{1}{2c}\Big(\left[\phi(cg_{i_0}(x),\lambda^*_{i_0})\right]_+-\lambda^{*2}_{i_0}\Big)\\ &\quad+\frac{1}{2c}\sum_{i\neq i_0}\left\{[\phi(cg_i(x),\lambda^*_i)]_+-\lambda^{*2}_i\right\}\\ &\quad-\frac{1}{2c}\sum_{k=1}^{n}(\xi^{*2}_k+\eta^{*2}_k)\\ &\ge -\sum_{j=1}^{l}\epsilon_1|\mu^*_j|+\frac{c}{2}g^2_{i_0}(x)+\lambda^*_{i_0}g_{i_0}(x)\\ &\quad-\frac{1}{2c}\left\{\sum_{i=1}^{m}\lambda^{*2}_i+\sum_{k=1}^{n}(\xi^{*2}_k+\eta^{*2}_k)\right\}\\ &\ge -\sum_{j=1}^{l}\epsilon_1|\mu^*_j|+\frac{c}{2}\epsilon^2_1+\lambda^*_{i_0}\epsilon_1\\ &\quad-\frac{1}{2c}\left\{\sum_{i=1}^{m}\lambda^{*2}_i+\sum_{k=1}^{n}(\xi^{*2}_k+\eta^{*2}_k)\right\}\\ &\to \infty,\quad \text{as } c\to\infty. \end{aligned} \tag{22}$$

Subcase (a)-3: There exists a k_0 such that $u_{k_0}(x,t)>\epsilon_1$ and $|h_j(x)|\le\epsilon_1$ for $j=1,\ldots,l$, $g_i(x)\le\epsilon_1$ for $i=1,\ldots,m$. Since ϕ is convex, then $\phi(cu_{k_0}(x),\xi^*_{k_0})\ge cu_{k_0}(x)+\xi^*_k>0$. Hence

$$\begin{aligned} &\mathcal{L}_2(x,t,\lambda^*,\mu^*,\xi^*,\eta^*,c)\\ &\ge -\sum_{j=1}^{l}\epsilon_1|\mu^*_j|+\frac{c}{2}\epsilon^2_1+\xi^*_{k_0}\epsilon_1\\ &\quad-\frac{1}{2c}\left\{\sum_{i=1}^{m}\lambda^{*2}_i+\sum_{k=1}^{n}(\xi^{*2}_k+\eta^{*2}_k)\right\}\\ &\to \infty,\quad \text{as } c\to\infty. \end{aligned} \tag{23}$$

Subcase (a)-4: $v_{k_0}(x,t)>\epsilon_1$ for some k_0. It is similar to above case.

Summarizing the above cases, we know that there exists $c_2\ge c_0$ such that for all $(x,t)\in U_{\epsilon_1}\backslash N\left((x^*,t^*),\delta\right)$

$$\mathcal{L}_2(x,t,\lambda^*,\mu^*,\xi^*,\eta^*,c)\ge\mathcal{L}_2(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*,c),$$

whenever $c\ge c_2$.

Case (b): $(x,t)\in[X\times\mathbb{R}^n_+]\backslash[U_{\epsilon_1}\cup N\left((x^*,t^*),\delta\right)]$. Since

$t^* \geq |x^*|$ and $e^T t > e^T t^* + \epsilon_1$, then

$$\begin{aligned}
&\mathcal{L}_2(x, t, \lambda^*, \mu^*, \xi^*, \eta^*, c) \qquad (24)\\
> \;& \|x^*\|_1 + \epsilon_1 + \sum_{j=1}^{l} \mu_j^* h_j(x) + \frac{c}{2}\sum_{j=1}^{l} h_j^2(x)\\
&-\frac{1}{2c}\left\{\sum_{i=1}^{m} \lambda_i^{*2} + \sum_{k=1}^{n}(\xi_k^{*2} + \eta_k^{*2})\right\}\\
> \;& \|x^*\|_1 + \epsilon_1 - \frac{1}{2c}\sum_{j=1}^{l} \mu_j^*\\
&-\frac{1}{2c}\left\{\sum_{i=1}^{m} \lambda_i^{*2} + \sum_{k=1}^{n}(\xi_k^{*2} + \eta_k^{*2})\right\}.
\end{aligned}$$

Pick

$$c_3 \geq \max\left\{c_0, \frac{\sum_{i=1}^{m} \lambda_i^{*2} + \sum_{k=1}^{n}(\xi_k^{*2} + \eta_k^{*2}) + \sum_{j=1}^{l} \mu_j^{*2}}{2\epsilon_1}\right\}.$$

It then follows from (24) that

$$\begin{aligned}
\mathcal{L}_2(x, t, \lambda^*, \mu^*, \xi^*, \eta^*, c) &\geq e^T x^*\\
&= \mathcal{L}_2(x^*, t^*, \lambda^*, \mu^*, \xi^*, \eta^*, c),
\end{aligned}$$

whenever $c \geq c_3$. Hence the second inequality of (17) holds whenever $(x, t) \in X \times \mathbb{R}_+^n$ as c is larger than $c_2 + c_3$. ■

Combining Theorem 2 and Theorem 6 together yields the following result.

Corollary 2. *Under the assumption of Theorem 6, (x^*, λ^*, μ^*) is a global saddle point of $\mathcal{L}_1(x, \lambda, \mu, c)$.*

*2) **Multiple optimal solutions**:* To remove the restriction on the uniqueness of optimal solutions, we resort to perturbation analysis of the primal problem. Firstly, the following lemmas are needed.

Lemma 5. *If $c_\omega \nearrow +\infty$, then for $\epsilon > 0$ there exists $\omega_\epsilon > 0$ such that*

$$\{(x, t)|\mathcal{L}_2(x, t, \lambda^*, \mu^*, \xi^*, \eta^*, c_\omega) \leq val(P)\} \subseteq S(\epsilon),$$

whenever $\omega \geq \omega_\epsilon$.

Proof: We prove it by contradiction. Suppose that there exist $\epsilon_0 > 0$, $\tilde{N} \subseteq \{1, 2, \ldots, n\}$, (x^ω, t^ω) with $\omega \in N$ such that

$$\begin{aligned}
&\mathcal{L}_2(x^\omega, t^\omega, \lambda^*, \mu^*, \xi^*, \eta^*, c_\omega) \leq val(P),\\
&(x^\omega, t^\omega) \notin S(\epsilon_0).
\end{aligned} \qquad (25)$$

Hence,

$$\begin{aligned}
&val(P) \qquad (26)\\
\geq\;& \mathcal{L}_2(x^\omega, t^\omega, \lambda^*, \mu^*, \xi^*, \eta^*, c_\omega)\\
=\;& e^T t^\omega + \sum_{j=1}^{l} \mu_j^* h_j(x^\omega) + \frac{c_\omega}{2}\sum_{j=1}^{l} h_j(x^\omega)\\
&+\frac{1}{2c_\omega}\sum_{i=1}^{m}\left\{\Big(\phi(c_\omega g_i(x^\omega), \lambda_i^*)\Big)_+^2 - \lambda_i^{*2}\right\}\\
&+\frac{1}{2c}\sum_{k=1}^{n}\left\{\Big(\phi(c_\omega u_k(x^\omega, t^\omega), \xi_k^*)\Big)_+ - (\xi_k^*)^2\right\}\\
&+\frac{1}{2c}\sum_{k=1}^{n}\left\{\Big(\phi(c_\omega v_k(x^\omega, t^\omega), \eta_k^*)\Big)_+ - (\eta_k^*)^2\right\}.
\end{aligned}$$

Recall that $\Omega = \{1, 2, \ldots, l\}$. It follows from (25) that there exists $N_0 \subseteq \tilde{N}$ satisfies one of the following cases.

Case 1: $\Omega_\omega := \{j \in \Omega \| h_j(x^\omega)| > \epsilon_0\} \neq \emptyset$ for $k \in N_0$. As $\omega \in N_0$ sufficiently large,

$$\frac{|\mu_j^*|}{c_\omega} \leq \frac{1}{4}\epsilon_0, \qquad \forall j \in \Omega.$$

This together with (21) and (26) implies that

$$\begin{aligned}
val(P) &\geq \mathcal{L}_2(x^\omega, t^\omega, \lambda^*, \mu^*, \xi^*, \eta^*, c_\omega)\\
&\geq -\epsilon_0\sum_{j=1}^{l}|\mu_j^*| + \frac{1}{4}c_\omega\epsilon_0^2|\Omega_\omega|\\
&\quad -\frac{1}{2c_\omega}\left\{\sum_{i=1}^{m} \lambda_i^{*2} + \sum_{k=1}^{n}(\xi_k^{*2} + \eta_k^{*2})\right\}.
\end{aligned}$$

Taking limit as $\omega \in N_0$ yields $val(P) = +\infty$. This is a contradiction. Thus,

$$|h_j(x^\omega)| \leq \epsilon_0, \quad \forall j \in \Omega = \{1, 2, \ldots, l\}. \qquad (27)$$

Case 2: $g_{i_0}(x_\omega) > \epsilon_0$ for some i_0 and $\omega \in N_0$. It follows from (22), (26), and (27) that

$$\begin{aligned}
val(P) &\geq \mathcal{L}_2(x^\omega, t^\omega, \lambda^*, \mu^*, \xi^*, \eta^*, c_\omega)\\
&\geq -\sum_{j=1}^{l}\epsilon_0|\mu_j^*| + \frac{c_\omega}{2}\epsilon_0^2 + \lambda_{i_0}^*\epsilon_0\\
&\quad -\frac{1}{2c_\omega}\left\{\sum_{i=1}^{m} \lambda_i^{*2} + \sum_{k=1}^{n}(\xi_k^{*2} + \eta_k^{*2})\right\}.
\end{aligned}$$

Taking limits also yielders $val(P) = +\infty$. A contradiction is obtained.

Case 3: there exists k_0 such that $u_{k_0}(x_\omega) > \epsilon_0$ for $\omega \in N_0$. By taking limits and using (23), (26), and (27), we obtain $val(P) = +\infty$. This is a contradiction.

Case 4: there exists k_0 such that $v_{k_0}(x_\omega) > \epsilon_0$ for $\omega \in N_0$. The analysis is similar to the above cases. ■

Lemma 6. *If $c_\omega \nearrow +\infty$, then for $\epsilon > 0$ there exists $\omega_\epsilon > 0$ such that*

$$\{(x, t)|\mathcal{L}_2(x, t, \lambda^*, \mu^*, \xi^*, \eta^*, c_\omega) \leq val(P)\} \subseteq U(\epsilon),$$

whenever $\omega \geq \omega_\epsilon$.

Proof: Let $\Gamma := \left(\sum_{j=1}^{l}|\mu_j^*| + 1\right)$ and $\alpha := \epsilon/(3\Gamma)$. Lemma 5 ensures

$$\{(x, t)|\mathcal{L}_2(x, t, \lambda^*, \mu^*, \xi^*, \eta^*, c_\omega) \leq val(P)\} \subseteq S(\alpha), \qquad (28)$$

and

$$\frac{1}{2c_\omega}\left\{\sum_{i=1}^{m} \lambda_i^{*2} + \sum_{k=1}^{n}(\xi_k^{*2} + \eta_k^{*2})\right\} \leq \frac{\epsilon}{2}, \qquad (29)$$

whenever c_w sufficiently large. Pick (x, t) such that

$$\mathcal{L}_2(x, t, \lambda^*, \mu^*, \xi^*, \eta^*, c_\omega) \leq val(P).$$

Therefore

$$\begin{aligned}
e^T t &= \mathcal{L}_2(x, t, \lambda^*, \mu^*, \xi^*, \eta^*, c_\omega)\\
&\quad -\sum_{j=1}^{l}\mu_j^* h_j(x) - \frac{c_\omega}{2}\sum_{j=1}^{l} h_j^2(x)
\end{aligned}$$

$$
\begin{aligned}
& -\frac{1}{2c_\omega}\sum_{i=1}^{m}\left\{\left(\phi\left(c_\omega g_i(x),\lambda_i^*\right)\right)_+ - \lambda_i^{*2}\right\} \\
& -\frac{1}{2c_\omega}\sum_{k=1}^{n}\left\{\left(\phi\left(cu_k(x,t),\xi_k^*\right)\right)_+ - (\xi_k^*)^2\right\} \\
& -\frac{1}{2c_\omega}\sum_{k=1}^{n}\left\{\left(\phi\left(cv_k(x,t),\eta_k^*\right)\right)_+ - (\eta_k^*)^2\right\} \\
\leq\ & val(P) + \sum_{j=1}^{l}|\mu_j^*|\cdot\left|h_j^2(x)\right| + \frac{1}{2c_\omega}\sum_{i=1}^{m}\lambda_i^{*2} \\
& +\frac{1}{2c_\omega}\sum_{k=1}^{n}\xi_k^{*2} + \frac{1}{2c_\omega}\sum_{k=1}^{n}\eta_k^{*2} \\
\leq\ & val(P) + \alpha\sum_{j=1}^{l}|\mu_j^*| + \frac{\epsilon}{2} \\
\leq\ & val(P) + \epsilon,
\end{aligned}
$$

where the last third inequality is by (28) and (29). ■

The following result is applicable to the multiple optimal solution case.

Theorem 7. *Assume that $S(\alpha_0)\cap U(\alpha_0)$ is bounded for $\alpha_0 > 0$ and that there exists $(\lambda^*,\mu^*,\xi^*,\eta^*)$ such that $(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*)$ is a local saddle point of $\mathcal{L}_2$ for each $(x^*,t^*)\in S^*$. Then, for each $(\bar{x},\bar{t})\in S^*$, $(\bar{x},\bar{t},\lambda^*,\mu^*,\xi^*,\eta^*)$ is a global saddle point of $\mathcal{L}_2$.*

Proof: By assumption, there exists a neighborhood $N((x^*,t^*),\delta^*)$ of (x^*,t^*) such that

$$
\begin{aligned}
& \mathcal{L}_2\left(x^*,t^*,\lambda,\mu,\xi,\eta,c\right) \\
\leq\ & \mathcal{L}_2\left(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*,c\right) \\
\leq\ & \mathcal{L}_2\left(x,t,\lambda^*,\mu^*,\xi^*,\eta^*,c\right). \qquad (30)
\end{aligned}
$$

According to Corollary 1,

$$\mathcal{L}_2\left(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*,c\right) = \|x^*\|_1, \qquad (31)$$

from which and x^* is feasible, it follows

$$
\begin{aligned}
\mathcal{L}_2\left(x^*,t^*,\lambda,\mu,\xi,\eta,c\right) &\leq \|x^*\|_1 \\
&= \mathcal{L}_2\left(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*,c\right).
\end{aligned}
$$

It remains to show that the second inequality of (30) holds true even if x does not belongs to $N(x^*,\delta^*)$ by increasing c.

If there exist $c_\omega \nearrow +\infty$ and $(x^\omega,t^\omega)\in X/N(x^*,\delta^*)\times\mathbb{R}_+^n$ such that

$$
\begin{aligned}
& \mathcal{L}_2\left(x^\omega,t^\omega,\lambda^*,\mu^*,\xi^*,\eta^*,c_\omega\right) \\
<\ & \mathcal{L}_2\left(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*,c_\omega\right) \\
=\ & \|x^*\|_1, \qquad (32)
\end{aligned}
$$

which further implies that (x^ω,t^ω) belongs to the following set

$$\left\{(x,t)|\mathcal{L}_2\left(x^\omega,t^\omega,\lambda^*,\mu^*,\xi^*,\eta^*,c_k\right)\leq\|x^*\|_1\right\}.$$

Given any $\varepsilon\in(0,\alpha_0]$, it follows from Lemma 5 and Lemma 6 that

$$(x^\omega,t^\omega)\in S(\varepsilon)\cap U(\varepsilon).$$

Since $S(\varepsilon)\cap U(\varepsilon)\subset S(\alpha_0)\cap U(\alpha_0)$, then $\{(x^\omega,t^\omega)\}$ is bounded. Hence its any accumulate point, say $(\bar{x},\bar{t})$, satisfies

$$(\bar{x},\bar{t})\in S(\varepsilon)\cap U(\varepsilon).$$

Since $\varepsilon > 0$ is arbitrary, $(\bar{x},\bar{t})\in S(0)\cap U(0) = S^*$, i.e., $e^T\bar{t} = e^Tt^*$.

According to assumption $(\bar{x},\bar{t},\lambda^*,\mu^*,\xi^*,\eta^*)$ is also a local saddle point. By a similar argument as above, it is readily verified that

$$
\begin{aligned}
\mathcal{L}_2\left(\bar{x},\bar{t},\lambda^*,\mu^*,\xi^*,\eta^*,c\right) &= e^T\bar{t} = e^Tt^* \\
&= \mathcal{L}_2\left(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*,c\right),
\end{aligned}
$$

which further implies that for any $(x,t)\in(X\times\mathbb{R}_+^n)\cap N((\bar{x},\bar{t}),\bar{\delta})$

$$\mathcal{L}_2\left(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*,c\right)\leq\mathcal{L}_2\left(x,t,\lambda^*,\mu^*,\xi^*,\eta^*,c\right).$$

Note that $(x^\omega,t^\omega)\in(X\times\mathbb{R}_+^n)\cap N((\bar{x},\bar{t}),\bar{\delta})$. Hence

$$\mathcal{L}_2\left(x^*,t^*,\lambda^*,\mu^*,\xi^*,\eta^*,c_\omega\right)\leq\mathcal{L}_2\left(x^\omega,t^\omega,\lambda^*,\mu^*,\xi^*,\eta^*,c_\omega\right),$$

contradicting (32). ■

Corollary 3. *Under the assumption of Theorem 7, for each $(\bar{x},\bar{t})\in S^*$, $(\bar{x},\lambda^*,\mu^*)$ is a global saddle point of $\mathcal{L}_1$.*

Proof: The desired result follows by combining Theorem 2 and Theorem 7 together. ■

Example 1.

$$
\begin{aligned}
\min\quad & \|x\|_1 \\
s.t.\quad & x_1 + x_2 - 1 = 0 \\
& x_1^2 + x_2^2 \geq 1.
\end{aligned}
$$

The optimal solutions are $x^{*,1} = (1,0)$ and $x^{*,2} = (0,1)$. By introducing variables, we have

$$
\begin{aligned}
\min\quad & e^Tt \\
s.t.\quad & x_1 + x_2 - 1 = 0 \\
& x_1^2 + x_2^2 \geq 1 \\
& x_k - t_k \leq 0,\quad k = 1,2 \\
& -x_k - t_k \leq 0,\quad k = 1,2.
\end{aligned}
$$

For a given $\alpha > 0$,

$$S(\alpha) := \left\{(x,t)\ \middle|\ \begin{array}{l} |x_1+x_2-1|\leq\alpha; \\ x_1^2+x_2^2\geq 1-\alpha; \\ x_k-t_k\leq\alpha, -x_k-t_k\leq\alpha, k=1,2 \end{array}\right\},$$

$$U(\alpha) := \left\{(x,t)|e^Tt\leq 1+\alpha\right\}.$$

It is easy to see that $S(\alpha)\cap U(\alpha)$ is bounded. The KKT conditions are

$$
\left\{
\begin{aligned}
0 &= \lambda^*\begin{pmatrix}1\\1\end{pmatrix} + \mu^*\begin{pmatrix}2x_1\\2x_2\end{pmatrix} + \begin{pmatrix}\xi_1^*\\\xi_2^*\end{pmatrix} - \begin{pmatrix}\eta_1^*\\\eta_2^*\end{pmatrix} \\
0 &= \begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}\xi_1^*\\\xi_2^*\end{pmatrix} - \begin{pmatrix}\eta_1^*\\\eta_2^*\end{pmatrix} \\
0 &= \lambda^*g(x^*) \\
0 &= \xi_k^*u_k(x^*,t^*),\quad \forall k = 1,2 \\
0 &= \eta_k^*v_k(x^*,t^*),\quad \forall k = 1,2.
\end{aligned}
\right.
$$

A common Lagrangian multiplier is $(\lambda^*,\mu^*,\xi^*,\eta^*) = \big(-1,0,(1,1),(0,0)\big)$ at $x^{*,i}$ for $i = 1,2$. Moreover, the second-order sufficiency conditions holds at $(x^{*,i},t^{*,i})(i = 1,2)$. Hence, the assumptions given in Theorem 7 are satisfied. Therefore, $(x^{*,i},t^{*,i},\lambda^*,\mu^*,\xi^*,\eta^*)$ is a global saddle point for $L_i(i = 1,2)$.

V. Conclusions

In this paper, we mainly deal with the existence theory on saddle points of l_1-minimization problems. The local saddle points are established by using the second-order sufficient conditions, while the global saddle points are established by two different approaches depending on whether the solution is unique. Saddle point theory plays an important role in the theoretical analysis for many primal-dual type algorithms. Hence, there are several interesting topics for further research, such as developing augmented Lagrangian multiplier methods for l_1-minimization problems, or studying the exact penalty representation of $\mathcal{L}_i$ for $i = 1, 2$.

References

[1] A. Beck and Y. C. Eldar, "Sparsity constrained nonlinear optimization: optimality conditions and algorithms," *SIAM Journal on Optimization*, vol. 23, pp. 1480-1509, 2013.

[2] J. U. Bouchot, S. Foucart, and P. Hitczenki, "Hard thresholding pursuit algorithms: number of iterations," *Applied and computational harmonic analysis*, vol. 41, pp. 412-435, 2016.

[3] R. S. Burachik and A. Rubinov, "Abstract convexity and augmented Lagrangians," *SIAM Journal on Optimization*, vol. 18, pp. 413-436, 2007.

[4] R. S. Burachik, X. Q. Yang, and Y. Y. Zhou, "Existence of augmented Lagrange multipliers for semi-infinite programming problems," *Journal of Optimization Theory and Applications*, vol. 173, pp. 471-503, 2017.

[5] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33-61, 1998.

[6] M. V. Dolgopolik, "Existence of augmented Lagrange multipliers: reduction to exact penalty functions and localization principle," *Mathematical Programming*, vol. 166, pp. 297-326, 2017.

[7] M. V. Dolgopolik, "Augmented Lagrangian functions for cone constrained optimization: the existence of global saddle points and exact penalty property," *Journal of Global Optimization*, vol. 71, pp. 237-296, 2018.

[8] C. D. Enyi, and M. E. Soh, "Modified gradient-projection algorithm for solving convex minimization problem in Hilbert spaces," *IAENG International Journal of Applied Mathematics*, vol. 44, no. 3, pp. 144-150, 2014.

[9] S. Foucart and H. Rauhut, "A Mathematical introduction to compressive sensing," Springer, New York, 2013.

[10] E. H. Fukuda and B. F. Lourenco, "Exact augmented Lagrangian functions for nonlinear semidefinite programming," *Computational Optimization and Applications*, vol. 71, pp. 457-482, 2018.

[11] N. Q. Huy and D. S. Kim, "Stability and augmented Lagrangian duality in nonconvex semi-infinite programming," *Nonlinear Analysis*, vol. 75, pp. 163-176, 2012.

[12] Y. J. Liu and L. W. Zhang, " Convergence analysis of the augmented Lagrangian method for nonlinear second-order cone optimization problems," *Nonlinear Analysis*, vol. 67, pp. 1359-1373, 2007.

[13] H. Luo, H. Wu, and J. Liu, "On saddle points in semidefinite optimization via separation scheme," *Journal of Optimization Theory and Applications*, vol. 165, pp. 113-150, 2015.

[14] S. Meo, "Nonlinear convex optimization of the energy management for hybrid electric vehicles," *Engineering Letters*, vol. 22, no. 4, pp. 170-182, 2014.

[15] R. T. Rockafellar and J. B. Wets, "Variational Analysis," Springer, Berlin, 1998.

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

[17] J. J. Ruckmann and A. Shapiro, "Augmented Lagrangians in semi-infinite programming," *Mathematical Programming*, vol. 116, pp. 499-512, 2009.

[18] A. Shapiro and J. Sun, "Some properties of the augmented Lagrangian in cone constrained optimization," *Mathematics of Operations Research*, vol. 29, pp. 479-491, 2004.

[19] D. F. Sun, J. Sun, and L. W. Zhang, "The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming," *Mathematical Programming*, vol. 114, pp. 349-391, 2008.

[20] J. Sun, L. W. Zhang, and Y. Wu, "Properties of the augmented Lagrangian in nonlinear semidefinite optimization," *Journal of Optimization Theory and Applications*, vol. 129, pp. 437-456, 2006.

[21] X. L. Sun, D. Li, and K. Mckinnon, "On saddle points of augmented Lagrangians for constrained non-convex optimization," *SIAM Journal on Optimization*, vol. 15, pp. 1128-1146, 2005.

[22] G. S. Wan, X. H. Song, and R. Bettati, "An improved EZW algorithm and its application in intelligent transportation systems," *Engineering Letters*, vol. 22, no. 2, pp. 63-69, 2014.

[23] C. Y. Wang, Q. Liu, and B. Qu," Global saddle points of nonlinear augmented Lagrangian functions," *Journal of Global Optimization* vol. 68, pp. 125-146, 2017.

[24] C. Y. Wang, X. Q. Yang, and X. M. Yang, "Nonlinear augmented Lagrangian and duality theory," *Mathematics of Operations Research* vol. 38, pp. 740-760, 2012.

[25] H. Wu, H. Luo, X. Ding, and G. Chen, "Global convergence of modified augmented Lagrangian methods for nonlinear semidefintie programming," *Computational Optimization and Applications*, vol. 56, pp. 531-558, 2013.

[26] L. W. Zhang, J. Gu, and X. T. Xiao, "A class of nonlinear Lagrangians for nonconvex second-order cone programming," *Computational Optimization and Applications*, vol. 49, pp. 61-99, 2011.

[27] X. Y. Zhao, D. F. Sun, and K. C. Toh, "A Newton-CG augmented Lagrangian method for semidefinite programming," *SIAM Journal on Optimization*, vol. 20, pp. 1737-1765, 2010.

[28] Y. B. Zhao, "Sparse Optimization Theory and Methods," CRC Press, Taylor, Francis Group, 2018.

[29] Y. B. Zhao, "Optimal k-thresholding algorithms for sparse optimization problems," *SIAM Journal on Optimization*, vol. 30, pp. 31-55, 2020.

[30] Y. B. Zhao and M. Kocvara, "A new computational method for the sparsest solutions to systems of linear equations," *SIAM Journal on Optimization*, vol. 25, pp. 1110-1134, 2015.

[31] Y. B. Zhao and Z. Q. Luo, "Constructing new reweighted l_1-algorithms for sparsest points of polyhedral sets," *Mathematics of Operations Research*, vol. 42, pp. 57-76, 2017.

[32] J. C. Zhou and J. S. Chen, "On the existence of saddle points for nonlinear second-order cone programming problems," *Journal of Global Optimization*, vol. 62, pp. 459-480, 2015.

[33] J. C. Zhou, N. H. Xiu, and C. Y. Wang, "Saddle point and exact penalty representation for generalized proximal Lagrangians," *Journal of Global Optimization*, vol. 56, pp. 669-687, 2012.

[34] Y. Y. Zhou and X. Q. Yang, "Augmented Lagrangian functions for constrained optimization problems," *Journal of Global Optimization*, vol. 52, pp. 95-108, 2012.

[35] Y. Y. Zhou, J. C. Zhou, and X. Q. Yang, "Existence of augmented Lagrange multipliers for cone constrained optimization problems," *Journal of Global Optimization*, vol. 58, pp. 243-260, 2014.