ON THE STRONG CONCAVITY OF THE DUAL FUNCTION OF AN OPTIMIZATION PROBLEM

Vincent Guigues
School of Applied Mathematics, FGV
Rio de Janeiro, Brazil
(vincent.guigues@fgv.br)

Abstract. We provide three new proofs of the strong concavity of the dual function of some convex optimization problems. For problems with nonlinear constraints, we show that the assumption of strong convexity of the objective cannot be weakened to convexity and that the assumption that the gradients of all constraints at the optimal solution are linearly independent cannot be further weakened. Finally, we illustrate our results with several examples.

Keywords: Convex Analysis, Duality, Strong Convexity, Optimization.

AMS subject classifications: 26B25, 49N15.

1. Introduction

Consider the optimization problem

$$\left\{ \begin{array}{l} \inf f(x) \\ Ax \leq b \end{array} \right. \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $b \in \mathbb{R}^q$, and A is a $q \times n$ real matrix.

We make the following assumptions:

(H1) f is convex, differentiable, and ∇f is Lipschitz continuous on $\mathbb{R}^n$: there is $L(f) > 0$ such that for every $x, y \in \mathbb{R}^n$ we have

$$\|\nabla f(y) - \nabla f(x)\|_2 \leq L(f)\|y - x\|_2.$$

(H2) The rows of the matrix A are linearly independent.

Let θ be the dual function of (1.1) given by

$$\theta(\lambda) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda) := f(x) + \lambda^T (Ax - b), \tag{1.2}$$

for $\lambda \in \mathbb{R}^q$. The function θ is concave [5, 7] and it was shown in [4] that under Assumptions (H1), (H2), it is strongly concave on $\mathbb{R}^q$ with constant of strong concavity $\frac{\lambda_{\min}(AA^T)}{L(f)}$. This property was also shown in [8] with stronger assumptions, namely assuming f strongly convex and twice continuously differentiable.

For more general convex problems of the form

$$\inf_{x \in \mathbb{R}^n} \{f(x) : Ax \leq b, g_i(x) \leq 0, i = 1, \ldots, p\}, \tag{1.3}$$

where $f, g_i : \mathbb{R}^n \to \mathbb{R}$ are convex, $b \in \mathbb{R}^q$, and A is a $q \times n$ real matrix, the local strong concavity of the dual function obtained dualizing all constraints was shown in [4] in a neighborhood of an optimal dual solution. Such a property was also shown in [8] with stronger assumptions, namely assuming functions f and g twice continuously differentiable whereas in [4] it was only assumed that functions f, g_i have Lipschitz continuous gradients (in both proofs, strong convexity of f was also used).

The strong concavity of the dual function can be used to design efficient solution methods on the dual problem, for instance the Drift-Plus-Penalty Algorithm described in [8]. It was also used in [4] to compute inexact cuts for the recourse function of a two-stage convex stochastic program. These inexact cuts are useful to design Inexact Stochastic Mirror Descent (ISMD) Method, introduced in [4], which is an inexact variant of Stochastic Mirror Descent (SMD, see [6]).

In this paper, we provide three new proofs for the strong concavity of θ given by (1.2). The first one uses the assumptions from [4], the second one applies when f is coercive while the third ones applies when f is twice continuously differentiable. We also show that for problems of the form (1.3), the assumption that the gradients of all constraints at the optimal solution are linearly independent cannot be further weakened and that the assumption of strong convexity of the objective cannot be weakened to convexity. Finally, several examples are given.

2. Preliminaries

Given a convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$, as usual, $\operatorname{dom} f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$ is its *(effective) domain* and for any $x \in \mathbb{R}^n$ with $|f(x)| < +\infty$ the subdifferential $\partial f(x)$ is defined by

$$\partial f(x) := \{s \in \mathbb{R}^n : \langle s, y - x\rangle \leq f(y) - f(x),\ \forall y \in \mathbb{R}^n\},$$

and $\partial f(x) := \emptyset$ if $|f(x)| = +\infty$. The function f is said to be *proper* if $\operatorname{dom} f \neq \emptyset$ and f does not take the value $-\infty$. The Legendre-Fenchel conjugate f^* on $\mathbb{R}^n$ is defined by

$$f^*(y) := \sup_{x \in \mathbb{R}^n} \{\langle y, x\rangle - f(x)\} \quad \text{for all } y \in \mathbb{R}^n.$$

Similarly, given a concave function $g : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$ (i.e., $-g$ is convex), its *(effective) domain* is the set $\{x \in \mathbb{R}^n : g(x) > -\infty\}$. We refer to [5, 7] for the above concepts.

In what follows, $X \subset \mathbb{R}^n$ is a nonempty convex set, and its relative interior will be denoted by $\operatorname{ri} X$.

Definition 2.1 (Strongly convex functions)**.** *A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is strongly convex on X with constant of strong convexity $\alpha > 0$ with respect to a norm $\|\cdot\|$ if for any $x, y \in X \cap \operatorname{dom} f$ we have*

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{\alpha t(1-t)}{2}\|y - x\|^2, \tag{2.4}$$

for all $0 \leq t \leq 1$.

We can show that for lower semicontinuous convex functions, strong convexity on the relative interior of the domain implies strong convexity everywhere. More precisely, we have the following:

Lemma 2.2. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous convex function which is strongly convex on the relative interior of its domain with constant of strong convexity $\alpha > 0$ with respect to norm $\|\cdot\|$, i.e., f satisfies (2.4) for $x, y \in \operatorname{ri}(\operatorname{dom}(f))$ and $0 \leq t \leq 1$. Then f is strongly convex on $\mathbb{R}^n$ with the same constant of strong convexity $\alpha > 0$ with respect to norm $\|\cdot\|$*

Proof. Take any $x, y \in \operatorname{dom}(f)$ and $0 < t < 1$. We want to show (2.4). Since $\operatorname{ri}(\operatorname{dom}(f))$ is nonempty, we can take an arbitrary point $x_0 \in \operatorname{ri}(\operatorname{dom}(f))$. Observe that since f is lower semicontinuous and convex with $x, y \in \operatorname{dom} f$, using Proposition 1.2.5 in [5] we have

$$\begin{aligned} f(x) &= \varliminf_{u \to x} f(u) = \lim_{\theta \to 0, \theta > 0} f(x + \theta(x_0 - x)), \\ f(y) &= \varliminf_{u \to y} f(u) = \lim_{\theta \to 0, \theta > 0} f(y + \theta(x_0 - y)), \end{aligned} \tag{2.5}$$

and by lower semicontinuity of f at $tx + (1-t)y$ we also have

$$\begin{aligned} f(tx + (1-t)y) &\leq \lim_{\theta \to 0, \theta > 0} f(tx + (1-t)y + \theta(x_0 - tx - (1-t)y)) \\ &= \lim_{\theta \to 0, \theta > 0} f(t(x + \theta(x_0 - x)) + (1-t)(y + \theta(x_0 - y))) \\ &\leq \lim_{\theta \to 0, \theta > 0} tf(x + \theta(x_0 - x)) + (1-t)f(y + \theta(x_0 - y)) - \frac{\alpha t(1-t)}{2}\|(1-\theta)(x - y)\|^2 \\ &\overset{(2.5)}{=} tf(x) + (1-t)f(y) - \tfrac{\alpha t(1-t)}{2}\|y - x\|^2, \end{aligned} \tag{2.6}$$

where in the inequality above we have used the fact that $x + \theta(x_0 - x), y + \theta(x_0 - y) \in \operatorname{ri}(\operatorname{dom}(f))$ and that f is strongly convex on $\operatorname{ri}(\operatorname{dom}(f))$. ☐

It is well known (see for instance Proposition 6.1.2 in [5]) that if f is strongly convex with constant α with respect to norm $\|\cdot\|$ and subdifferentiable on X (i.e., the subdifferential $\partial f(x)$ of f at x is nonempty for every $x \in X$) then for all $x, y \in X$, we have

$$f(y) \geq f(x) + s^T(y-x) + \frac{\alpha}{2}\|y-x\|^2, \ \forall s \in \partial f(x).$$

Therefore, using the notation (here and in what follows) $\langle x, y\rangle = x^T y$ for $x, y \in \mathbb{R}^n$, since for a function f satisfying (H1) we must have

$$f(y) \leq f(x) + \langle \nabla f(x), y-x\rangle + \frac{L(f)}{2}\|y-x\|_2^2,$$

for all $x, y \in \mathbb{R}^n$, if f satisfies (H1) and is strongly convex on $\mathbb{R}^n$ with constant of strong convexity α with respect to norm $\|\cdot\|_2$ then we must have $\alpha \leq L(f)$.

We also recall that a convex function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is strongly convex on $\mathrm{ri}(\mathrm{dom}(f))$ with constant of strong convexity $\alpha > 0$ with respect to norm $\|\cdot\|$ if and only if for every $x, y \in \mathrm{ri}(\mathrm{dom}(f))$ we have

$$\langle \sigma - s, y-x\rangle \geq \alpha\|y-x\|^2, \ \text{ for all } \sigma \in \partial f(y), s \in \partial f(x). \tag{2.7}$$

For a proof of the above characterization (2.7), see the proof of Theorem 6.1.2 in [5].

Finally, if $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable then f is strongly convex on $\mathbb{R}^n$ with constant of strong convexity $\alpha > 0$ with respect to norm $\|\cdot\|_2$ if and only if for every $x \in \mathbb{R}^n$, we have $\nabla^2 f(x) \succeq \alpha I_n$(see for instance Proposition 1 in [2]).

Definition 2.3 (Strongly concave functions)**.** *A function $f : \mathbb{R}^n \to \mathbb{R}\cup\{-\infty\}$ is strongly concave on X with constant of strong concavity $\alpha > 0$ with respect to norm $\|\cdot\|$ if and only if $-f$ is strongly convex on X with constant of strong convexity $\alpha > 0$ with respect to norm $\|\cdot\|$.*

We recall two well known results of convex analysis that will be used in the sequel.

Proposition 2.4. *Let $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be a convex and lower semicontinuous function. Then f^* is strongly convex on $\mathbb{R}^n$ with constant of strong convexity $\alpha > 0$ for norm $\|\cdot\|_2$ if and only if f is differentiable and ∇f is Lipschitz continuous on $\mathbb{R}^n$ with constant $1/\alpha$ for norm $\|\cdot\|_2$.*

For a proof of the previous proposition, see the proof of Proposition 12.60 in [7]. The following result is known as Baillon-Haddad Theorem that we specialize to functions $f : \mathbb{R}^n \to \mathbb{R}$:

Theorem 2.5. [1] *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex, differentiable and satisfying Assumption (H1) for some $0 < L(f) < \infty$. Then ∇f is $1/L(f)$-co-coecive, meaning that for all $x, y \in \mathbb{R}^n$ we have*

$$\langle y-x, \nabla f(y) - \nabla f(x)\rangle \geq \frac{1}{L(f)}\|\nabla f(y) - \nabla f(x)\|_2^2. \tag{2.8}$$

Remark 2.6. *Theorem 2.5 follows from Proposition 2.4 and property* (2.7)*.*

3. Problems with linear constraints

3.1. **Proofs of the strong concavity of the dual function.** In this section we provide several proofs of the following proposition, first proved in [4].

Proposition 3.1. *Let Assumptions (H1) and (H2) hold. Then the dual function θ given by* (1.2) *is strongly concave on $\mathbb{R}^q$ with constant of strong concavity $\frac{\lambda_{\min}(AA^T)}{L(f)}$ with respect to norm $\|\cdot\|_2$.*

We first recall the proof of Proposition 3.1 given in [4].

Proof of Proposition 3.1 from [4]**.** The dual function of (1.1) given by (1.2) can be written

$$\begin{array}{rcl}\theta(\lambda) & = & \inf_{x\in\mathbb{R}^n}\{f(x) + \lambda^T(Ax-b)\} = -\lambda^T b - \sup_{x\in\mathbb{R}^n}\{-x^T A^T\lambda - f(x)\} \\ & = & -\lambda^T b - f^*(-A^T\lambda) \text{ by definition of } f^*.\end{array} \tag{3.9}$$

From Assumption (H1) and Proposition 2.4, $-f^*$ is strongly concave with constant of strong concavity $1/L(f)$. Assumption (H2) implies $\mathrm{Ker}(AA^T) = \{0\}$ which, together with the strong concavity of $-f^*$, easily implies that $\lambda \to -f^*(-A^T\lambda)$ is strongly concave with constant of strong concavity $\frac{\lambda_{\min}(AA^T)}{L(f)}$ (see Proposition 2.5 in [4] for details) and therefore so is θ which is the sum of a linear function and of $-f^*(-A^T\lambda)$. □

Our second proof is based on Theorem 2.5.

Second proof of Proposition 3.1. If $\text{dom}(-\theta)$ is empty there is nothing to show. Let us now assume that $\text{dom}(-\theta)$ is nonempty. Since $-\theta$ is convex, the subdifferential $\partial(-\theta)(\lambda)$ is nonempty for every λ in the relative interior of $\text{dom}(-\theta)$. We first show that for $x, y \in \text{ri}(\text{dom}(-\theta))$ relation (2.7) holds for $f = -\theta$, $\alpha = \frac{\lambda_{\min}(AA^T)}{L(f)}$ and $\|\cdot\| = \|\cdot\|_2$. Let us take $\lambda_1, \lambda_2 \in \text{ri}(\text{dom}(-\theta))$. Then $\partial(-\theta)(\lambda_1)$ and $\partial(-\theta)(\lambda_2)$ are nonempty. Using for instance Lemma 2.1 in [3] or Corollary 4.5.3 p.273 in [5], we have

$$-\partial(-\theta)(\lambda) = \{Ax - b : x \in S(\lambda)\} \tag{3.10}$$

where $S(\lambda)$ is the set of optimal solutions of (1.2). The latter equality yields that $S(\lambda_1)$ and $S(\lambda_2)$ are nonempty. Next, for any $x_1 \in S(\lambda_1)$ and $x_2 \in S(\lambda_2)$, by first order optimality conditions, we have

$$\nabla f(x_i) + A^T\lambda_i = 0,\ i = 1, 2. \tag{3.11}$$

Now for any $s_2 \in \partial(-\theta)(\lambda_2)$ and $s_1 \in \partial(-\theta)(\lambda_1)$ we can find $x_1 \in S(\lambda_1)$ and $x_2 \in S(\lambda_2)$ such that $s_i = -(Ax_i - b)$, $i = 1, 2$, which implies

$$\begin{array}{rcl}\langle s_2 - s_1, \lambda_2 - \lambda_1\rangle & = & -\langle A(x_2 - x_1), \lambda_2 - \lambda_1\rangle \text{ using (3.10)}\\ & = & -\langle x_2 - x_1, A^T(\lambda_2 - \lambda_1)\rangle\\ & = & \langle x_2 - x_1, \nabla f(x_2) - \nabla f(x_1)\rangle \text{ using (3.11)}\\ & \geq & (1/L(f))\|\nabla f(x_2) - \nabla f(x_1)\|_2^2 \text{ using (2.8)}\\ & = & (1/L(f))\|A^T(\lambda_2 - \lambda_1)\|_2^2 \text{ using (3.11)}\\ & \geq & \dfrac{\lambda_{\min}(AA^T)}{L(f)}\|\lambda_2 - \lambda_1\|_2^2.\end{array} \tag{3.12}$$

Recalling characterization (2.7) of strong convexity, we have shown that $-\theta$ is strongly convex with constant of strong convexity $\frac{\lambda_{\min}(AA^T)}{L(f)}$ with respect to norm $\|\cdot\|_2$ on $\text{ri}(\text{dom}(-\theta))$. Since $-\theta$ is convex and lower semicontinuous on $\mathbb{R}^q$, we can apply Lemma 2.2 with $X = \mathbb{R}^q$ to obtain the strong convexity of $-\theta$ (or equivalently the strong concavity of θ) on $\mathbb{R}^q$ with constant of strong convexity $\frac{\lambda_{\min}(AA^T)}{L(f)}$ with respect to norm $\|\cdot\|_2$. □

Our next proof of the strong concavity of the dual function applies when the objective function f is coercive. It is based on properties of the value function.

Proposition 3.2. *Let Assumptions (H1) and (H2) hold and assume that f is coercive in the sense that $f(x) \to +\infty$ as $\|x\| \to +\infty$. Then the dual function θ given by* (1.2) *is strongly concave on $\mathbb{R}^q_+$ with constant of strong concavity $\frac{\lambda_{\min}(AA^T)}{L(f)}$ with respect to norm $\|\cdot\|_2$.*

Proof. Let v be the value function given by

$$v(c) = \left\{\begin{array}{l}\inf f(x)\\ Ax - b + c \leq 0,\end{array}\right. \tag{3.13}$$

for $c \in \mathbb{R}^q$.

We first show that (i) v is differentiable and (ii) ∇v is Lipschitz continuous with Lipschitz constant $\frac{L(f)}{\lambda_{\min}(AA^T)}$.

Let us show (i). By Assumption (H2), A is surjective meaning that for any $y \leq b - c$ we can find x satisfying $Ax = y \leq b - c$ which is feasible for (3.13) (due to (H2), problem (3.13) is in fact strictly feasible for every c). Therefore, for any c, problem (3.13) is convex, with polyhedral nonempty feasible set, and continuous coercive objective function, implying that it has a finite optimal value $v(c)$ and optimal solutions. Also since v is convex and finite for any c, it is a continuous convex function, hence its subdifferential is nonempty at any c and is given by the set of optimal dual solutions of the dual problem

$$\sup_{\lambda \geq 0} \theta_c(\lambda), \tag{3.14}$$

where

$$\theta_c(\lambda) = \inf_{x \in \mathbb{R}^n} f(x) + \lambda^T (Ax - b + c).$$

Take $c \in \mathbb{R}^q$. To show that v is differentiable at c, if suffices to show that $\partial v(c)$ is a singleton. Take $\lambda_1, \lambda_2 \in \partial v(c)$, which, as we recall, are optimal solutions to dual problem (3.14). Let $x(c)$ be an optimal solution of (3.13) (recall that (3.13) has optimal primal and dual solutions). By the optimality conditions, we have $\nabla f(x(c)) + A^T \lambda_1 = 0$ and $\nabla f(x(c)) + A^T \lambda_2 = 0$. Therefore $\lambda_1 - \lambda_2 \in \mathrm{Ker}(A^T)$ and by Assumption (H2) we have $\mathrm{Ker}(A^T) = \{0\}$ which implies $\lambda_1 = \lambda_2$, i.e., $\partial v(c)$ is a singleton and (i) is shown. Therefore, for each $c \in \mathbb{R}^q$ there is a unique multiplier $\lambda(c) = \nabla v(c)$.

Let us now show (ii). We will still denote by $x(c)$ an optimal solution of (3.13). Take $c_1, c_2 \in \mathbb{R}^q$. By the optimality conditions, we get

$$\nabla f(x(c_i)) + A^T \lambda(c_i) = 0, i = 1, 2, \tag{3.15}$$

and by complementary slackness

$$\langle \lambda(c_i), Ax(c_i) - b + c_i \rangle = 0, i = 1, 2. \tag{3.16}$$

Therefore

$$\begin{array}{rcl} \langle A(x(c_2) - x(c_1)), \lambda(c_1) - \lambda(c_2) \rangle & = & \langle x(c_2) - x(c_1), A^T(\lambda(c_1) - \lambda(c_2)) \rangle \\ & \overset{(3.15)}{=} & \langle x(c_2) - x(c_1), \nabla f(x(c_2)) - \nabla f(x(c_1)) \rangle \\ & \overset{(2.8)}{\geq} & (1/L(f)) \|\nabla f(x(c_2)) - \nabla f(x(c_1))\|_2^2 \\ & \overset{(3.15)}{=} & (1/L(f)) \|A^T(\lambda(c_2) - \lambda(c_1))\|_2^2 \\ & \geq & \dfrac{\lambda_{\min}(AA^T)}{L(f)} \|\lambda(c_2) - \lambda(c_1)\|_2^2. \end{array} \tag{3.17}$$

Observe that in the first inequality above, (2.8) can be used because (H1) is satisfied. Next since $\lambda(c_1), \lambda(c_2) \geq 0$, we have

$$\langle \lambda(c_1), Ax(c_2) - b + c_2 \rangle \leq 0, \ \langle \lambda(c_2), Ax(c_1) - b + c_1 \rangle \leq 0. \tag{3.18}$$

It follows that

$$\begin{array}{l} \langle A(x(c_2) - x(c_1)), \lambda(c_1) - \lambda(c_2) \rangle = \langle Ax(c_2) - b - (Ax(c_1) - b), \lambda(c_1) - \lambda(c_2) \rangle \\ \overset{(3.16)}{=} \langle \lambda(c_1), c_1 \rangle + \langle \lambda(c_2), c_2 \rangle + \langle \lambda(c_1), Ax(c_2) - b \rangle + \langle \lambda(c_2), Ax(c_1) - b \rangle \\ \overset{(3.18)}{\leq} \langle \lambda(c_2) - \lambda(c_1), c_2 - c_1 \rangle \leq \|\lambda(c_2) - \lambda(c_1)\|_2 \|c_2 - c_1\|_2, \end{array} \tag{3.19}$$

where the last inequality is due to the Cauchy-Schwartz inequality.

Combining (3.17) and (3.19) we get

$$\|\nabla v(c_2) - \nabla v(c_1)\|_2 = \|\lambda(c_2) - \lambda(c_1)\|_2 \leq \frac{L(f)}{\lambda_{\min}(AA^T)} \|c_2 - c_1\|_2.$$

Therefore, we have shown that v is differentiable and has Lipschitz continuous gradient with Lipschitz constant $\frac{L(f)}{\lambda_{\min}(AA^T)}$. Recalling that v is convex, using Proposition 2.4 we deduce that v^* is strongly convex on $\mathbb{R}^q$ with constant of strong convexity $\frac{\lambda_{\min}(AA^T)}{L(f)}$. Since $v^* = -\theta$ on $\mathbb{R}^q_+$, this shows the strong concavity of θ on $\mathbb{R}^q_+$ with the constant of strong concavity $\frac{\lambda_{\min}(AA^T)}{L(f)}$. □

Proofs of the strong concavity of θ for strongly convex f. The proof of the strong concavity of θ when, additionally to (H1) and (H2), the function f is strongly convex is known. It can be seen as a special case of Theorem 10 in [8]. For completeness, we provide below a simple proof of this result and provide a new proof when f is twice continuously differentiable.

Proposition 3.3. *Let Assumptions (H1) and (H2) hold. Assume that f is strongly convex on $\mathbb{R}^n$ with constant of strong convexity α with respect to $\|\cdot\|_2$. Then the dual function θ given by* (1.2) *is strongly concave on $\mathbb{R}^q$ with constant of strong concavity $\frac{\alpha\lambda_{\min}(AA^T)}{L(f)^2}$ with respect to norm $\|\cdot\|_2$.*

Proof. For any $\lambda \in \mathbb{R}^q$, due to the strong convexity of f, optimization problem (1.2) has a unique optimal solution denoted by $x(\lambda)$. By (3.10), θ is differentiable with $\nabla\theta(\lambda) = Ax(\lambda) - b$. It follows that

$$\begin{array}{rcl} -\langle \nabla\theta(\lambda_2) - \nabla\theta(\lambda_1), \lambda_2 - \lambda_1\rangle & = & -\langle A(x(\lambda_2) - x(\lambda_1)), \lambda_2 - \lambda_1\rangle, \\ & = & -\langle x(\lambda_2) - x(\lambda_1), A^T(\lambda_2 - \lambda_1)\rangle \\ & \stackrel{(3.11)}{=} & \langle x(\lambda_2) - x(\lambda_1), \nabla f(x(\lambda_2)) - \nabla f(x(\lambda_1))\rangle \\ & \geq & \alpha\|x(\lambda_2) - x(\lambda_1)\|_2^2 \\ & \stackrel{(H1)}{\geq} & (\alpha/L(f)^2)\|\nabla f(x(\lambda_2)) - \nabla f(x(\lambda_1))\|_2^2 \\ & = & (\alpha/L(f)^2)\|A^T(\lambda_2 - \lambda_1)\|_2^2 \\ & \geq & \dfrac{\alpha\lambda_{\min}(AA^T)}{L(f)^2}\|\lambda_2 - \lambda_1\|_2^2. \end{array} \tag{3.20}$$

In (3.20), the first equality comes from $\nabla\theta(\lambda) = Ax(\lambda) - b$, the first inequality comes from the strong convexity of f, while the last equality comes from the optimality conditions. This achieves the proof. □

Since we must have $\alpha \leq L(f)$, we get a smaller constant of strong concavity than in the previous case where f was not necessarily strongly convex. We now provide a new proof when f is strongly convex and twice continuously differentiable on $\mathbb{R}^n$.

Proposition 3.4. *Let Assumptions (H1) and (H2) hold. Assume that f is strongly convex on $\mathbb{R}^n$ with constant of strong convexity α with respect to $\|\cdot\|_2$ and twice continuously differentiable on $\mathbb{R}^n$. Then the dual function θ given by* (1.2) *is strongly concave on $\mathbb{R}^q$ with constant of strong concavity not larger than $\frac{1}{\alpha}\lambda_{\min}(AA^T)$ with respect to norm $\|\cdot\|_2$.*

Proof. By the Implicit Function Theorem, θ is twice continuously differentiable with

$$\nabla^2\theta(\lambda) = -H_{x\lambda}^T H_{xx}^{-1} H_{x\lambda}$$

where

$$H_{x\lambda} = \nabla^2_{x\lambda}\mathcal{L}(x(\lambda), \lambda) = A^T,\ H_{xx} = \nabla^2_{xx}\mathcal{L}(x(\lambda), \lambda) = \nabla^2 f(x(\lambda)).$$

Hence,

$$\nabla^2\theta(\lambda) = -A[\nabla^2 f(x(\lambda))]^{-1}A^T.$$

The function f being strongly convex with constant of strong convexity α we have that $\nabla^2 f(x) \succeq \alpha I_n$ for all x and therefore $\frac{1}{\alpha}I_n \succeq [\nabla^2 f(x)]^{-1}$. Using Assumption (H2), matrix $A[\nabla^2 f(x(\lambda))]^{-1}A^T$ is invertible for all λ and satisfies $\frac{1}{\alpha}AA^T \succeq A[\nabla^2 f(x(\lambda))]^{-1}A^T$ implying $\lambda_{\min}(A[\nabla^2 f(x(\lambda))]^{-1}A^T) \leq \frac{1}{\alpha}\lambda_{\min}(AA^T)$ which implies that θ is strongly concave with constant of strong concavity not larger than $\frac{1}{\alpha}\lambda_{\min}(AA^T)$ with respect to norm $\|\cdot\|_2$. □

3.2. **Applications.** We illustrate Proposition 3.1 with 3 examples. The first example is a *degenerate* one and corresponds to linear programs which indeed satisfy (H1) and can satisfy (H2). However, as discussed in Example 3.5 below, for such problems the domain $\mathrm{dom}(\theta) = \{\lambda : \theta(\lambda) > -\infty\}$ of the dual function θ is either a singleton or the empty set and such functions are indeed, by definition, strongly concave even if this property will not, in this case, be enlightening in practice.

Example 3.5 (Linear programs)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by*

$$f(x) = c^T x + c_0 \tag{3.21}$$

where $c \in \mathbb{R}^n$, $c_0 \in \mathbb{R}$. Clearly f is convex differentiable with Lipschitz continuous gradients; any $L(f) > 0$ being a valid Lipschitz constant. Proposition 3.1 tells us that if the rows of A are linearly independent then the dual function θ of (1.1) *given by* (1.2) *is strongly concave on $\mathbb{R}^q$. In this case, the strong concavity can be checked directly by computing θ. Indeed, we have*

$$f^*(x) = \begin{cases} -c_0 & \textit{if } x = c, \\ +\infty & \textit{if } x \neq c, \end{cases}$$

and plugging this expression of f^* into (3.9), *we get*[1]

$$\theta(\lambda) = \begin{cases} -\lambda^T b + c_0 & \text{if } A^T\lambda = -c, \\ -\infty & \text{if } A^T\lambda \neq -c. \end{cases}$$

Therefore if $c \in Im(A^T)$ then there is $\lambda \in \mathbb{R}^q$ such that

$$A^T\lambda = -c, \tag{3.22}$$

and if the rows of A are linearly independent then there is only one λ, let us call it λ_0, satisfying (3.22). *In this situation, the domain of θ is a singleton: $dom(\theta) = \{\lambda_0\}$, and θ indeed is strongly concave (see Definition 2.1). If $c \notin Im(A^T)$ then $dom(\theta) = \emptyset$ and θ is again strongly concave.*

The example which follows gives a class of problems where the dual function is strongly concave on $\mathbb{R}^q$ with f not necessarily strongly convex.

Example 3.6 (Quadratic convex programs)**.** *Consider a problem of form* (1.1) *where $f(x) = \frac{1}{2}x^T Q_0 x + a_0^T x + b_0$, Q_0 is an $n \times n$ nonnull semidefinite positive matrix, A is a $q \times n$ real matrix, $a_0 \in Im(Q_0)$, and $b_0 \in \mathbb{R}$. Clearly, f is convex, differentiable, and ∇f is Lipschitz continuous with Lipschitz constant $L(f) = \|Q_0\|_2 = \lambda_{\max}(Q_0) > 0$ with respect to $\|\cdot\|_2$ on $\mathbb{R}^n$. If the rows of A are linearly independent, using Proposition 3.1 we obtain that the dual function of* (1.1) *is strongly concave with constant of strong concavity $\frac{\lambda_{\min}(AA^T)}{\lambda_{\max}(Q_0)} > 0$ with respect to norm $\|\cdot\|_2$ on $\mathbb{R}^q$. Observe that strong concavity holds in particular if Q_0 is not definite positive, in which case f is not strongly convex. For this example, strong concavity of θ is driven by the greatest eigenvalue of Q_0 and by the lowest eigenvalue of AA^T.*

Since f is convex, differentiable, its gradient being Lipschitz continuous with Lipschitz constant $\lambda_{\max}(Q_0)$, from Proposition 2.4, we know that f^* is strongly convex with constant of strong convexity $1/\lambda_{\max}(Q_0)$. This can be checked by direct computation. Indeed, let $\lambda_{\max}(Q_0) = \lambda_1(Q_0) \geq \lambda_2(Q_0) \geq \ldots \geq \lambda_r(Q_0) > \lambda_{r+1}(Q_0) = \lambda_{r+2}(Q_0) = \ldots = \lambda_n(Q_0) = 0$ be the ordered eigenvalues of Q_0 where r is the rank of Q_0. Let P be a corresponding orthogonal matrix of eigenvectors for Q_0, i.e., $\text{Diag}(\lambda_1(Q_0), \ldots, \lambda_n(Q_0)) = P^T Q_0 P$ *with $PP^T = P^T P = I_n$. Defining*

$$Q_0^+ = P\text{Diag}\Big(\frac{1}{\lambda_1(Q_0)}, \ldots, \frac{1}{\lambda_r(Q_0)}, \underbrace{0, \ldots, 0}_{n\text{-}r \text{ times}}\Big)P^T,$$

it is straightforward to check that

$$f^*(x) = \begin{cases} -b_0 + \frac{1}{2}(x - a_0)^T Q_0^+ (x - a_0) & \text{if } x \in Im(Q_0), \\ +\infty & \text{otherwise}, \end{cases} \tag{3.23}$$

and plugging expression (3.23) *of f^* into* (3.9), *we get*

$$\theta(\lambda) = \begin{cases} b_0 - \lambda^T b - \frac{1}{2}(a_0 + A^T\lambda)^T Q_0^+ (a_0 + A^T\lambda) & \text{if } A^T\lambda \in Im(Q_0), \\ -\infty & \text{otherwise}. \end{cases}$$

If $x' = (x_1', \ldots, x_n')$ is the vector of the coordinates of x in the basis $(v_1, v_2, \ldots, v_n)$ where v_i is ith column of $P = [v_1, v_2, \ldots, v_n]$ (i.e., $(v_1, \ldots, v_r)$ is a basis of $Im(Q_0)$ and $(v_{r+1}, \ldots, v_n)$ is a basis of $Ker(Q_0)$) and writing $a_0 = \sum_{i=1}^r a_{0i}' v_i$, we obtain

$$f^*(x) = \begin{cases} g(P^T x) \text{ where } g : \mathbb{R}^n \rightarrow \mathbb{R} \text{ is given by } g(x') = -b_0 + \sum_{i=1}^r \frac{(x_i' - a_{0i}')^2}{2\lambda_i(Q_0)} & \text{if } x \in Im(Q_0), \\ +\infty \text{ otherwise}. \end{cases}$$

Observe that for $x', y' \in \mathbb{R}^r \times \{\underbrace{(0, \ldots, 0)}_{n\text{-}r \text{ times}}\}$ we have

$$g(y') \geq g(x') + \nabla g(x')^T (y' - x') + \frac{1}{2\lambda_1(Q_0)}\|y' - x'\|_2^2$$

[1] In this simple case, the dual function is well known and can also be obtained without using the conjugate of f

and g is strongly convex with constant of strong convexity $\frac{1}{\lambda_1(Q_0)}$ *with respect to norm* $\|\cdot\|_2$ *on* $\mathbb{R}^r \times \{\underbrace{(0,\ldots,0)}_{n\text{-}r\ times}\}$. *Recalling that* $f^*(x) = g(P^T x)$ *for* $x \in dom(f^*) = Im(Q_0)$ *and that* $P^T x \in \mathbb{R}^r \times \{\underbrace{(0,\ldots,0)}_{n\text{-}r\ times}\}$ *for* $x \in Im(Q_0)$, *we deduce that* f^* *is strongly convex with constant of strong convexity*

$$\frac{\lambda_{\min}(PP^T)}{\lambda_1(Q_0)} = \frac{\lambda_{\min}(I_n)}{\lambda_{\max}(Q_0)} = \frac{1}{\lambda_{\max}(Q_0)}$$

with respect to norm $\|\cdot\|_2$.

Example 3.7. *Let* $f(x) = \sum_{k=1}^M \alpha_k f_k(x)$ *for* $\alpha_k \in \mathbb{R}$ *and* $f_k : \mathbb{R}^n \to \mathbb{R}$ *convex differentiable with Lipschitz constant* $L_k > 0$ *with respect to norm* $\|\cdot\|_2$ *on* $\mathbb{R}^n$ *for* $k = 1,\ldots,M$. *Let* A *be a* $q \times n$ *matrix with independent rows. Then the dual function* (1.2) *of* (1.1) *is strongly concave on* $\mathbb{R}^q$ *with constant of strong concavity* $\lambda_{\min}(AA^T)/\sum_{k=1}^M \alpha_k L_k$ *with respect to* $\|\cdot\|_2$.

4. Problems with linear and nonlinear constraints

We now consider problems of form (1.3) with corresponding dual function

$$\theta(\lambda,\mu) = \left\{ \begin{array}{l} \inf\ f(x) + \lambda^T(Ax-b) + \mu^T g(x) \\ x \in \mathbb{R}^n \end{array} \right. \tag{4.24}$$

where $g(x) = (g_1(x),\ldots,g_p(x))$. For this class of problems, the local strong concavity of dual function (4.24) is given by the following theorem, which was shown in [4].

Theorem 4.1. *Consider the optimization problem*

$$\inf_{x\in\mathbb{R}^n} \{f(x) : Ax \le b, g_i(x) \le 0, i = 1,\ldots,p\}, \tag{4.25}$$

where A *is a* $q \times n$ *real matrix. We assume that*

- (A1) $f : \mathbb{R}^n \to \mathbb{R}$ *is strongly convex and has Lipschitz continuous gradient on* $\mathbb{R}^n$*;*
- (A2) $g_i : \mathbb{R}^n \to \mathbb{R}, i = 1,\ldots,p$, *are convex and have Lipschitz continuous gradients;*
- (A3) *if* x_* *is the optimal solution of* (4.25) *then the rows of matrix* $\begin{pmatrix} A \\ J_g(x_*) \end{pmatrix}$ *are linearly independent where* $J_g(x)$ *denotes the Jacobian matrix of* $g(x) = (g_1(x),\ldots,g_p(x))$ *at* x*;*
- (A4) *there is* $x_0 \in ri(\{g \le 0\})$ *such that* $Ax_0 \le b$.

Let θ *be the dual function of problem* (4.25)*:*

$$\theta(\lambda,\mu) = \left\{ \begin{array}{l} \inf\ f(x) + \lambda^T(Ax-b) + \mu^T g(x) \\ x \in \mathbb{R}^n. \end{array} \right. \tag{4.26}$$

Let $(\lambda_*,\mu_*) \ge 0$ *be an optimal solution of the dual problem*[2]

$$\sup_{\lambda\ge 0,\mu\ge 0} \theta(\lambda,\mu).$$

Then there is some neighborhood $\mathcal{N}$ *of* (λ_*,μ_*) *such that* θ *is strongly concave on* $\mathcal{N} \cap \mathbb{R}_+^{p+q}$.

Comparing Theorem 4.1 where strong convexity of the objective is required with Proposition 3.1 which applies to problems with convex (and possibly non strongly convex) objectives, we can wonder if strong convexity can be relaxed to convexity in Theorem 4.1. The answer is negative, as shown by the following example.

[2]Observe that the primal problem has a finite optimal value and the assumptions of the Convex Duality Theorem are satisfied (Slater Assumption (A4) is satisfied and the primal objective is bounded from below on the feasible set), implying that the dual problem is feasible, has an optimal solution, and the same optimal value as the primal problem.

Example 4.2. *Consider the optimization problem*

$$(P_1)\ \min_{x\in\mathbb{R}^n}\{c : x_i^2 \le 1, i=1,\dots,n\}$$

which is of form (1.3) with $f(x)=c$ constant, without linear constraints, and with constraint functions $g_i(x)=x_i^2-1$, $i=1,\dots,n$, which satisfy Assumption (A2). Any feasible x_* with all components nonnull is an optimal solution of (P_1) satisfying Assumption (A3) since the rows of $J_g(x_*)=2\mathrm{Diag}(x_*)$ are linearly independent. Clearly (A4) is also satisfied. However, (A1) is not satisfied. For this example, for $\mu\ge 0$ dual function θ is given by

$$\theta(\mu)=c+\min_{x\in\mathbb{R}^n}\sum_{i=1}^n \mu_i(x_i^2-1)=c-\sum_{i=1}^n \mu_i$$

and is therefore not strongly concave. This shows that the conclusion of Theorem 4.1 may fail if we replace strong convexity by convexity in Assumption (A1). Observe also that nonlinear constraints of (P_1) can be written as $Ax\le b$ where b is a vector of ones of size $2n$ and where the rows of $A=[I_n;-I_n]$ are not linearly independent.

It is also natural to wonder if in Proposition 3.1 and Theorem 4.1, Assumptions (H2) and (A3) can be relaxed assuming that the gradients of the active constraints at an optimal solution are linearly independent, instead of assuming that the gradients of all constraints at an optimal solution are linearly independent.

For problems with linear constraints of form (1.1), from representation (3.9), if θ is strongly concave on $\mathbb{R}^q$ then $\lambda\to f^*(-A^T\lambda)$ is strongly convex. If $0\in\mathrm{dom}(f^*)$ this implies that $\mathrm{Ker}(A^T)=\{0\}$ and therefore that Assumption (H2) must hold otherwise $f^*(-A^T\lambda)$ would be constant equal to $f^*(0)$ on the vector space $\mathrm{Ker}(A^T)$ of positive dimension which is not possible for a strongly convex function with $0\in\mathrm{dom}(f^*)$. Similarly, the following example shows that in Theorem 4.1, Assumption (A3) cannot be relaxed assuming that the gradients of the active constraints at the optimal solution are linearly independent.

Example 4.3. *Consider the optimization problem*

$$(P_2)\left\{\begin{array}{l}\displaystyle\min_{x\in\mathbb{R}^n}\ \frac{1}{2}\sum_{i=1}^n x_i^2\\ \displaystyle -\sum_{i=1}^n x_i\le -1,\\ \displaystyle \sum_{i=1}^n x_i^2-1\le 0,\end{array}\right.$$

of form (1.3) satisfying (A4), with $A=-e^T$ where e is a vector of ones of dimension n, $f(x)=\frac{1}{2}\sum_{i=1}^n x_i^2$ satisfying (A1), and $p=1$, $g_1(x)=\sum_{i=1}^n x_i^2-1$ satisfying (A2). The optimal solution of this problem is $x_*=\frac{1}{n}e$ with corresponding optimal value $\frac{1}{2n}$ and only the constraint $-\sum_{i=1}^n x_i\le -1$ is active at x_*. For this problem, for $\lambda,\mu\ge 0$, dual function θ is given by

$$\begin{array}{rcl}\theta(\lambda,\mu)&=&\displaystyle\min_{x\in\mathbb{R}^n}\frac{1}{2}\sum_{i=1}^n x_i^2+\lambda(1-\sum_{i=1}^n x_i)+\mu(\sum_{i=1}^n x_i^2-1)\\ &=&\displaystyle\lambda-\mu-\frac{n}{2}\frac{\lambda^2}{1+2\mu}.\end{array}$$

The Hessian matrix of θ at $(\lambda,\mu)\ge 0$ is given by

$$\nabla^2\theta(\lambda,\mu)=\left(\begin{array}{cc}\dfrac{-n}{1+2\mu} & \dfrac{2n\lambda}{(1+2\mu)^2}\\ \dfrac{2n\lambda}{(1+2\mu)^2} & \dfrac{-4n\lambda^2}{(1+2\mu)^3}\end{array}\right).$$

Observe that 0 is an eigenvalue of $\nabla^2\theta(\lambda,\mu)$ with $(\frac{2\lambda}{1+2\mu},1)$ a corresponding eigenvector, the other eigenvalue being $-\frac{n}{1+2\mu}-\frac{4n\lambda^2}{(1+2\mu)^3}$ which is negative for $\lambda,\mu\ge 0$. Therefore for all $\lambda,\mu\ge 0$ we have that $\nabla^2\theta(\lambda,\mu)$ is semidefinite negative but not definite negative implying that θ is not strongly concave on any set of positive measure contained in $\mathbb{R}_+^2$ and in particular there is no neighborhood $\mathcal{N}_*$ of the optimal dual solution λ_*,μ_*

such that θ is strongly concave on $\mathcal{N}_* \cap \mathbb{R}^2_+$. Finally, observe that strong duality holds and $\lambda_* = \frac{1}{n}$, $\mu_* = 0$ since the dual problem is

$$\begin{aligned}
\max_{\lambda,\mu \geq 0} \theta(\lambda,\mu) &= \max_{\lambda\geq 0}\max_{\mu \geq 0} \lambda - \mu - \frac{n}{2}\frac{\lambda^2}{1+2\mu} \\
&= \max\left(\max_{\lambda \geq \frac{1}{\sqrt{n}}} \frac{1}{2} + \lambda(1-\sqrt{n}), \max_{0\leq\lambda\leq\frac{1}{\sqrt{n}}} \lambda - \frac{n}{2}\lambda^2\right) \\
&= \max\left(-\frac{1}{2} + \frac{1}{\sqrt{n}}, \frac{1}{2n}\right) = \frac{1}{2n},
\end{aligned}$$

whose optimal value is indeed the optimal value $\frac{1}{2n}$ of the primal problem attained at $\lambda_* = \frac{1}{n}$, $\mu_* = 0$. Therefore for this problem, the gradient of the active constraint at x_* is $-e$ and is consequently linearly independent whereas the gradients of the constraints at x_* are $-e$ and $\frac{2}{n}e$ and are therefore not linearly independent. This shows that the conclusion of Theorem 4.1 does not hold if instead of assuming that the gradients of all constraint functions at the optimal solution x_* are linearly independent we assume that the gradients of the active constraint functions at the optimal solution are linearly independent.

5. Conclusion

In this paper we analyzed the strong concavity of the dual function of an optimization problem. A possible extension would be to show this property for some classes of problems when the dual function is obtained dualizing only some of the constraints.

Acknowledgments

Research of the author was supported by CNPq grant 304887/2019-6. The author would like to thank the referee and the Managing Editor in charge of the manuscript for helpful suggestions that improved the presentation.

References

[1] J.-B. Baillon and G. Haddad. Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones. *Israel J. Math.*, 26:137–150, 1977.

[2] H.H. Bauschke, Jerome Bolte, and Marc Teboulle. A descent lemma beyond lipschitz gradient continuity: first-order methods revisited and applications. *Mathematics of Operations Research*, 42:330–348, 2016.

[3] V. Guigues. Convergence analysis of sampling-based decomposition methods for risk-averse multistage stochastic convex programs. *SIAM Journal on Optimization*, 26:2468–2494, 2016.

[4] V. Guigues. Inexact Stochastic Mirror Descent for two-stage nonlinear stochastic programs. *Mathematical Programming*, to appear.

[5] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I.* Springer-Verlag, Berlin, 1993.

[6] G. Lan, A. Nemirovski, and A. Shapiro. Validation analysis of mirror descent stochastic approximation method. *Mathematical Programming*, 134:425–458, 2012.

[7] R. T. Rockafellar and R. J-B Wets. *Variational Analysis.* Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, 1997.

[8] H. Yu and J. Neely. On the Convergence Time of the Drift-Plus-Penalty Algorithm for Strongly Convex Programs. *54th IEEE Conference on Decision and Control (CDC)*, pages 2673–2679, 2015.