

# A perturbation view of level-set methods for convex optimization

Ron Estrin<sup>a</sup> and Michael P. Friedlander<sup>b</sup>

<sup>a</sup>Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA, USA (restrin@stanford.edu); <sup>b</sup>Department of Computer Science and Department of Mathematics, University of British Columbia, Vancouver, BC, Canada (mpf@cs.ubc.ca). Research supported by ONR award N00014-16-1-2242

July 31, 2018

**Level-set methods for convex optimization are predicated on the idea that certain problems can be parameterized so that their solutions can be recovered as the limiting process of a root-finding procedure. This idea emerges time and again across a range of algorithms for convex problems. Here we demonstrate that strong duality is a necessary condition for the level-set approach to succeed. In the absence of strong duality, the level-set method identifies  $\epsilon$ -infeasible points that do not converge to a feasible point as  $\epsilon$  tends to zero. The level-set approach is also used as a proof technique for establishing sufficient conditions for strong duality that are different from Slater’s constraint qualification.**

convex analysis | constrained optimization | duality | level-set methods

Duality in convex optimization may be interpreted as a notion of the sensitivity of an optimization problem to perturbations of its data. Similar notions of sensitivity appear in numerical analysis, where the effects of numerical errors on the stability of the computed solution are of central concern. Indeed, backward-error analysis (1, §1.5) describes the related notion that computed approximate solutions may be considered as exact solutions of perturbations of the original problem. It is natural, then, to ask if duality can help us to understand the behavior of a class of numerical algorithms for convex optimization. In this paper, we describe how the level-set method (2–4) produces an incorrect solution when applied to a problem for which strong duality fails to hold. In other words, the level-set method cannot succeed if there does not exist a dual pairing that is tight. This failure of strong duality indicates that the stated optimization problem is brittle, in the sense that its value as a function of small perturbations to its data is discontinuous; this violates a vital assumption needed for the level-set method to succeed.

Consider the convex optimization problem

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad g(x) \leq 0, \quad [\text{P}]$$

where  $f$  and  $g$  are closed proper convex functions that map  $\mathbb{R}^n$  to the extended real line  $\mathbb{R} \cup \{\infty\}$ , and  $\mathcal{X}$  is a convex set in  $\mathbb{R}^n$ . Let the optimal value be  $\tau_p^* := \inf(\text{P}) < \infty$ , so that (P) is feasible. In the context of level-set methods, we may think of  $g(x) \leq 0$  as representing a constraint that poses a computational challenge. For example, there may not exist any efficient algorithm that can compute the projection onto the constraint set  $\{x \in \mathcal{X} \mid g(x) \leq 0\}$ . In many important cases, the objective function has a useful structure that makes it computationally convenient to swap the roles of the objective  $f$  with the constraint  $g$ , and to instead solve the *level-set problem*

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad g(x) \quad \text{subject to} \quad f(x) \leq \tau, \quad [\text{Q}_\tau]$$

where  $\tau$  is an estimate of the optimal value  $\tau_p^*$ . If  $\tau \approx \tau_p^*$ , the level-set constraint  $f(x) \leq \tau$  ensures that a solution  $x_\tau \in \mathcal{X}$  of this problem causes  $f(x_\tau)$  to have a value near  $\tau_p^*$ . If, additionally,  $g(x_\tau) \leq 0$ , then  $x_\tau$  is a nearly optimal and feasible solution for (P). The trade-off for this potentially more convenient problem is that we must compute a sequence of parameters  $\tau_k$  that converges to the optimal value  $\tau_p^*$ .

Define the optimal-value function, or simply the *value function*, of  $(\text{Q}_\tau)$  by

$$v(\tau) = \inf_{x \in \mathcal{X}} \{g(x) \mid f(x) \leq \tau\}. \quad [1]$$

This definition suggests that if the constraint in (P) is *active* at a solution (i.e.,  $g(x) = 0$ ), then  $\tau_p^*$  is a root of the equation

$$v(\tau) = 0,$$

and in particular, the leftmost root:

$$\tau_p^* = \inf \{ \tau \mid v(\tau) = 0 \}. \quad [2]$$

The surprise is that this is not always true.

Fix  $\tau_d^*$  to be the optimal value of the Lagrange-dual problem (5, §1) to (P). If strong duality fails for (P)—i.e.,  $\tau_p^* > \tau_d^*$ —in fact it holds that

$$\tau_d^* = \inf \{ \tau \mid v(\tau) = 0 \} < \tau_p^*,$$

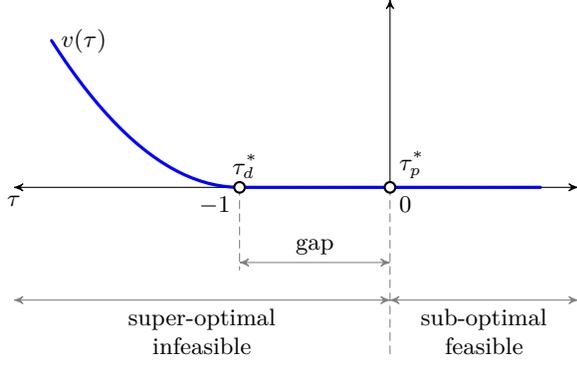
and the inequality is tight only when strong duality holds, i.e.,  $\tau_p^* = \tau_d^*$ . Thus, a root-finding algorithm—such as bisection or Newton’s method—implemented so as to yield the leftmost root of the equation Eq. (1), converges to a value of  $\tau$  that prevents  $(\text{Q}_\tau)$  from attaining an optimal solution. This idea is illustrated in Figure 1 and manifested by Example 2.

Our discussion identified  $\tau_d^*$  as the optimal value of the Lagrange dual problem. However, the same conclusion holds when we consider any dual pairing that arises from a convex perturbation. The following theorem, proven in a later section, encapsulates the main result of this paper. The inequality on  $v$  in the following statement accommodates the possibility that the constraint in (P) is not active at the solution, and so applies more generally than the preceding discussion.

**Theorem 1.** *For the value function  $v$ ,*

$$\tau_d^* = \inf \{ \tau \mid v(\tau) \leq 0 \} \quad \text{and} \quad v(\tau) \leq 0 \quad \text{for all} \quad \tau > \tau_d^*.$$

Note that the theorem does not address conditions under which  $v(\tau_p^*) \leq 0$ , which is true if and only if the solution set  $\text{argmin}(\text{P})$  is not empty. In particular, any  $x^* \in \text{argmin}(\text{P})$  is a solution of  $(\text{Q}_\tau)$  for  $\tau = \tau_p^*$ , and hence  $v(\tau_p^*) \leq 0$ . However, if  $\text{argmin}(\text{P})$  is empty, then there is no solution to  $(\text{Q}_\tau)$  and hence  $v(\tau_p^*) = +\infty$ .



**Fig. 1.** A sketch of the value function  $v$  for Example 2; cf. Eq. (9). The value function  $v(\tau)$  vanishes for all  $\tau \geq \tau_d^*$ , where  $\tau_d^* < \tau_p^*$ . This causes the level-set method to converge to an incorrect solution when strong duality fails to hold.

We do not assume that readers are experts in convex duality theory, so we present an abbreviated summary of the machinery needed to develop the proof of Theorem 1. We also describe a generalized version of the level-set pairing between the problems (P) and  $(Q_\tau)$  and thus establish Theorem 4, which generalizes Theorem 1. We show how this theory can be used as a theoretical device to establish sufficient conditions for strong duality.

## Level-set methods

The technique of exchanging the roles of the objective and constraint functions has a long history. For example, the isoperimetric problem—which dates back to the second century B.C.E.—seeks the maximum area that can be circumscribed by a curve of fixed length (6). The converse problem seeks the minimum-length curve that encloses a certain area. Both problems yield the same circular solution. The mean-variance model of financial portfolio optimization pioneered by Markovitz (7), is another example. It can be phrased as either the problem of allocating assets that minimize risk (i.e., variance) subject to a specified mean return, or as the problem of maximizing the mean return subject to a specified risk. The correct choice of parameters causes both problems to have the same solution.

The idea of rephrasing an optimization problem as root-finding problem appears often in the optimization literature. The celebrated Levenberg-Marquadt algorithm (8, 9), and trust-region methods (10) more generally, use a root-finding procedure to solve a parameterized version of the optimization problem. The widely-used SPGL1 software package for sparse optimization (11) implements the level-set method for obtaining sparse solutions of linear least-squares and underdetermined linear systems (12, 13).

In practice, only an approximate solution of the problem (P) is required, and the level-set method can be used to obtain an approximate root that satisfies  $v(\tau) \leq \epsilon$ . The solution  $x \in \mathcal{X}$  of the corresponding level-set problem  $(Q_\tau)$  is super-optimal and  $\epsilon$ -infeasible:

$$f(x) \leq \tau_p^* \quad \text{and} \quad g(x) \leq \epsilon. \quad [3]$$

Aravkin et al. (4) give a full description of the general algorithm, together with a complexity analysis showing only

$\mathcal{O}(\log(1/\epsilon))$  approximate evaluations of  $v$  are required to obtain an  $\epsilon$ -infeasible solution, where the subproblem accuracy is proportional to the final required accuracy specified by Eq. (3).

**Problem formulation.** The formulation (P) is very general, even though the constraint  $g(x) \leq 0$  represents only a single function of the full constraint set represented by  $\mathcal{X}$ . There are various avenues for reformulating any combination of constraints that lead to this formulation. For instance, Example 2 demonstrates how multiple linear constraints of the form  $Ax = b$  can be represented as a constraint on the norm of the residual, i.e.,  $g(x) = \|Ax - b\| \leq 0$ . More generally, for any set of constraints  $c(x) \leq 0$  where  $c = (c_i)$  is a vector of functions  $c_i$ , we may set  $g(x) = \rho(\max\{0, c(x)\})$  for any convenient nonnegative convex function  $\rho$  that vanishes only at the origin.

## Examples

Some examples will best illustrate the behavior of the value function. The first two are semidefinite programs (SDPs) for which strong duality fails to hold. These demonstrate that the level-set method can produce diverging iterates. The last example is a 1-norm regularized least-squares problem. It confirms that the level-set method succeeds in that important case, often used in solving problems that arise in applications of sparse recovery (14) and compressed sensing (15, 16).

For the first two examples below, let  $x_{ij}$  denote the  $(i, j)$ th entry of the  $n$ -by- $n$  symmetric matrix  $X = (x_{ij})$ . The notation  $X \succeq 0$  denotes the requirement that  $X$  is symmetric positive semidefinite.

**Example 1** (SDP with infinite gap). Consider the  $2 \times 2$  SDP

$$\underset{X \succeq 0}{\text{minimize}} \quad -2x_{21} \quad \text{subject to} \quad x_{11} = 0, \quad [4]$$

whose solution and optimal value are given, respectively, by

$$X_* = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \tau_p^* = 0.$$

The Lagrange dual is a feasibility problem:

$$\underset{y \in \mathbb{R}}{\text{maximize}} \quad 0 \quad \text{subject to} \quad \begin{bmatrix} y & -1 \\ -1 & 0 \end{bmatrix} \succeq 0.$$

Because the dual problem is infeasible, we assign the dual optimal value  $\tau_d^* = -\infty$ . Thus,  $\tau_d^* = -\infty < \tau_p^* = 0$ , and this dual pairing fails to have strong duality.

The application of the level-set method to the primal problem Eq. (4) can be accomplished by defining the functions

$$f(X) := -2x_{21} \quad \text{and} \quad g(X) := |x_{11}|,$$

which together define the value function of the level-set problem  $(Q_\tau)$ :

$$v(\tau) = \inf_{X \succeq 0} \left\{ |x_{11}| \mid -2x_{21} \leq \tau \right\}. \quad [5]$$

Because  $X^*$  is primal optimal,  $v(\tau) = 0$  for all  $\tau \geq \tau_p^* = 0$ . Now consider the parametric matrix

$$X(\tau, \epsilon) := \begin{bmatrix} \epsilon & \frac{\tau}{2} \\ \frac{\tau}{2} & \frac{\tau^2}{4\epsilon} \end{bmatrix} \quad \text{for all} \quad \tau < 0 \quad \text{and} \quad \epsilon > 0,$$

which is feasible for the level-set problem Eq. (5)—i.e.,  $v(\tau)$  is finite. The level-set problem clearly has a zero lower bound

that can be approached by sending  $\epsilon \downarrow 0$ . Thus,  $v(\tau) = 0$  for all  $\tau < 0$ .

In summary,  $v(\tau) = 0$  for all  $\tau$ , and so  $v(\tau)$  has roots less than the true optimal value  $\tau_p^*$ . Furthermore, for  $\tau < 0$ , there is no primal attainment for Eq. (1), because  $\lim_{\epsilon \downarrow 0} X(\tau, \epsilon)$  does not exist.  $\square$

**Example 2** (SDP with finite gap). Consider the  $3 \times 3$  SDP

$$\underset{X \succeq 0}{\text{minimize}} \quad -2x_{31} \quad \text{subject to} \quad x_{11} = 0, \quad x_{22} + 2x_{31} = 1. \quad [6]$$

The positive semidefinite constraint on  $X$ , together with the constraint  $x_{11} = 0$ , implies that  $x_{31}$  must vanish. Thus, the solution and optimal value are given, respectively, by

$$X^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \tau_p^* = 0. \quad [7]$$

The Lagrange dual problem is

$$\underset{y \in \mathbb{R}^2}{\text{maximize}} \quad -y_2 \quad \text{subject to} \quad \begin{bmatrix} y_1 & 0 & y_2 - 1 \\ 0 & y_2 & 0 \\ y_2 - 1 & 0 & 0 \end{bmatrix} \succeq 0.$$

The dual constraint requires  $y_2 = 1$ , and thus the optimal dual value is  $\tau_d^* = -1 < 0 = \tau_p^*$ .

For the application of the level-set method to the primal problem Eq. (6), we can assign

$$f(X) := -2x_{31} \quad \text{and} \quad g(X) := x_{11}^2 + (x_{22} + 2x_{31} - 1)^2, \quad [8]$$

which together define the value function

$$v(\tau) = \inf_{X \succeq 0} \{ x_{11}^2 + (x_{22} + 2x_{31} - 1)^2 \mid -2x_{31} \leq \tau \}. \quad [9]$$

As in [Example 1](#), any convex nonnegative  $g$  function that vanishes on the feasible set could have been used to define  $v$ . It follows from Eq. (7) that  $v(\tau) = 0$  for all  $\tau \geq 0$ . Also, it can be verified that  $v(\tau) = 0$  for all  $\tau \geq \tau_d^* = -1$ . To understand this, first define the parametric matrix

$$X_\epsilon = \begin{bmatrix} \epsilon & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{4\epsilon} \end{bmatrix} \quad \text{with} \quad \epsilon > 0,$$

which is feasible for the level-set problem Eq. (9), and has objective value  $g(X_\epsilon) = \epsilon^2$ . Because  $X_\epsilon$  is feasible for all positive  $\epsilon$ , the optimal value vanishes because  $v(\tau) = \inf \{ g(X_\epsilon) \mid \epsilon > 0 \} = 0$ . Moreover, the set of minimizers for Eq. (9) is empty for all  $\tau \in (-1, 0)$ . [Figure 1](#) illustrates the behavior of this value function.

The level-set method fails because the root of  $v(\tau)$  identifies an incorrect optimal primal value  $\tau_p^*$ , and instead identifies the optimal dual value  $\tau_d^* < \tau_p^*$ .  $\square$

The proof of [Theorem 1](#) reveals that the behavior exhibited by [Examples 1](#) and [2](#) stems from failure of strong duality with respect to perturbations in the linear constraints. In the case of [Example 2](#), we can produce a sequence of matrices  $X_\epsilon$  each of which is  $\epsilon$ -infeasible with respect to the infeasibility measure  $g$ —cf. Eq. (8). However, the limit as  $\epsilon \downarrow 0$  does not produce a feasible point, and the limit does not even exist because the entry  $x_{33}$  of  $X_\epsilon$  goes to infinity.

**Example 3** (Basis pursuit denoising ([14](#), [17](#))). The level-set method implemented in the SPGL1 software package solves the 1-norm regularized least-squares problem

$$\underset{x}{\text{minimize}} \quad \|x\|_1 \quad \text{subject to} \quad \|Ax - b\|_2 \leq \sigma$$

for any value of  $\sigma \geq 0$ , assuming that the problem remains feasible. (The case  $\sigma = 0$  is important, as it accommodates the case in which we seek a sparse solution to the under-determined linear system  $Ax = b$ .) The algorithm approximately solves a sequence of flipped problems

$$\underset{x}{\text{minimize}} \quad \|Ax - b\|_2 \quad \text{subject to} \quad \|x\|_1 \leq \tau_k,$$

where  $\tau_k$  is chosen so that the corresponding solution  $x_k$  satisfies  $\|Ax_k - b\|_2 \approx \sigma$ . Strong duality holds because the domains of the nonlinear functions (i.e., the 1- and 2-norms) cover the whole space; see [Rockafellar and Wets \(5, Theorem 11.39\)](#), which is also summarized by [Theorem 3](#).  $\square$

**Equivalence between value functions.** The level-set method is based on a kind of inverse-function relationship between the value function Eq. (1) of the flipped problem and the value function of the perturbed original problem

$$p(\sigma) = \inf_{x \in \mathcal{X}} \{ f(x) \mid g(x) \leq \sigma \}.$$

Clearly,  $\tau_p^* = p(0)$ . [Aravkin et al. \(18\)](#) describe the formal relationship between the value functions  $v$  and  $p$ , and their respective solutions. We summarize the key aspects here.

Use  $\operatorname{argmin} v(\tau)$  and  $\operatorname{argmin} p(\sigma)$ , respectively, to denote the set of solutions to the optimization problem underlying the value functions  $v$  and  $p$  (which may be empty). Thus, for example, if  $p(\sigma) < \infty$ ,

$$\operatorname{argmin} p(\sigma) = \{ x \in \mathcal{X} \mid f(x) = p(\sigma), \quad g(x) \leq 0 \},$$

and  $\operatorname{argmin} p(\sigma)$  is empty otherwise. Clearly,  $\operatorname{argmin} p(0) = \operatorname{argmin} (P)$ . Because  $p$  is defined via an infimum,  $\operatorname{argmin} p(\sigma)$  can be empty even if  $p$  is finite, in which case we say that the optimal value  $p(\sigma)$  is not attained.

Let  $\mathcal{S}$  be the set of parameters  $\tau$  for which the level-set constraint  $f(x) \leq \tau$  of  $(Q_\tau)$  holds with equality. Formally,

$$\mathcal{S} = \{ \tau \leq +\infty \mid \emptyset \neq \operatorname{argmin} v(\tau) \subseteq \{ x \in \mathcal{X} \mid f(x) = \tau \} \}.$$

**Theorem 2** (Value-function inverses ([18, Theorem 2.1](#))). *For every  $\tau \in \mathcal{S}$ , the following statements hold:*

- (a)  $(p \circ v)(\tau) = \tau$ ,
- (b)  $\operatorname{argmin} v(\tau) = \operatorname{argmin} p(v(\tau)) \subseteq \{ x \in \mathcal{X} \mid f(x) = v(\tau) \}$ .

The condition  $\tau \in \mathcal{S}$  means that the constraint of the level-set problem  $(Q_\tau)$  must be active in order for the result to hold. The following example illustrates the need for this condition.

**Example 4.** The univariate problem

$$\underset{x \in \mathbb{R}}{\text{minimize}} \quad |x| \quad \text{subject to} \quad |x| - 1 \leq 0$$

has the obvious solution  $x^* = 0$  and optimal value  $\tau_p^* = 0$ , and an inactive constraint. The corresponding level-set problem

$$\underset{x \in \mathbb{R}}{\text{minimize}} \quad |x| - 1 \quad \text{subject to} \quad |x| \leq \tau$$

shares the same solution  $x^*$  when  $\tau = \tau_p^*$ . However the corresponding value function for the level-set problem is

$$v(\tau) = \begin{cases} -1 & \text{if } \tau \geq 0 \\ +\infty & \text{if } \tau < 0, \end{cases}$$

so it has the wrong optimal value.  $\square$

This theorem is symmetric, and holds if the roles of  $f$  and  $g$ , and  $p$  and  $v$ , are reversed. We note that this result holds even if any of the underlying objects that define (P) are not convex.

Part (b) of the theorem confirms that if  $\tau_p^* \in \mathcal{S}$ —i.e., the constraint  $g(x) \leq 0$  holds with equality at a solution of (P)—then solutions of the level-set problem coincide with solution of the original problem defined by  $p(0)$ . More formally,

$$\operatorname{argmin} v(\tau_p^*) = \operatorname{argmin} (\mathbf{P}).$$

Again consider [Example 2](#), where we set  $\tau = -1/2$ , which falls midway between the interval  $(\tau_d^*, \tau_p^*) = (-1, 0)$ . Because the solution set  $\operatorname{argmin} v(\tau)$  is empty,  $\tau \notin \mathcal{S}$ . Thus,

$$(p \circ v)(\tau) = p(0) = 0 \neq \tau,$$

and the level-set method fails.

In order to establish an inverse-function-like relationship between the value functions  $p$  and  $v$  that always holds for convex problems, we provide a modified definition of the epigraphs for  $v$  and  $w$ .

**Definition 1** (Value function epigraph). The *value function epigraph* of the optimal value function  $p(\sigma)$

$$\operatorname{vfepi} p = \{ (\sigma, \tau) \mid \exists x \in \mathcal{X}, f(x) \leq \tau, g(x) \leq \sigma \}.$$

This definition is almost exactly the same as the regular definition for the epigraph of a function, given by

$$\operatorname{epi} p = \{ (\sigma, \tau) \mid p(\sigma) \leq \tau \},$$

except that if  $\tau = p(\sigma)$  but  $\operatorname{argmin} p(\sigma)$  is empty, then  $(\sigma, \tau) \notin \operatorname{vfepi} w$ .

The result below follows immediately from the definition of the value function epigraph. It establishes that [Eq. \(2\)](#) holds if  $(\mathbf{Q}_\tau)$  has a solution that attains its optimal value (as opposed to relying on the infimal operator to achieve that value).

**Proposition 1.** For the value functions  $p$  and  $v$ ,

$$(\sigma, \tau) \in \operatorname{vfepi} p \iff (\tau, \sigma) \in \operatorname{vfepi} v.$$

## Duality in convex optimization

Duality in convex optimization can be understood as describing the behavior of an optimization problem under perturbation to its data. From this point of view, dual variables describe the sensitivity of the problem's optimal value to that perturbation. The description that we give here summarizes a well-developed theory fully described by Rockafellar and Wets (5). We adopt a geometric viewpoint that we have found helpful for understanding the connection between duality and the level-set method.

For this section only, consider the generic convex optimization problem

$$\underset{x \in \mathcal{X}}{\operatorname{minimize}} \quad f(x)$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$  is an arbitrary closed proper convex function. The perturbation approach is predicated on fixing an arbitrary convex function  $F(x, u): \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{\infty\}$  with the property that

$$F(x, 0) = f(x) \quad \forall x.$$

([Example 5](#) below describes an instance of such a function.) The choice of  $F$  determines the perturbation function

$$p(u) := \inf_x F(x, u),$$

which describes how the optimal value of  $f$  changes under a perturbation  $u$ . We seek the behavior of the perturbation function about the origin, at which the value of  $p$  coincides with the optimal value  $\tau_p^*$ , i.e.,  $p(0) = \tau_p^*$ .

A dual optimization problem is constructed by considering the set of affine functions  $u \mapsto \langle \mu, u \rangle + q$  parameterized by  $(\mu, q) \in \mathbb{R}^m \times \mathbb{R}$  that minorize  $p$ :

$$\langle \mu, u \rangle + q \leq p(u) \quad \forall u. \quad [10]$$

The tightest lower minorant is obtained by maximizing the left-hand side  $\langle \mu, u \rangle + q$  over  $(\mu, q)$  subject to the constraint above. This can be accomplished by first eliminating  $q$  and setting it to

$$q \equiv q_\mu := -\sup_u \{ \langle \mu, u \rangle - p(u) \} \equiv -p^*(\mu),$$

where

$$p^*(\mu) = \sup_u \{ \langle \mu, u \rangle - p(u) \}$$

is the convex conjugate of  $p$ . [Figure 2\(a\)](#) illustrates how the choice of  $q_\mu$  causes the affine function to support the epigraph of  $p$ . Thus, [Eq. \(10\)](#) is equivalent to the expression

$$\langle \mu, u \rangle - p^*(\mu) \leq p(u) \quad \forall u.$$

Now take the supremum over  $\mu$  of the left-hand side, which results in the value

$$p^{**}(u) = \sup_\mu \{ \langle \mu, u \rangle - p^*(\mu) \}.$$

It is thus evident that the biconjugate  $p^{**}$ , which is necessarily closed convex, provides a global lower envelope for  $p$ , i.e.,

$$p^{**}(u) \leq p(u) \quad \forall u.$$

This inequality is tight at a point  $u$ —i.e.,  $p^{**}(u) = p(u)$ —if and only if  $p$  is lower-semicontinuous at  $u$ . Because of the connection between lower semicontinuity and the closure of the epigraph, we say that  $p$  is *closed* at such points  $u$ ; see Rockafellar (19, Theorem 7.1).

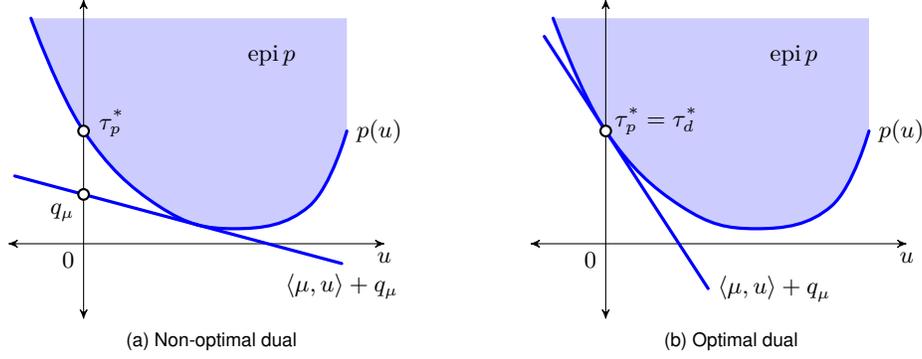
Weak duality expresses the following relationship between the optimal primal and dual values:

$$\tau_d^* \equiv p^{**}(0) \leq p(0) \equiv \tau_p^*. \quad [11]$$

Strong duality holds when  $\tau_p^* = \tau_d^*$ , which is equivalent to the closure of  $p$  at the origin.

The functions  $p$  and  $p^{**}$ , evaluated at the origin, define dual pairs of optimization problems given by

$$p(0) = \inf_x F(x, 0) \quad \text{and} \quad p^{**}(0) = \sup_y -F^*(0, y). \quad [12]$$



**Fig. 2.** The relationship between the primal perturbation  $p(u)$  and a single instance (with slope  $\mu$  and intercept  $q_\mu$ ) of the uncountably many minorizing affine functions that define the dual problem. The panel on the left depicts a non-optimal supporting hyperplane that generates a value  $q_\mu < \tau_p^*$ ; the panel on the right depicts an optimal supporting hyperplane that generates a slope  $\mu$  that causes  $q_\mu \equiv \tau_d^* = \tau_p^*$ .

The expression for  $p^{**}(0)$  is a straightforward application of the conjugate operator.

The following theorem confirms the *weak duality* that always holds between the optimal values  $\tau_p^*$  and  $\tau_d^*$ . It also establishes sufficient conditions that guarantee when the perturbation function  $p$  is closed at the origin, and hence strong duality holds. The complete version of this theorem includes statements regarding the relationships between the subdifferentials of  $p$  and  $q$  at the origin (5, Theorem 11.39).

**Theorem 3** (Weak and strong duality). *Consider the primal-dual pair Eq. (12).*

- (a) [Weak duality] *The inequality  $\tau_p^* \geq \tau_d^*$  always holds.*
- (b) [Strong duality] *If  $0 \in \text{int dom } p$ , then  $\tau_p^* = \tau_d^*$ .*

**Example 5.** Consider the convex optimization problem

$$\underset{x}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad c(x) \leq 0, \quad Ax = b, \quad [13]$$

where  $c$  is a vector-valued convex function and  $A$  is a matrix. Introduce perturbations  $u_1$  and  $u_2$  to the right-hand sides of the constraints, which gives rise to Lagrange duality, and corresponds to the perturbation function

$$p(u_1, u_2) = \inf_x \{ f(x) \mid c(x) \leq u_1, \quad Ax - b = u_2 \}.$$

One valid choice for the value function that corresponds to swapping both constraints with the objective to Eq. (13) can be expressed as

$$v(\tau) = \inf_{x, u_1, u_2} \left\{ \frac{1}{2} \| [u_1]_+ \|^2 + \frac{1}{2} \| u_2 \|^2 \mid \begin{array}{l} f(x) \leq \tau \\ c(x) \leq u_1 \\ Ax - b = u_2 \end{array} \right\},$$

where the operator  $[u_1]_+ = \max\{0, u_1\}$  is taken component-wise on the elements of  $u_1$ . This particular formulation of the value function makes explicit the connection to the perturbation function. We may thus interpret the value function as giving the minimal perturbation that corresponds to an objective value less than or equal to  $\tau$ .  $\square$

### A duality gap gives a false root

We discuss some intuition as to why we can expect level-set methods to return incorrect solutions when applied to problems without strong duality. Apply the level-set method to (P) and for simplicity, suppose that (P) attains its optimal value and that the constraint  $g(x) \leq 0$  is active at any solution. Consider the perturbation function

$$p(u) = \inf_{x \in \mathcal{X}} \{ f(x) \mid g(x) \leq u \}, \quad [14]$$

and the value function  $v(\tau)$  given by Eq. (1). First, suppose that strong duality holds for the problem under the given perturbation, so then  $p(0) = p^{**}(0)$ , which means that the perturbation function  $p$  is closed at the origin. We sketch an example  $p(u)$  and the corresponding  $v(\tau)$  in the top row of Figure 3.

To understand this picture, first consider the value  $\tau_1 < \tau_p^*$ , as in the top row of Figure 3. It is evident that  $v(\tau_1)$  is positive, because otherwise there must exist a vector  $x \in \mathcal{X}$  that is super-optimal ( $f(x) \leq \tau_1 < \tau_p^*$ ) and feasible ( $g(x) \leq 0$ ), which contradicts the definition of  $\tau_p^*$ . It then follows that the value  $u := v(\tau_1)$  yields  $p(u) = \tau_1$ . For  $\tau_2 > \tau^*$ , any solution to the original problem would be feasible (therefore requiring no perturbation  $u$ ) and would achieve objective value  $p(0) = \tau_p^* < \tau_2$ . Furthermore, notice that as  $\tau_1 \rightarrow \tau_p^*$ , the value  $p(u_1)$  varies continuously in  $\tau_1$ , where  $u_1$  is the smallest root of  $p(u) = \tau_1$ .

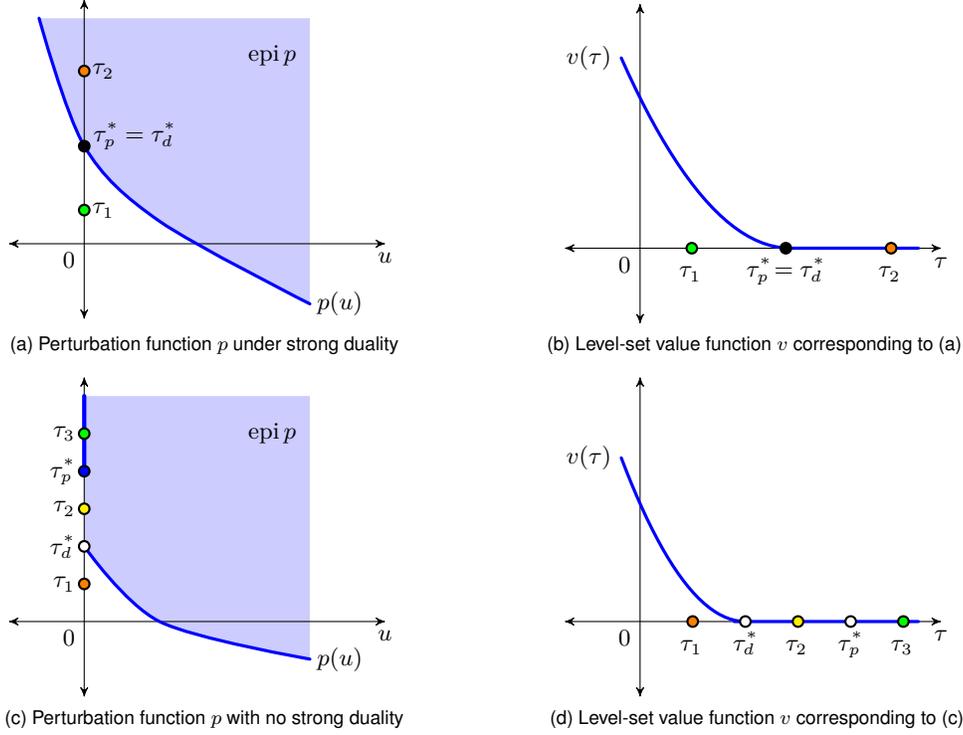
Next consider the second row of Figure 3. In this case, strong duality fails, which means that

$$\lim_{u \downarrow 0} p(u) = \tau_d^* \neq p(0).$$

With  $\tau = \tau_1$ , we have  $v(\tau_1) > 0$ . With  $\tau = \tau_3 > \tau_p^*$ , we have  $v(\tau) = 0$  because any solution to (P) causes  $(Q_\tau)$  to have zero value. But we see that for  $\tau_d^* < \tau_2 < \tau_p^*$ ,  $v(\tau_2) = 0$ , because for any positive  $\epsilon$  there exists positive  $u < \epsilon$  such that  $p(u) \leq \tau_2$ . Even though there is no feasible point that achieves a superoptimal value  $f(x) \leq \tau_2 < \tau_p^*$ , for any positive  $\epsilon$  there exists an  $\epsilon$ -infeasible point that achieves that objective value.

**Lagrange duality framework.** Define

$$F(x, u) = \begin{cases} f(x) & \text{if } g(x) \leq u, \\ +\infty & \text{otherwise.} \end{cases}$$



**Fig. 3.** The perturbation function  $p(u)$  and corresponding level-set value function  $v(\tau)$  for problems with strong duality (top row) and no strong duality (bottom row). Panel (c) illustrates the case when strong duality fails and the graph of  $p$  is open at the origin, which implies that  $\tau_d^* < \tau_p^* \equiv p(0)$ .

The function  $F$  is convex in both of its arguments, and gives rise to the perturbation function  $p$  defined by Eq. (14). In summary, this choice of perturbation leads to the pair

$$p(u) = \inf_{x \in \mathcal{X}} \{ f(x) \mid g(x) \leq u \}, \quad [15a]$$

$$v(\tau) = \inf_{x \in \mathcal{X}} \{ g(x) \mid f(x) \leq \tau \}, \quad [15b]$$

as defined by Eqs. (1) and (14).

**Proof of Theorem 1.** We first prove the second result that  $v(\tau) \leq 0$  if  $\tau > \tau_d^*$ . Suppose that strong duality holds, i.e.,  $\tau_p^* = \tau_d^*$ . Then the required result is immediate because if  $\tau_p^*$  is the optimal value, then for any  $\tau > \tau_p^*$ , there exists feasible  $x$  such that  $f(x) \leq \tau$ .

Suppose that strong duality does not hold, i.e.,  $\tau_p^* > \tau_d^*$ . If  $\tau > \tau_p^*$ , it is immediate that  $v(\tau) \leq 0$ . Assume, then, that  $\tau \in (\tau_d^*, \tau_p^*]$ . Note that the conditions  $g(x) \leq u$  and  $f(x) \leq \tau$  are equivalent to  $F(x, u) \leq \tau$ . We will therefore prove that

$$\forall \epsilon > 0, \exists x \in \mathcal{X} \text{ such that } F(x, u) \leq \tau, \quad u \leq \epsilon, \quad [16]$$

which is equivalent to the required condition  $v(\tau) \leq 0$ . It follows from the convexity of  $\text{epi } p$  and from Eq. (11) that  $(0, \tau_d^*) \in \text{epi } p^{**} = \text{cl epi } p$ . Thus,

$$\forall \eta > 0, \exists (u, \omega) \in \text{epi } p \text{ such that } \|(u, \omega) - (0, \tau_d^*)\| < \eta.$$

Note that

$$\begin{aligned} \liminf_{\epsilon \downarrow 0} \{ p(u) \mid |u| \leq \epsilon \} &\stackrel{(i)}{=} \liminf_{\epsilon \downarrow 0} \{ p^{**}(u) \mid |u| \leq \epsilon \} \\ &\stackrel{(ii)}{=} p^{**}(0) \stackrel{(iii)}{=} \tau_d^*, \end{aligned} \quad [17]$$

where equality (i) follows from the fact that  $p(u) = p^{**}(u)$  for all  $u \in \text{dom } p$ , equality (ii) follows from the closure of  $p^{**}$ , and (iii) follows from Eq. (11). This implies that

$$\forall \eta > 0, \exists (u, \omega) \in \text{epi } p \text{ such that } \|(u, p(u)) - (0, \tau_d^*)\| < \eta.$$

For any fixed positive  $\epsilon$  define  $\mu = \min \{ \epsilon, \frac{1}{4}(\tau - \tau_d^*) \}$ . Choose  $\hat{u} \in \text{dom } p$  such that  $\|(\hat{u}, p(\hat{u})) - (0, \tau_d^*)\| < \mu$ , and so

$$\epsilon \geq \mu > \|(\hat{u}, p(\hat{u})) - (0, \tau_d^*)\| \geq \max \{ \|\hat{u}\|, |p(\hat{u}) - \tau_d^*| \}.$$

Thus,

$$p(\hat{u}) < \tau_d^* + \mu. \quad [18]$$

Moreover, it follows from the definition of  $p(\hat{u})$ —cf. Eq. (14)—that

$$\forall \nu > 0, \exists x \in \mathcal{X} \text{ such that } F(x, \hat{u}) \leq p(\hat{u}) + \nu.$$

Choose  $\nu = \mu$ , and so there exists  $\hat{x}$  such that  $F(\hat{x}, \hat{u}) \leq p(\hat{u}) + \mu$ . Together with Eq. (18), we have

$$f(\hat{x}) \leq p(\hat{u}) + \mu < \tau_d^* + 2\mu \leq \tau.$$

Therefore, for each  $\epsilon > 0$ , we can find a pair  $(\hat{x}, \hat{u})$  that satisfies Eq. (16), which completes the proof of the second result.

Next we prove the first result, which is equivalent to proving that  $v(\tau) > 0$  if  $\tau < \tau_d^*$  because  $v(\tau)$  is convex. Observe that  $\tau < \tau_d^* \equiv p^{**}(0)$  is equivalent to  $(0, \tau) \notin \text{cl epi } p$ , which implies that

$$\begin{aligned} 0 &< \inf_u \{ u \mid (u, \tau) \in \text{cl epi } p \} \\ &= \inf_u \{ u \mid (u, \tau) \in \text{epi } p \} \\ &= \inf_u \{ u \mid \exists x \in \mathcal{X} \text{ such that } F(x, u) \leq \tau \} = v(\tau), \end{aligned} \quad [19]$$

which completes the proof.  $\square$

**General duality framework.** We generalize the notion of the level-set method to arbitrary perturbations and therefore more general notions of duality. In this case we are interested in the value function pair

$$p(u) = \inf_{x \in \mathcal{X}} F(x, u), \quad [20a]$$

$$v(\tau) = \inf_{x \in \mathcal{X}} \{ \|u\| \mid F(x, u) \leq \tau \}, \quad [20b]$$

where  $\|\cdot\|$  is any norm. Because  $p$  is parameterized by a vector  $u$ , we must consider the norm of the perturbation, rather than just its value. Therefore,  $v(\tau)$  is necessarily non-negative. We are thus interested in the leftmost root of the equation  $v(\tau) = 0$ , rather than an inequality as in [Theorem 1](#).

**Theorem 4.** *For the functions  $p$  and  $v$  defined by Eq. (20),*

$$\tau_d^* = \inf \{ \tau \mid v(\tau) = 0 \} \quad \text{and} \quad v(\tau) = 0 \quad \text{for all } \tau > \tau_d^*.$$

The proof is almost identical to that of [Theorem 1](#), except that we treat  $u$  as a vector, and replace  $u$  by  $\|u\|$  in Eqs. (16), (17), and (19).

[Theorems 1](#) and [4](#) imply that  $v(\tau) \leq 0$  for all values larger than the optimal dual value (the inequality  $\tau > \tau_d^*$  is strict, as  $v(\tau_d^*)$  may be infinite). Thus if strong duality does not hold, then  $v(\tau)$  identifies the wrong optimal value for the original problem that is being solved. This means that the level-set method may provide a point that is arbitrarily close to feasibility, but is at least a fixed distance away from the true solution independent of how close to feasibility the returned point may be.

### Sufficient conditions for strong duality

The condition that  $0 \in \text{dom } p$  may be interpreted as Slater's constraint qualification ([20](#), §3.2), which in the context of (P) requires that there exist a point  $\hat{x}$  in the domain of  $f$  and for which  $g(\hat{x}) < 0$ . This condition is sufficient to establish strong duality. Here we show how [Theorem 1](#) can be used as a device to characterize an alternative set of sufficient conditions that continue to ensure strong duality even for problems that do not satisfy Slater's condition.

**Proposition 2.** *Problem (P) satisfies strong duality if either one of the following conditions hold:*

- (a) *the objective  $f$  is coercive, i.e.,  $f(x) \rightarrow \infty$  as  $\|x\| \rightarrow \infty$ ;*
- (b)  *$\mathcal{X}$  is compact.*

*Proof.* Consider the level-set problem  $(Q_\tau)$  and its corresponding optimal-value function  $v(\tau)$  given by Eq. (1). In either case (a) or (b), the feasible set  $\{x \in \mathcal{X} \mid f(x) \leq \tau\}$  of Eq. (1) is compact because either  $\mathcal{X}$  is compact or the level sets of  $f$  are compact. Therefore,  $(Q_\tau)$  always attains its minimum for all  $\tau \geq \inf \{f(x) \mid x \in \mathcal{X}\}$ .

Suppose strong duality does not hold. [Theorem 1](#) then confirms that there exists a parameter  $\tau \in (\tau_d^*, \tau_p^*)$  such that  $v(\tau) = 0$ . However, because  $(Q_\tau)$  always attains its minimum, there must exist a point  $\hat{x} \in \mathcal{X}$  such that  $f(\hat{x}) \leq \tau < \tau_p^*$  and  $g(\hat{x}) \leq 0$ , which contradicts the fact that  $\tau_p^*$  is the optimal value of (P). We have therefore established that  $\tau_d^* = \tau_p^*$  and hence that (P) satisfies strong duality.  $\square$

We can use [Proposition 2](#) to establish that certain optimization problems that do not satisfy a Slater constraint qualification still enjoy strong duality. As an example, consider the conic optimization problem

$$\underset{x}{\text{minimize}} \quad \langle c, x \rangle \quad \text{subject to} \quad \mathcal{A}x = b, \quad x \in \mathcal{K}, \quad [21]$$

where  $\mathcal{A} : \mathcal{E}_1 \rightarrow \mathcal{E}_2$  is a linear map between Euclidean spaces  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , and  $\mathcal{K} \subseteq \mathcal{E}_1$  is a closed proper convex cone. This wide class of problems includes linear programming (LP), second-order programming (SOCP), and SDPs, and has many important scientific and engineering applications ([21](#)). If  $c$  is contained in the interior of the dual cone  $\mathcal{K}^* = \{y \in \mathcal{E}_1 \mid \langle x, y \rangle \geq 0 \forall x \in \mathcal{K}\}$ , then  $\langle c, x \rangle > 0$  for all feasible  $x \in \mathcal{K}$ . Equivalently, the function  $f(x) := \langle c, x \rangle + \delta_{\mathcal{K}}(x)$  is coercive. Thus, Eq. ([21](#)) is equivalent to the problem

$$\underset{x}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad \mathcal{A}x = b,$$

which has a coercive objective. Thus, Part (a) of [Proposition 2](#) applies, and strong duality holds.

A concrete application of this model problem is the SDP relaxation of the celebrated phase-retrieval problem ([22](#), [23](#))

$$\underset{X}{\text{minimize}} \quad \text{tr}(X) \quad \text{subject to} \quad \mathcal{A}X = b, \quad X \succeq 0, \quad [22]$$

where  $\mathcal{K}$  is now the cone of Hermitian positive semidefinite matrices (i.e., all the eigenvalues are real-valued and nonnegative) and  $c = I$  is the identity matrix, so that  $\langle C, X \rangle = \text{tr}(X)$ . In that setting, ([22](#)) prove that with high probability, the feasible set of Eq. ([21](#)) is a rank-1 singleton (the desired solution), and thus we cannot use Slater's condition to establish strong duality. However, because  $\mathcal{K}$  is self dual ([24](#), Example 2.24), clearly  $c \in \text{int } \mathcal{K}$ , and by the discussion above, we can use [Proposition 2](#) to establish that strong duality holds Eq. ([22](#)).

A consequence of [Proposition 2](#) is that it is possible to modify (P) in order to guarantee strong duality. In particular, we may regularize the objective, and instead consider a version of the problem with the objective as  $f(x) + \mu\|x\|$ , where the parameter  $\mu$  controls the degree of regularization contributed by the regularization term  $\|x\|$ . If, for example,  $f$  is bounded below on  $\mathcal{X}$ , the regularized objective is then coercive and [Proposition 2](#) asserts that the revised problem satisfies strong duality. Thus, the optimal value function of the level-set problem has the correct root, and the level-set method is applicable. For toy problems such as [Examples 1](#) and [2](#), where all of the feasible points are optimal, regularization would not perturb the solution; however, in general we expect that the regularization will perturb the resulting solution—and in some cases this may be the desired outcome.

**ACKNOWLEDGMENTS.** The authors are indebted to Professor Bruno F. Lourenço of Seikei University for fruitful discussions that followed the second author's course on first-order methods at the summer school associated with the 2016 International Conference on Continuous Optimization, held in Tokyo. Professor Lourenço asked if level-set methods could be applied to solve degenerate SDPs. His thinking was that the level-set problems  $(Q_\tau)$  might satisfy Slater's constraint qualification even if the original problem (P) did not, and therefore the level-set method might be useful as a way to alleviate numerical difficulties that can arise when an algorithm is applied directly to an SDP without strong duality. The conclusion of this paper suggests that this is not always the case.

1. Higham NJ (2002) *Accuracy and stability of numerical algorithms*. (Society for Industrial and Applied Mathematics) Vol. 80.
2. van den Berg E, Friedlander MP (2007) SPGL1: A solver for large-scale sparse reconstruction. <http://www.cs.ubc.ca/labs/scl/spgl1>.
3. van den Berg E, Friedlander MP (2008) Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing* 31(2):890–912.
4. Aravkin AY, Burke JV, Drusvyatskiy D, Friedlander MP, Roy S (2017) Level-set methods for convex optimization. *ArXiv e-prints*. Submitted to Mathematical Programming Journal.
5. Rockafellar RT, Wets RJB (1998) *Variational Analysis*. (Springer) Vol. 317. 3rd printing.
6. Wiegert J (2010) The sagacity of circles: A history of the isoperimetric problem (<https://www.maa.org/press/periodicals/convergence/the-sagacity-of-circles-a-history-of-the-isoperimetric-problem>). Accessed: 2018-07-25.
7. Markowitz HM (1987) *Mean-Variance Analysis in Portfolio Choice and Capital Markets*. (Frank J. Fabozzi Associates, New Hope, Pennsylvania).
8. Marquardt D (1963) An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* 11:431–441.
9. Morrison DD (1960) Methods for nonlinear least squares problems and convergence proofs in *Proceedings of the Seminar on Tracking Programs and Orbit Determination*, eds. Lorell J, Yagi F. (Jet Propulsion Laboratory, Pasadena, USA), pp. 1–9.
10. Conn AR, Gould NIM, Toint PhL (2000) *Trust-Region Methods*, MPS-SIAM Series on Optimization. (Society of Industrial and Applied Mathematics, Philadelphia).
11. van den Berg E, Friedlander MP (2013) SPGL1: A solver for large-scale sparse reconstruction.
12. van den Berg E, Friedlander MP (2011) Sparse optimization with least-squares constraints. *SIAM J. Optim.* 21(4):1201–1229.
13. van den Berg E, Friedlander MP (2008) Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.* 31(2):890–912.
14. Chen SS, Donoho DL, Saunders MA (2001) Atomic decomposition by basis pursuit. *SIAM Rev.* 43(1):129–159.
15. Candès EJ (2006) Compressive sampling in *Proceedings of the International Congress of Mathematicians*.
16. Donoho DL (2006) For most large underdetermined systems of linear equations, the minimal  $\ell_1$ -norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics* 59.
17. Chen SS, Donoho DL, Saunders MA (1998) Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* 20(1):33–61.
18. Aravkin AY, Burke J, Friedlander MP (2013) Variational properties of value functions. *SIAM J. Optim.* 23(3):1689–1717.
19. Rockafellar RT (1970) *Convex Analysis*. (Princeton University Press, Princeton).
20. Borwein J, Lewis AS (2010) *Convex analysis and nonlinear optimization: theory and examples*. (Springer Science & Business Media).
21. Ben-Tal A, Nemirovski A (2001) *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, MPS/SIAM Series on Optimization. (Society of Industrial and Applied Mathematics, Philadelphia) Vol. 2.
22. Candès EJ, Strohmer T, Vershynina V (2013) PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming. *Comm. Pure Appl. Math.* 66(8):1241–1274.
23. Waldspurger I, d’Apremont A, Mallat S (2015) Phase recovery, maxcut and complex semidefinite programming. *Math. Program.* 149(1-2):47–81.
24. Boyd S, Vandenberghe L (2004) *Convex optimization*. (Cambridge University Press, Cambridge), pp. xiv+716.