COPYRIGHT NOTICE

Federation
UNIVERSITY·AUSTRALIA

FedUni ResearchOnline
http://researchonline.federation.edu.au

Canonical duality for solving general nonconvex constrained problems

Vittorio Latorre, David Yang Gao

Abstract This paper presents a canonical duality theory for solving a general nonconvex constrained optimization problem within a unified framework to cover Lagrange multiplier method and KKT theory. It is proved that if both target function and constraints possess certain patterns necessary for modeling real systems, a perfect dual problem (without duality gap) can be obtained in a unified form with global optimality conditions provided. While the popular augmented Lagrangian method may produce more difficult nonconvex problems due to the nonlinearity of constraints.

1 Introduction

We are interested in solving the following nonconvex constrained minimization problem:

$$
\begin{aligned}
&\min \ f(x) \\
&s.t. \ \ g_i(x) \le 0 \ \ i = 1, \ldots, m \\
&\qquad h_j(x) = 0 \ j = 1, \ldots, p,
\end{aligned}
\tag{1}
$$

where f, g_i and h_j are smooth, real-valued functions on a subset of $\mathbb{R}^n$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, p$. For notational convenience, we use vector form for constraints $g(x)$ and $h(x)$ (without the subscript):

$$
\begin{aligned}
g(x) &= (g_1(x), \ldots, g_m(x)), \\
h(x) &= (h_1(x), \ldots, h_p(x)).
\end{aligned}
$$

V. Latorre
Department of Computer, Control and Management Engineering, University of Rome Sapienza, via Ariosto 25, Rome, Italy
E-mail: latorre@dis.uniroma1.it

D. Y. Gao
School of Science, Information Technology and Engineering, University of Ballarat, Mt Helen, Ballarat, Victoria 3350
E-mail: d.gao@ballarat.edu.au

Therefore, the feasible space can be defined as

$$\mathcal{X}_a := \{x \in \mathbb{R}^n | g(x) \leq 0, h(x) = 0\}.$$

Lagrange multiplier method was originally proposed by J-L. Lagrange from analytical mechanics in 1811 [13]. During the past two hundred years, this method and the associated Lagrangian duality theory have been well-developed with extensively applications to many fields of physics, mathematics and engineering sciences. Strictly speaking, the Lagrange multiplier method can be used only for equilibrium constraints. For inequality constraints, the additional KKT conditions should be considered. In order to solve inequality constrained problems, penalty methods and augmented Lagrangian methods have been studied extensively during the past fifty years (see [17,1,12]). However, these well-developed methods can be used mainly for solving linear inequality constrained problems. For nonlinear constraints, say even the most simple quadratic constraint $\|x\|^2 \leq r$ which is essential for virtually any real-world system [2], the (external) penalty/augmented Lagrangian methods produce a nonconvex term $\frac{1}{2}\alpha(\|x\|^2 - r)_+^2$ in the problem.

Canonical duality theory is potentially powerful methodological method, which was developed originally from nonconvex analysis/mechanics [3,4]. This theory has been used successfully for solving a large class of challenging problems in nonconvex/nonsmooth/discrete systems [5,18,19], recently in network communications [7,15] and radial basis neural networks [14]. It was shown in [10] that both the Lagrange multiplier method and KKT conditions can be unified within a framework of the canonical duality theory. This unified framework leads to an elegant and simple way to handle nonlinear constrained optimization problems. The associated triality theory provides global optimal conditions which can be used to develop efficient algorithms for solving general nonconvex constrained problems (see [8,11]).

The canonical duality theory for solving nonconvex constrained quadratic minimization problem has been discussed in [8]. The main goal of this paper is to demonstrate how to use the canonical duality theory for solving the general non-convex constrained problem (1).

2 Unity for Convex Problems

For a given convex feasible set $\mathcal{E}$, its indicator function $\Psi(\epsilon)$ is defined by

$$\Psi(\epsilon) = \begin{cases} 0 & \text{if } \epsilon \in \mathcal{E} \\ +\infty & \text{otherwise.} \end{cases} \tag{2}$$

The Legendre conjugate of $\Psi(\epsilon)$ is defined by using the Fenchel transformation

$$\Psi^*(\epsilon^*) = \sup_{\epsilon \in \mathcal{E}}\{\epsilon^T \epsilon^* - \Psi(\epsilon)\} \quad \forall \epsilon^* \in \mathcal{E}^*, \tag{3}$$

where $\mathcal{E}^*$ is a dual space of $\mathcal{E}$. Clearly, $\Psi^*(\epsilon^*)$ is convex and lower semi-continuous. By the theory of convex analysis, the following canonical duality relations hold on $\mathcal{E} \times \mathcal{E}^*$:

$$\epsilon^* \in \partial\Psi(\epsilon) \quad \Leftrightarrow \quad \epsilon \in \partial\Psi^*(\epsilon^*) \quad \Leftrightarrow \quad \Psi(\epsilon) + \Psi^*(\epsilon^*) = \epsilon^T \epsilon^*. \qquad (4)$$

A real-valued function is called the *canonical function* if the canonical duality relations (4) hold. Based on the standard canonical dual transformation, we choose the geometrical operator $\xi_0 = \Lambda_0(x) = \{g(x), h(x)\} : \mathbb{R}^n \to \mathbb{R}^2$ and let

$$V_0(\xi_0) = \Psi_1(g) + \Psi_2(h),$$

where

$$\Psi_1(g) = \begin{cases} 0 & \text{if } g \le 0 \\ +\infty & \text{otherwise,} \end{cases} \qquad \Psi_2(h) = \begin{cases} 0 & \text{if } h = 0 \\ +\infty & \text{otherwise,} \end{cases} \qquad (5)$$

the constrained problem (1) can be written in the following canonical form

$$\min \{P(x) = f(x) + V_0(\Lambda_0(x)) | \ \forall x \in \mathbb{R}\}. \qquad (6)$$

By the Fenchel transformation, the conjugate of $V_0(\xi_0)$ can be easily obtained as $V_0^*(\xi_0^*) = \Psi_1^*(\lambda) + \Psi_2^*(\mu)$, where $\xi_0^* = (\lambda, \mu)$ and

$$\Psi_1^*(\lambda) = \sup_{g \in \mathbb{R}^m} \{g^T \lambda - \Psi_1(g)\} = \begin{cases} 0 & \text{if } \lambda \ge 0 \\ +\infty & \text{otherwise,} \end{cases}$$

$$\Psi_2^*(\mu) = \sup_{h \in \mathbb{R}^p} \{h^T \mu - \Psi_2(h)\} = 0 \quad \forall \quad \mu \in \mathbb{R}^p.$$

By using the Fenchel-Young equality $V_0(\xi_0) = \xi_0^T \xi_0^* - V_0^*(\xi_0^*)$ to replace $V_0(\Lambda_0(x))$ in (6), the so called *total complementarity function* in the canonical duality theory can be obtained in the following form

$$\Xi_0(x, \lambda, \mu) = f(x) + [\lambda^T g(x) - \Psi_1^*(\lambda)] + [\mu^T h(x) - \Psi_2^*(\mu)]. \qquad (7)$$

For the indicator $\Psi_1(g)$, the canonical duality relations in (4) lead to

$$\begin{aligned} \lambda_i &\in \partial\Psi_1(g_i) & &\Longrightarrow \lambda \ge 0 & &i = 1, \ldots, m \\ g(x) &\in \partial\Psi_1^*(\lambda) & &\Longrightarrow g_i \le 0 & &i = 1, \ldots, m \\ \lambda^T g(x) &= \Psi_1(g(x)) + \Psi_1^*(\lambda) & &\Longrightarrow \lambda^T g = 0, \end{aligned} \qquad (8)$$

which are the KKT conditions for the inequality constrains $g(x) \le 0$. While for $\Psi_2(h)$, the same relations in (4) lead to

$$\begin{aligned} \mu &\in \partial\Psi_2(h_j) & &\Longrightarrow \mu \in \mathbb{R}^p \\ h_j(x) &\in \partial\Psi_2^*(\lambda) & &\Longrightarrow h_j = 0 & &j = 1, \ldots, p \\ \mu^T h(x) &= \Psi_2(g(x)) + \Psi_2^*(\mu) & &\Longrightarrow \mu^T h = 0. \end{aligned} \qquad (9)$$

From the second and third equation in the (9), it is clear that in order to enforce the constrain $h(x) = 0$, the dual variables μ_i must be not zero for $i = 1, \ldots, p$. This is a special complementarity condition for equality constrains, generally

not mentioned in many textbooks. However, the implicit constraint $\mu \neq 0$ is important in nonconvex optimization. Let $\boldsymbol{\sigma}_0 = (\lambda, \mu)$. The dual feasible spaces should be defined as

$$\mathcal{S}_0 = \{\boldsymbol{\sigma}_0 = (\lambda, \mu) \in \mathbb{R}^{m \times p} | \ \lambda_i \geq 0 \ \forall i = 1, \ldots, m, \ \ \mu_j \neq 0 \ \forall j = 1, \ldots, p\}.$$

Thus, on the feasible space $\mathbb{R}^n \times \mathcal{S}_0$, the total complementary function (7) can be simplified as

$$\Xi_0(x, \boldsymbol{\sigma}_0) = f(x) + \lambda^T g(x) + \mu^T h(x) = \mathcal{L}(x, \lambda, \mu), \tag{10}$$

which is the classical Lagrangian form, and we have

$$P(x) = \sup\{\Xi_0(x, \boldsymbol{\sigma}_0)| \ \forall \boldsymbol{\sigma}_0 \in \mathcal{S}_0\}.$$

This shows that the canonical duality theory is an extension of the Lagrangian theory (actually, the total complementary function was called the extended Lagrangian in [3]). With the canonical duality theory it is possible to formulate the optimality conditions for both inequality and equality constraints in an unified framework.

If f, g are convex and h is linear, the Lagrangian (10) is a saddle function, i.e. $\mathcal{L}(x, \lambda, \mu)$ is convex in the primal variable x and concave(linear) in the dual variables λ and μ. In this case, the Lagrangian dual can be defined by

$$P^*(\lambda, \mu) = \inf_{x \in \mathcal{X}_a} \mathcal{L}(x, \lambda, \mu)$$

on a subspace $\mathcal{S}_a \subset \mathcal{S}_0$ and the saddle Lagrangian duality leads to the following strong duality relation

$$\inf_{x \in \mathcal{X}_a} \mathcal{L}(x, \lambda, \mu) = \sup_{(\lambda,\mu) \in \mathcal{S}_a} P^*(\lambda, \mu).$$

It is well-known that this Lagrangian duality holds only for convex problems. For general nonconvex constrained problems, only the weak duality relation is available, i.e. there is a duality gap between the primal problem and its Lagrangian dual. With the canonical duality theory, it is possible to close the duality gap to obtain global optimal solutions.

3 Sequential Transformation for Nonconvex Problems

In order to solve nonconvex constrained problems in a unified way, the nonconvex functions should be assumed to have certain patterns in order to model real-world problems. In this paper, we need the following assumption.

Assumption 1 *The nonconvex functions f, g_i and h_j for $i = 1, \ldots, m$ and $j = 1, \ldots, p$ can be expressed in the following way:*

$$\begin{aligned} f(x) &= V_f(\Lambda_f(x)) + \tfrac{1}{2}x^T A x - c^T x \\ g_i(x) &= V_{g_i}(\Lambda_{g_i}(x)) \quad i = 1, \ldots, m \\ h_j(x) &= V_{h_j}(\Lambda_{h_j}(x)) \quad j = 1, \ldots, p \end{aligned}$$

where $\xi_f = \Lambda_f(x)$, $\xi_{g_i} = \Lambda_{g_i}(x)$ and $\xi_{h_j} = \Lambda_{h_j}(x)$ are quadratic geometrical operators such that $V_f(\xi_f)$, $V_{g_i}(\xi_{g_i})$, $V_{h_j}(\xi_{h_j})$ are differentiable canonical functions for every $i = 1, \ldots, m$ and $j = 1, \ldots, p$.

Based on this assumption, we can define the following second-level geometrical operators

$$\xi_g = \Lambda_g(x) = (\Lambda_{g_i}, \ldots, \Lambda_{g_m}), \qquad \xi_h = \Lambda_g(x) = (\Lambda_{h_j}, \ldots, \Lambda_{h_p}).$$

Let $\xi_1 = (\xi_f, \xi_g, \xi_h) = \Lambda_1(x)$, $V_g(\xi_g) = \{V_{g_i}(\xi_{g_i})\}$, and $V_h(\xi_h) = \{V_{h_i}(\xi_{h_i})\}$. By Assumption 1, the following duality relations are invertible on their domains, respectively,

$$\xi_f^* = \nabla V_f(\xi_f), \;\; \xi_g^* = \nabla V_g(\xi_g), \;\; \xi_h^* = \nabla V_h(\xi_h). \tag{11}$$

Also, the Legendre conjugates $V_f^*(\xi_f^*)$, $V_g^*(\xi_g^*)$ and $V_h^*(\xi_h^*)$ can be defined uniquely.

Denote $\boldsymbol{\sigma}_1 = (\sigma_f, \sigma_g, \sigma_h) = (\xi_f^*, \xi_g^*, \xi_h^*)$ and let $\mathcal{S}_1$ be a domain such that on which, the inverse duality relations (11) hold. By using the Fenchel-Young equalities, the first-level total complementary function Ξ_0 (10) can be written in the following second-level form:

$$\begin{aligned}\Xi_1(x, \boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) &= \Lambda_f(x)\sigma_f - V_f^*(\sigma_f) + \lambda^T(\Lambda_g(x) \circ \sigma_g - V_g^*(\sigma_g)) \\ &\quad + \mu^T(\Lambda_h(x) \circ \sigma_h - V_h^*(\sigma_h)) - U(x),\end{aligned} \tag{12}$$

where $U(x) = c^T x - \frac{1}{2}x^T A x$, and the symbol $\circ$ indicates the Hadamard product between the primal and dual variables, i.e.,

$$\xi_g \circ \sigma_g = (\xi_{g_1}\sigma_{g_1}, \ldots, \xi_{g_m}\sigma_{g_m}).$$

Based on (12), the canonical dual function can be obtained by

$$P^d(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) = U^\Lambda(\lambda, \mu, \sigma) - \left(V_f^*(\sigma_f) + \lambda^T V_g^*(\sigma_g) + \mu^T V_h^*(\sigma_h)\right), \tag{13}$$

where $U^\Lambda(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1)$ is the Λ-conjugate of $U(x)$ defined by (see [3])

$$U^\Lambda(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) = \mathrm{sta}\{\Lambda_f(x)\sigma_f + \lambda^T(\Lambda_g(x) \circ \sigma_g) + \mu^T(\Lambda_h(x) \circ \sigma_h) - U(x) : x \in \mathbb{R}^n\} \tag{14}$$

Let $\mathcal{S}_a \subset \mathcal{S}_0 \times \mathcal{S}_1$ be the canonical dual feasible space such that on which, $U^\Lambda(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1)$ is well-defined. The canonical dual problem can be proposed as

$$(\mathcal{P}^d): \quad \mathrm{sta}\{P^d(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1): \;\; (\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) \in \mathcal{S}_a\}.$$

Theorem 1 *(Complementarity Dual Principle) Suppose that the point $(\bar{x}, \bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1)$ is a critical point for the total complementarity function (12), then $\bar{x}$ is a KKT point of the primal problem (1), $(\bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1)$ is a KKT point of the dual problem (13) and*

$$P(\bar{x}) = \Xi_1(\bar{x}, \bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1) = P^d(\bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1)$$

$$x \qquad\qquad\qquad x^*$$

$$\Lambda_1(x) \Bigg\downarrow \qquad\qquad \Bigg\uparrow \nabla\Lambda_1^*$$

$$(\xi_f, \xi_g, \xi_h) = \xi_1 \xleftarrow{\quad \xi_1^* = \nabla V_1(\xi_1) \quad} \xi_1^* = \boldsymbol{\sigma}_1 = (\sigma_f, \sigma_g, \sigma_h)$$

$$\Lambda_0 \Bigg\downarrow \qquad\qquad \Bigg\uparrow \xi_1^* = \boldsymbol{\sigma}_1(\boldsymbol{\sigma}_0)$$

$$(g, h) = \xi_0 \xleftarrow{\quad \xi_0^* \in \partial V_0(\xi_0) \quad} \xi_0^* = \boldsymbol{\sigma}_0 = (\lambda, \mu)$$
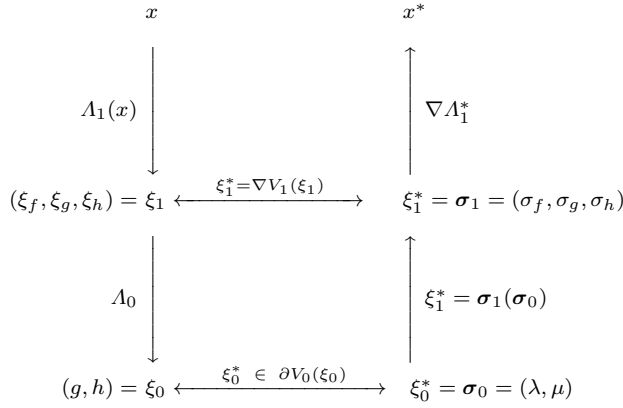
Fig. 1: The scheme of the sequential canonical dual transformation.

Proof . If $(\bar{x}, \bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1)$ is a critical point for the total complementarity function (12) then it must satisfy the following first order conditions:

$$
\begin{aligned}
\nabla_x \Xi_1(\bar{x}, \bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1) =\ & \nabla\Lambda_f(\bar{x})\bar{\sigma}_f + \bar{\lambda}^T (\nabla\Lambda_g(\bar{x}) \circ \bar{\sigma}_g) \\
& + \bar{\mu}^T (\nabla\Lambda_h(\bar{x}) \circ \bar{\sigma}_h) + Ax - c = 0, \\
\nabla_{\sigma_f} \Xi_1(\bar{x}, \bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1) =\ & \Lambda_f(\bar{x}) - \nabla V_f^*(\bar{\sigma}_f) = 0, \\
\nabla_{\sigma_g} \Xi_1(\bar{x}, \bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1) =\ & \Lambda_g(\bar{x}) - \nabla V_g^*(\bar{\sigma}_h) = 0, \\
\nabla_{\sigma_h} \Xi_1(\bar{x}, \bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1) =\ & \Lambda_h(\bar{x}) - \nabla V_h^*(\bar{\sigma}_h) = 0.
\end{aligned}
\tag{15}
$$

The last three conditions in the (15) are equivalent to

$$\bar{\sigma}_f = \nabla V_f\left(\Lambda_f(\bar{x})\right), \quad \bar{\sigma}_g = \nabla V_g\left(\Lambda_g(\bar{x})\right), \quad \bar{\sigma}_h = \nabla V_h\left(\Lambda_f(\bar{x})\right).$$

By substituting these conditions in the first equation of the (15) and using the chain rule of derivation on f, g_i and h_j for every $i = 1, \ldots, m$ and $j = 1, \ldots, p$, we obtain

$$\nabla f(\bar{x}) + \bar{\lambda}^T \nabla g(\bar{x}) + \bar{\mu}^T \nabla h(\bar{x}) = \nabla\mathcal{L}(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0.$$

This condition plus the conditions coming from the (8) prove that $\bar{x}$ is a KKT point for the (1). Furthermore, from these complementarity conditions we obtain that $f(\bar{x}) = \Xi_1(\bar{x}, \bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1)$.

The first equation of the (15) leads to the satisfaction of the stationarity condition (14) that is:

$$U^\Lambda(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) = \Lambda_f(x)\sigma_f + \lambda^T \left(\Lambda_g(x) \circ \sigma_g\right) + \mu^T \left(\Lambda_h(x) \circ \sigma_h\right) - U(x).$$

This together with the property that the first order conditions of the dual are equivalent to the last three conditions of the (15) proves that $(\bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1) = (\bar{\lambda}, \bar{\mu}, \bar{\sigma})$ is a KKT point of the dual and $\Xi_1(\bar{x}, \bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1) = P^d(\bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1)$. $\quad\square$

This theorem shows that with the canonical duality theory and the sequential canonical dual transformation it is possible to close the duality gap between the nonconvex primal problem and its canonical dual problem.

4 Global Optimality Solutions

In order to have conditions for the global minimum of the original constrained problem (1), we make the following assumptions

Assumption 2 *The canonical functions $V_f(\xi_f)$, $V_{g_i}(\xi_{g_i})$, and $V_{h_j}(\xi_{h_j})$ are convex for all $i = 1, \ldots, m$ and $j = 1, \ldots, p$. Furthermore, for any Lagrange multiplier $\mu \in \mathbb{R}^p$, we assume that*

$$\mu^T h(x) > -\infty \quad \forall x \in \mathbb{R}^n.$$

Since Ξ_1 is a quadratic function of x, its Hessian matrix is x-free and can be defined by $G(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) = \nabla_x^2 \Xi_1(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1)$. Let

$$\mathcal{S}_a^+ = \{(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) \in \mathcal{S}_a |\ G(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) \succ 0,\ \mu_i > 0 \ \forall i = 1, \ldots, p\}. \quad (16)$$

Theorem 2 *(Global Optimality Conditions) Suppose that Assumptions 1 and 2 are satisfied, and $\mathcal{S}_a^+$ is convex. Then if the point $(\bar{x}, \bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1)$ is a critical point of the Ξ_1 and $(\bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1) \in \mathcal{S}_a^+$, then $(\bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1)$ is the global maximizer of P^d on $\mathcal{S}_a^+$ and $\bar{x}$ is the global minimizer of P on $\mathcal{X}_a$, that is*

$$P(\bar{x}) = \min_{x \in \mathcal{X}_a} P(x) = \max_{(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) \in \mathcal{S}_a^+} P^d(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) = P^d(\bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1)$$

Proof . By Assumption 2, the functions $V_f(\xi_f)$, $V_g(\xi_g)$ and $V_h(\xi_h)$ are convex. This implies that their Legendre conjugates are also convex. Because of the positivity of both λ and μ, the total complementarity function Ξ_1 is concave in the dual variables σ_f, σ_g and σ_h. Also these variables are decoupled. This implies that the following relation

$$\max_{\sigma_f} \max_{\sigma_g} \max_{\sigma_h} \Xi_1(x, \boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) = \max_{\boldsymbol{\sigma}_1} \Xi_1(x, \boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1),$$

is always verified in $\mathcal{S}_1$. By the fact that Ξ_0 is linear in both λ and μ we have

$$\max_{(\lambda,\mu) \in \mathcal{S}_0} \max_{\boldsymbol{\sigma}_1 \in \mathcal{S}_1} \Xi_1(x, \lambda, \mu, \boldsymbol{\sigma}_1) = \max_{(\lambda,\mu) \in \mathcal{S}_0} \mathcal{L}(x, \lambda, \mu) = \begin{cases} P(x) \text{ if } & x \in \mathcal{X}_a \\ \infty & \text{otherwise.} \end{cases}$$

Furthermore if $(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) \in \mathcal{S}_a^+$, then the total complementarity function is convex in x and concave in $\boldsymbol{\sigma}_1$. For this reason the min and max statements can be exchanged in the total complementarity function and we obtain

$$\begin{aligned} \min_{x \in \mathcal{X}_a} P(x) &= \min_{x \in \mathbb{R}^n} \max_{(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) \in \mathcal{S}_a^+} \Xi_1(x, \boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) \\ &= \max_{(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) \in \mathcal{S}_a^+} \min_{x \in \mathbb{R}^n} \Xi_1(x, \boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) = \max_{(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) \in \mathcal{S}_a^+} P^d(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1). (17) \end{aligned}$$

This proves the theorem. $\quad \square$

Remark 1 Since the geometrical operator $\Lambda_1(x)$ is a quadratic vector-valued function of x, by Assumption 1, the canonical dual function $P^d(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1)$ can be written in the following standard form:

$$P^d(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) = -\frac{1}{2}\mathbf{F}(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1)G^{-1}(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1)\mathbf{F}(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) - V^*(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1), \qquad (18)$$

where $V^*(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) = \left(V_f^*(\sigma_f) + \lambda^T V_g^*(\sigma_g) + \mu^T V_h^*(\sigma_h) \right)$, and $\mathbf{F}(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) \in \mathbb{R}^n$ depends on the linear terms in $\Lambda_1(x)$ and in $U(x)$ (see, for example the Eqn (81) in [9]). By the fact that the canonical dual variables $\boldsymbol{\sigma}_0$ and $\boldsymbol{\sigma}_1$ are generally not independent (see Eqn (4.26) in [3]), even if $P^d(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1)$ is concave in $\boldsymbol{\sigma}_0$ and $\boldsymbol{\sigma}_1$ respectively, it may not be concave in $(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1)$ on $\mathcal{S}_a^+$. Detailed studies on the convexity of $P^d(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1)$ for polynomial optimization and neural network problems have been discussed in [6,14]

Remark 2 Similarly to Theorem 2, it is possible to find global maximum conditions by defining

$$\mathcal{S}_a^- = \{(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) \in \mathcal{S}_a|\ G(\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1) \prec 0,\ \mu_i < 0\ \forall i = 1, \ldots, p\}.$$

Thus, if $(\bar{x}, \bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1)$ is a critical point of the function Ξ_1 and such that $(\bar{\boldsymbol{\sigma}}_0, \bar{\boldsymbol{\sigma}}_1)$ is the global minimizer of P^d in $\mathcal{S}_a^-$, then $\bar{x}$ is the biggest local maximizer of P on $\mathcal{X}_a$.

In particular, if the problem is only composed of a quadratic objective function and equality constraints, it is possible to put together these conditions in order to find both the global minimum and global maximum.

Example 1. Let us consider the following one-dimensional constrained problem

$$\min\left\{ \frac{1}{2}qx^2 - cx\ \mid\ s.t.\quad \frac{1}{2}\left(\frac{1}{2}x^2 - d\right)^2 - e = 0 \right\}. \qquad (19)$$

Since the constraint $h(x)$ is a fourth-order polynomial (double well function), we let $\Lambda_h(x) = \frac{1}{2}x^2$, the canonical dual function can be obtained as

$$P^d(\mu, \sigma) = -\frac{c^2}{2(q + \mu\sigma)} - \mu\left(\frac{1}{2}\sigma^2 + \sigma d + e\right).$$

In this particular example with only one equality constraint, we have $\boldsymbol{\sigma}_0 = \mu,\ \lambda = 0$ and $\boldsymbol{\sigma}_1 = \sigma$. If we let $q = 1$, $c = 1$, $d = 6$, $e = 15$, there are total four KKT points as reported in Table 1. It is easy to see that there is no duality gap between the solutions of the primal and the dual problems just as reported in Theorem 1. From the values of the multipliers μ at the optimum, we can say that the first two critical points are the solutions of the minimization problem, while the last two are the solutions for the maximization problem. If we check to which domain the solutions belong, we have that (μ_1, σ_1) is the global maximum in $\mathcal{S}_a^+$ while (μ_4, σ_4) is the local minimum in $\mathcal{S}_a^-$. This means that x_1 is the global minimum of the original constrained problem, while x_4 is the biggest local maximum of the original constrained problem. This example

	x	μ	σ	$f(x)$	$P^d(\mu,\sigma)$	$G(\mu,\sigma)$
(x_1,μ_1,σ_1)	1.023	0.004	-5.48	-0.5	-0.5	0.98
(x_2,μ_2,σ_2)	-1.023	0.36	-5.48	1.55	1.55	-0.98
(x_3,μ_3,σ_3)	4.791	-0.14	5.48	6.69	6.69	0.21
(x_4,μ_4,σ_4)	-4.791	-0.22	5.48	16.27	16.27	-0.21

Table 1: Critical points of the primal and dual problems for example (19) with $q = 1$, $c = 1$, $d = 6$, $e = 15$.

shows once again that thanks to canonical duality theory, not only we are able to close the gap created by dropping the convexity assumptions in the Lagrangian function, but we are also able to obtain the conditions for finding the global minimum.

5 Augmented Lagrangian

We want to compare the approach of the Lagrangian with the one of augmented Lagrangian by using canonical duality theory. We will consider the problem only with one non-convex equality constraint $h(x) = 0$ (i.e. $p = 1$):

$$\mathcal{L}_\nu(x,\mu;\nu) = f(x) + \mu h(x) + \frac{1}{2\nu}\|h(x)\|^2, \tag{20}$$

Where ν is a penalty parameter. The principal framework of Augmented Lagrangian consists in solving a sequence of sub-problems with both the penalty constant ν_k and the Lagrangian multiplier μ_k fixed. At each iteration, a local minimum in x of the function (20) with fixed μ_k is found. The penalty constant is generally updated by $\nu_{k+1} = \alpha\nu_k$ with $\alpha \in (0,1)$, while the multipliers are updated in the following way:

$$\mu_{k+1} = \mu_k + \frac{h(x)}{\nu}. \tag{21}$$

Then a new sub-problem with updated parameters is generated and a new iteration begins.

We analyze both the general case in which μ is considered as variable and the sub-problem in which μ_k is fixed. Differently from the augmented Lagrangian approach, with canonical duality theory it is possible to consider μ as a variable.

Similarly with the previous sections we make the assumption that every equality constraint can be written in the following way

$$h(x) = V_h(\xi_h) = V_h(\Lambda_h(x)),$$

where V_h is convex canonical function and Λ_h is a quadratic operator. The augmented Lagrangian can be written as:

$$\mathcal{L}_\nu(x,\mu) = f(x) + \mu V_h(\Lambda_h(x)) + \frac{1}{2\nu}\|V_h(\Lambda_h(x))\|^2, \tag{22}$$

This function is different than the Lagrangian, as the penalty term adds a further level of complexity, but with a simple canonical transformation we can go back to a form similar to the (10). We choose as non-linear operator $\xi_0 = h(x)$ and by following the same procedure for canonical duality transformation in the previous sections we obtain:

$$V_0(\xi_0) = \frac{1}{2\nu}\xi_0^2, \quad \tau = \nabla V_0(\xi_0) = \frac{\xi_0}{\nu}, \quad V_0^*(\tau) = \frac{\tau^2\nu}{2}. \quad (23)$$

It is important to notice that the dual variable τ at the optimum has the value of the increment that should be applied to μ at every iteration as described in the (21). By using the Fenchel-Young equality we obtain:

$$\Xi_0^\nu(x, \mu, \tau) = f(x) + (\mu + \tau)^T V_h(\Lambda_h(x)) - V_0^*(\tau) \quad (24)$$

This formula is similar in its structure to the (10). By looking at the (24), it is clear that because of the assumptions made on the constrains $h(x)$, the quantity $(\mu+\tau)$ must be positive in order to ensure that $\Xi_0^\nu(x, \mu, \tau)$ is bounded below in x. Furthermore, the quantity $(\mu + \tau)$ must not be zero otherwise the constrain would be ignored. By using the same procedure showed in the previous section we obtain:

$$\Xi_1^\nu(x, \mu, \tau, \sigma) = \Lambda_f(x)\sigma_f - V_f^*(\sigma_f) + (\mu+\tau)^T(\Lambda_h(x)\sigma_h - V_h^*(\sigma_h)) - V_0^*(\tau) - U(x) \quad (25)$$

and the dual formulation is:

$$P_\nu^d(\mu, \tau, \sigma) = U^\Lambda(\mu, \sigma) - \left(V_f^*(\sigma_f) + (\mu + \tau)^T V_h^*(\sigma_h) + \frac{\tau^2\nu}{2}\right) \quad (26)$$

Remark 3 The complementary-dual principle proved in Theorem 1 for the Lagrangian function can be easily extended to the critical points of $\mathcal{L}_a^\nu(x, \mu)$ and $P_\nu^d(\mu, \tau, \sigma)$ as well.

Theorem 3 *If $(\bar{\mu}, \bar{\tau}, \bar{\sigma})$ is a critical point for $P_\nu^d(\mu, \tau, \sigma)$, then $\bar{\tau} = 0$. Furthermore we have*

$$P_\nu^d(\bar{\mu}, 0, \bar{\sigma}) = P^d(\bar{\mu}, \bar{\sigma}),$$

that is P_ν^d and P^d are equivalent in their stationary points and Theorem 2 can be applied to find the global minimum.

Proof From the second conditions in the (23) we have that in critical points $\bar{\tau} = \frac{h(\bar{x})}{\nu}$. As $(\bar{x}, \bar{\mu})$ is a feasible KKT point with associated multipliers $\bar{\mu}$, we have that $\frac{h(\bar{x})}{\nu} = 0$. If $\bar{\tau}$ is zero for every critical point, then by plugging this value in every τ of the (26) we obtain the (13). $\square$

Remark 4 Theorem 3 shows that, from canonical duality point of view, the use of the penalty term is not necessary in the problems considered in this paper because it increases both the complexity of the primal problem and the dimensionality of the dual problem. By solving the dual problem in both the Lagrange multiplier μ and dual variable σ it is possible to find the global solution of the original problem.

5.1 Solution to the Sub-Problem

Like we have stated in the previous section, the strategy of the augmented Lagrangian creates a succession of sub-problems with solutions are convergent to a stationary point of $\mathcal{L}(x, \mu)$. In these sub-problems both μ and ν are fixed to certain values and then updated once the sub-problem is solved and before a new iteration starts. In this section we want to apply canonical duality theory to the subproblem. The primal problem is

$$\mathcal{L}_{\nu,\mu_k}(x) = f(x) + \mu_k h(x) + \frac{1}{2\nu}\|h(x)\|^2,$$

with associated dual similar to the (26), that is

$$P_{\nu,\mu_k}^d(\tau, \sigma) = U^\Lambda(\mu_k, \sigma) - \left(V_f^*(\sigma_f) + (\mu_k + \tau)^T V_h^*(\sigma_h) + \frac{\tau^2\nu}{2} \right)$$

We also define the following matrix:

$$G(\tau, \sigma) = \nabla_x^2 \Xi_1^{\nu,\mu_k}(x, \tau, \sigma),$$

where $\Xi_1^{\nu,\mu_k}(x, \tau, \sigma)$ is the total complementarity function that connects the primal and dual problem that can be easily obtained by the (25). Let

$$\mathcal{S}_{a,\mu_k}^+ = \{(\tau, \sigma) \in \mathcal{S}_a |\ G(\tau, \sigma) \succ 0\}. \tag{27}$$

In this case the solution of the sub-problem $\mathcal{L}_\nu(x, \mu_k)$ are not KKT points of the original problem (1) and Theorem 3 cannot be applied due to the additional penalty term. By the canonical duality we have the following Corollary.

Corollary 1 *Suppose that the point $\bar{x}$ is a stationary point of $\mathcal{L}_{\nu,\mu_k}(x)$, then $\bar{x}$ has a corresponding $(\bar{\tau}, \bar{\sigma})$ that is a stationary point of the P_{ν,μ_k}^d and*

$$\mathcal{L}_{\nu,\mu_k}(\bar{x}) = P_{\nu,\mu_k}^d(\bar{\tau}, \bar{\sigma}).$$

Furthermore if $\mu_k + \tau > 0$ and $(\bar{\tau}, \bar{\sigma}) \in \mathcal{S}_a^+$ then $\bar{x}$ is the global minimizer of $\mathcal{L}_\nu(x, \mu_k)$.

Proof This proof is similar to those of Theorem 1 and Theorem 2 and can be omitted. □

Because of this Corollary, it is possible to find the global solution x^* to $\mathcal{L}_{\nu,\mu_k}(x)$ for any value of ν and μ_k. Furthermore, as $\tau = \frac{h(x)}{\nu}$, it is possible to update the current value of the multiplier $\mu_{k+1} = \mu_k + \tau^*$, where τ^* is the dual variable corresponding to x^*, to get closer to the Lagrangian multiplier μ^* of the global solution.
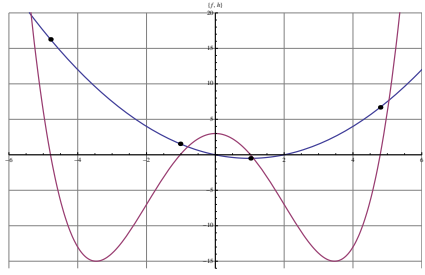
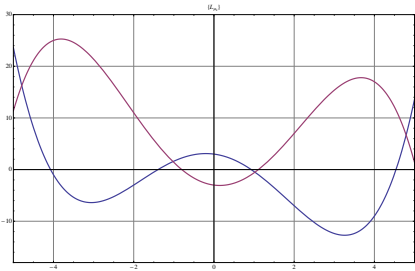Fig. 2: Graphs of the target function (blue) and constrain (red) with KKT points highlighted

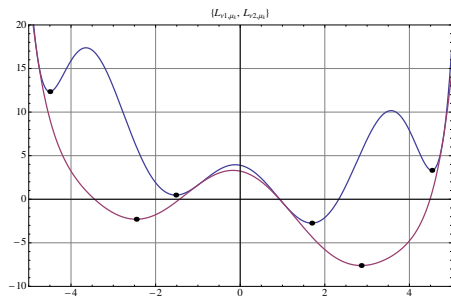Fig. 3: Comparison for the La-grangian with positive (blue) and negative multipliers(red)

Fig. 4: Augmented Lagrangian for $\mu = 5$ (blue) and $\mu = 20$ (red).

5.2 Sub-Problem Example

In this subsection we study the same example already proposed in Section 4 but with the augmented Lagrangian. First we show how the penalty term, in the case of non-convex constraints, greatly increases the complexity of the problem. From Figure 2 it is possible to see the target function and the constrain. The black dots in the picture highlight the four KKT points for this problem. Figure 3 shows the Lagrangian function for positive multiplier $\mu = 1$ and negative multiplier $\mu = -1$. In both cases we observe the presence of a double well. In the case of positive multiplier there are the two local minima, while in the case of negative multipliers the two local maxima can be seen.

Finally in Figure 4 two augmented Lagrangian functions are shown. The blue function has a relatively smaller value of the penalty parameter, $\nu = 5$, while the red function has a big value of the penalty parameter, $\nu = 20$. The small values ν produce nonconvex augmented Lagrangian, and the points corresponding to local maxima of the original problem are made into local minima by the penalty term. This produces much more difficulties in numerical computation for finding the global optimal solution.

We have already showed in section 4 that the canonical duality theory is able to find the global minimum of the Lagrangian function, and at the

beginning of this section we showed that the same solution is valid if the dual problem of the augmented Lagrangian is solved with also considering μ as a variable. Now we show the results for the dual when μ_k is fixed, by Corollary 1 the global solution of the sub-problem can be found.

We solve the problem of the augmented Lagrangian with the same parameters of the problem in Section 4, with $\mu = 1$ and $\nu = 5$. The function in blue of Figure 4 is the problem we want to solve. In this case the dual is:

$$P^d_{\nu,\mu_k}(\tau,\sigma) = -\frac{c^2}{2\left(q + (\mu_k + \tau)\sigma\right)} - (\mu_k + \tau)\left(\frac{1}{2}\sigma^2 + \sigma d + e\right) - \frac{\tau^2\nu}{2}.$$

Table 2 lists all critical points of the primal problem and the dual problem.

	x	τ	σ	$\mathcal{L}_{\nu,\mu_k}(x)$	$P^d_{\nu,\mu_k}(\tau,\sigma)$	$G(\tau,\sigma)$	$(\mu + \tau)$
(x_1,τ_1,σ_1)	1.69	-0.91	-4.57	-2.74	-2.74	0.59	0.09
(x_2,τ_2,σ_2)	-1.52	-0.66	-4.84	0.48	0.48	-0.66	0.34
(x_3,τ_3,σ_3)	4.53	-1.18	0.36	3.32	3.32	1.88	-0.18
(x_4,τ_4,σ_4)	-4.50	-1.30	4.13	12.35	12.35	-0.22	-0.30
(x_5,τ_5,σ_5)	-0.12	0.59	-5.99	3.72	3.72	-8.54	1.59
(x_6,τ_6,σ_6)	-3.65	-2.96	0.65	17.38	17.38	-0.27	-1.96
(x_7,τ_7,σ_7)	3.57	-2.99	0.36	10.16	10.16	0.28	-1.99

Table 2: Critical points of the augmented Lagrangian. The first four points correspond to the KKT points of the original problem, while the last three to the to the maxima of the Lagrangian function

From these results we can see that there is no duality gap between the primal solutions and their canonical dual solutions. By the fact that the point (x_1,τ_1,σ_1) satisfies both the conditions: $G(\tau,\sigma) \succeq 0$ and $(\mu + \tau) > 0$, it is the point corresponding to the global minimum of the primal problem, just as it is reported in Corollary 1. Moreover by updating $\mu_{k+1} = \mu_k + \tau_1 = 0.09$ for the next iteration, the value of the multiplier gets closer to the one corresponding to the global minimum, as reported in Table 1. Furthermore, by the conditions in Remark 2 adapted for this sub-problem, the point (x_4,τ_4,σ_4) is the biggest local maximum of the original problem.

This example shows that even if the problem with non-convex constraints becomes more complicated due to the additional penalty term, the canonical duality theory is still able to find the global solution. It is also important to note that for a problem with nonlinear constraints, the augmented Lagrangian methods usually produce a nonconvex sub-problem with double local minimizers. Traditional direct methods and algorithms for solving such highly nonconvex problems have great difficulties to find a good solution.

6 Conclusions

In this paper we have shown that the canonical duality theory presents a unified framework to cover traditional Lagrangian duality and KKT theory.

For general nonlinear constrained problems, the popular penalty methods and augmented Lagrangian theory may produce nonconvex sub-problems. Theorem 3 shows that as long as the nonconvex constraints satisfy the conditions in Assumption 1 and 2, the canonical duality theory can be used to solve the problem and the augmented Lagrangian method is indeed not necessary.

Finally we showed that even with the unnecessary nonconvex term produced by the penalty method, the canonical duality theory is still able to find the the best solution of the problem.

References

1. Birgin, E.G., Castillo, R.A., Martnez, J.M.: "Numerical Comparison of Augmented Lagrangian Algorithms for Nonconvex Problems", Computational Optimization and Applications **31**, 31-55 (2005).
2. Chen, Y. and Gao, D.Y. : Global solutions to large-scale spherical constrained quadratic minimization via canonical dual approach, to appear in *Optimization Methods and Software*.
3. Gao, D.Y.: Duality principles in nonconvex systems: Theory, methods and applications. Kluwer Academic Publishers, Dordrecht (2000)
4. Gao, D.Y.: Canonical dual transformation method and generalized triality theory in nonsmooth global optimization. J. Glob. Optim. **17**(1/4), 127–160 (2000)
5. Gao, D.Y.: Canonical duality theory: theory, method, and applications in global optimization. Comput. Chem. **33**, 1964–1972 (2009)
6. Gao, T.K.: Complete solutions to a class of eighth-order polynomial optimization problems, *IMA J. Appl. Math*, (2013). doi:10.1093/imamat/hxt033
7. Gao, D.Y., Ruan, N., Pardalos, P.M.: Canonical dual solutions to sum of fourth-order polynomials minimization problems with applications to sensor network localization. In: Pardalos, P.M., Ye, Y.Y., Boginski, V., Commander, C. (eds) *Sensors: Theory, Algorithms and Applications*, Springer (2010)
8. Gao, D.Y., Ruan, N., Sherali, H.D.: Solutions and optimality criteria for nonconvex constrained global optimization problems. J. Global Optim. 45, 473-497 (2009)
9. Gao, D. Y., Ruan, N., Sherali, H. D.: "Solutions and optimality criteria for nonconvex constrained global optimization problems with connections between canonical and Lagrangian duality." Journal of Global Optimization 45.3 473-497 (2009).
10. Gao, D.Y. and Sherali, H.D.: Canonical duality: Connection between nonconvex mechanics and global optimization, in *Advances in Appl. Mathematics and Global Optimization*, 249-316, Springer (2009).
11. Gao, D. Y. and Yang, Wei-Chi. *Minimal distance between two non-convex surfaces.* Optimization, Vol. 57, Issue 5, pp. 705-714 (2008).
12. Glowinski, R.: "Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems", Elsevier Science (2000).
13. Lagrange, Joseph-Louis: *Mecanique Analytique*. Courcier (1811), (reissued by Cambridge University Press, 2009; ISBN 978-1-108-00174-8).
14. Latorre, V., Gao, D.Y.: Canonical dual solution to nonconvex radial basis neural network optimization problem. submitted to Neurocomputing (2013)
15. Ruan, N., Gao, D.Y.: Global optimal solutions to a general sensor network localization problem. to appear in Perform. Eval. (2013) published online at http://arxiv.org/submit/654731
16. Nocedal, J., Wright, S. J.: "Numerical Optimization (2nd edition)". Springer (2006).
17. Powell, M.J.D.: "The Lagrange method and SAO with bounds on the dual variables", Optimization Methods and Software, (2013).
18. Wang, Z.B., Fang, S.C., Gao, D.Y., Xing, W.X.: Canonical dual approach to solving the maximum cut problem. J. Glob. Optim. **54**, 341–352 (2012)

19. Zhang, J., Gao, D.Y., Yearwood, J.: A novel canonical dual computational approach for prion AGAAAAGA amyloid fibril molecular modeling. J. Theor. Biol. **284**, 149–157 (2011)