
A perturbation view of level-set methods for convex optimization

Ron Estrin · Michael P. Friedlander

January 17, 2020 (revised May 15, 2020)

Abstract 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 ϵ -infeasible points that do not converge to a feasible point as ϵ 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.

Keywords convex analysis · duality · level-set methods

1 Introduction

Duality in convex optimization may be interpreted as a notion of 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 (Higham 2002, §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 understand the behavior of a class of numerical algorithms for convex optimization. In this paper, we describe how the level-set method (van den Berg and Friedlander 2007, 2008a; Aravkin et al. 2018) 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

R. Estrin
Institute for Computational and Mathematical Engineering
Stanford University, Stanford, CA, USA
E-mail: restrin@stanford.edu

M. P. Friedlander
Department of Computer Science and Department of Mathematics
University of British Columbia
Vancouver, BC, V6R 1Y8, Canada
E-mail: michael@friedlander.io

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 τ_p^* of (P) be finite, which indicates that (P) is feasible. In the context of level-set methods, we may think of the constraint $g(x) \leq 0$ as representing a computational challenge. For example, there may not exist any efficient algorithm to 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 instead to 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 τ is an estimate of the optimal value τ_p^* . The term “level set” points to the feasible set of problem (Q $_\tau$), which is the τ level set $\{x \mid f(x) \leq \tau\}$ of the function f .

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 the optimal value τ_p^* . If, additionally, $g(x_\tau) \leq 0$, then x_τ 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 τ_k that converges to τ_p^* .

1.1 Objective and constraint reversals

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 (Wiegert 2010). 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 Markowitz (1987), 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 parameter choice, such as τ in the case of the level-set problem (Q $_\tau$), causes both problems to have the same solution.

The idea of rephrasing an optimization problem as a root-finding problem appears often in the optimization literature. The celebrated Levenberg-Marquardt algorithm (Marquardt 1963; Morrison 1960), and trust-region methods (Conn et al. 2000) more generally, use a root-finding procedure to solve a parameterized version of the optimization problem. Lemaréchal et al. (1995) develop a root-finding procedure for a level-bundle method for general convex optimization. The widely used SPGL1 software package for sparse optimization (van den Berg and Friedlander 2013) implements the level-set method for obtaining sparse solutions of linear least-squares and underdetermined linear systems (van den Berg and Friedlander 2008b, 2011).

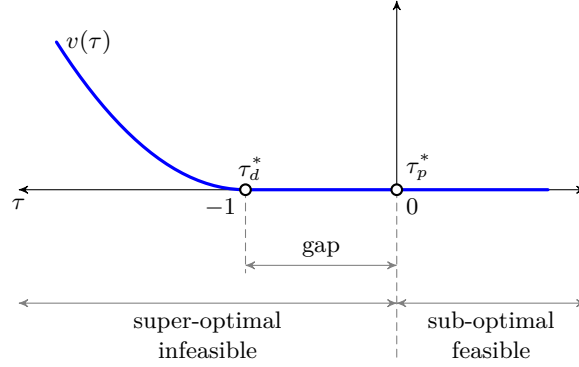


Fig. 1.1 A depiction of a value function v that exhibits the strict inequality described by (1.3); see also Example 2.2. In this example, the value function $v(\tau)$ vanishes for all $\tau \geq \tau_d^*$, where $\tau_d^* < \tau_p^*$. Solutions of (1.1) for values of $\tau < \tau_p^*$ are necessarily super-optimal and infeasible for (P). The difference between τ_d^* and τ_p^* corresponds to the gap between the optimal values of (P) and its dual problem.

1.2 Duality of the value function root

Define the optimal-value function, or simply the *value function*, of (Q_τ) by

$$v(\tau) = \inf_{x \in \mathcal{X}} \{ g(x) \mid f(x) \leq \tau \}. \quad (1.1)$$

If the constraint in (P) is *active* at a solution, that is, $g(x) = 0$, this definition then suggests that the optimal value τ_p^* of (P) is a root of the equation

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

and in particular, is the leftmost root:

$$\tau_p^* = \inf \{ \tau \mid v(\tau) = 0 \}. \quad (1.2)$$

The surprise is that this is not always true.

In fact, as we demonstrate in this paper, the failure of strong duality for (P) implies that

$$\tau_d^* := \inf \{ \tau \mid v(\tau) = 0 \} < \tau_p^*. \quad (1.3)$$

Thus, a root-finding algorithm, such as bisection or Newton's method, implemented so as to yield the leftmost root of the equation $v(\tau) = 0$ will converge to a value of τ that prevents (Q_τ) from attaining a meaningful solution. This phenomenon is depicted in Figure 1.1, and is manifested by the semidefinite optimization problem in Example 2.2. Moreover, the infimal value in (1.3), defined here as τ_d^* , coincides with the optimal value of any dual pairing of (P) that arises from Fenchel-Rockafellar convex duality (Rockafellar and Wets 1998, Theorem 11.39). These results are established by Theorems 5.1 and 5.2.

We do not assume that our readers are experts in convex duality theory, and so we present an abbreviated summary of the machinery needed to develop our main results. We also describe a generalized version of the level-set pairing between the problems (P) and (Q_τ) , and thus establish Theorem 5.2. We show in Section 2 how these theoretical results can be used to establish sufficient conditions for strong duality.

1.3 Level-set methods

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_τ) is super-optimal and ϵ -infeasible:

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

Aravkin et al. (2018) describe the general level-set approach, and establish a complexity analysis that asserts that $\mathcal{O}(\log \epsilon^{-1})$ approximate evaluations of v are required to obtain an ϵ -infeasible solution. These root-finding procedures are based on standard approaches, including bisection, secant, and Newton methods. The efficiency of these approaches hinges on the accuracy required of each evaluation of the value function v . Aravkin et al. also demonstrate that the required complexity can be achieved by requiring a bound on error in each evaluation of v that is proportional to ϵ .

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 a single functional-constraint formulation such as (P). For instance, 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 convex functions c_i , we may set $g(x) = \rho(\max\{0, c(x)\})$ for any convenient nonnegative convex function ρ that vanishes only at the origin, thus ensuring that $g(x) \leq 0$ if and only if $c(x) \leq 0$.

2 Examples

We provide concrete examples that exhibit the behavior shown in (1.3). These semidefinite programs (SDPs) demonstrate that the level-set method can produce diverging iterates.

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 2.1 (SDP with infinite gap) Consider the 2×2 SDP

$$\underset{X \succeq 0}{\text{minimize}} \quad -2x_{21} \quad \text{subject to} \quad x_{11} = 0, \quad (2.1)$$

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 (2.1) 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_τ) :

$$v(\tau) = \inf_{X \succeq 0} \{ |x_{11}| \mid -2x_{21} \leq \tau \}. \quad (2.2)$$

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

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

which is feasible for the level-set problem (2.2). Thus, $v(\tau)$ 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 τ , and so $v(\tau)$ has roots less than the true optimal value τ_p^* . Furthermore, for $\tau < 0$, there is no primal attainment for (1.1), because $\lim_{\epsilon \downarrow 0} X(\tau, \epsilon)$ does not exist. \square

Example 2.2 (SDP with finite gap) Consider the 3×3 SDP

$$\underset{X \succeq 0}{\text{minimize}} \quad -2x_{31} \quad \text{subject to} \quad x_{11} = 0, \quad x_{22} + 2x_{31} = 1. \quad (2.3)$$

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

$$X^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \tau_p^* = 0. \quad (2.4)$$

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 primal problem (2.3), we assign

$$f(X) := -2x_{31} \quad \text{and} \quad g(X) := x_{11}^2 + (x_{22} + 2x_{31} - 1)^2, \quad (2.5)$$

which together define the value function

$$v(\tau) = \inf_{X \succeq 0} \{ x_{11}^2 + (x_{22} + 2x_{31} - 1)^2 \mid -2x_{31} \leq \tau \}. \quad (2.6)$$

As in Example 2.1, any convex nonnegative g function that vanishes on the feasible set could have been used to define v . It follows from (2.4) 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 level-set problem (2.6), and has objective value $g(X_\epsilon) = \epsilon^2$. Because X_ϵ is feasible for all positive ϵ , the optimal value vanishes because $v(\tau) = \inf \{g(X_\epsilon) \mid \epsilon > 0\} = 0$. Moreover, the set of minimizers for (2.6) is empty for all $\tau \in (-1, 0)$. Figure 1.1 illustrates the behavior of this value function.

Thus, we can produce a sequence of matrices X_ϵ each of which is ϵ -infeasible with respect to the infeasibility measure given by (2.5). 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_ϵ goes to infinity.

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

3 Value functions

The level-set method based on (1.1) is founded on the inverse-function relationship between the pair of “flipped” value functions

$$p(u) = \inf_{x \in \mathcal{X}} \{f(x) \mid g(x) \leq u\} \quad (3.1a)$$

$$v(\tau) = \inf_{x \in \mathcal{X}} \{g(x) \mid f(x) \leq \tau\}. \quad (3.1b)$$

Clearly, $\tau_p^* = p(0)$. Here we summarize the key aspects of the relationship between the value functions v and p , and their respective solutions. Aravkin et al. (2013) provide a complete description.

Let $\text{argmin } v(\tau)$ and $\text{argmin } p(u)$, respectively, denote the set of solutions to the optimization problem underlying the value functions v and p . Thus, for example, if the value $p(u)$ is finite,

$$\text{argmin } p(u) = \{x \in \mathcal{X} \mid f(x) = p(u), g(x) \leq 0\};$$

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

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

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

The following theorem establishes the relationships between the value functions p and v , and their respective solution sets. This result is reproduced from Aravkin et al. (2013, Theorem 2.1).

Theorem 3.1 (Value-function inverses) *For every $\tau \in \mathcal{S}$, the following statements hold:*

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

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

Example 3.1 (Failure of value-function inverse) The univariate problem

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

has the trivial solution $x^* = 0$ with optimal value $\tau_p^* = 0$. Note that the constraint is inactive at the solution, which violates the hypothesis of Theorem 3.1. Now consider the value functions

$$\begin{aligned} p(u) &= \inf \{ |x| : |x| - 1 \leq u \}, \\ v(\tau) &= \inf \{ |x| - 1 : |x| \leq \tau \}, \end{aligned}$$

which correspond, respectively, to a parameterization of the original problem, and to the level-set problem. The level-set value function v evaluates to

$$v(\tau) = \begin{cases} -1 & \text{if } \tau \geq \tau_p^* \\ +\infty & \text{if } \tau < \tau_p^*. \end{cases}$$

Because p is nonnegative over its domain, there is no value τ for which the inverse-function relationship shown by Theorem 3.1(a) holds.

Theorem 3.1 is symmetric, and holds if the roles of f and g , and p and v , are reversed. Aravkin et al. (2013) show that this result holds even if the underlying functions and sets 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,

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

Again consider Example 2.2, 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 3.1 (Value function epigraph) The *value function epigraph* of the optimal value function p in (3.1a) is defined by

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

This definition is similar to the regular definition for the epigraph of a function, given by

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

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

The result below follows immediately from the definition of the value function epigraph. It establishes that (1.2) holds if (Q_τ) has a solution that attains its optimal value (as opposed to relying on the infimal operator to achieve that value).

Proposition 3.1 For the value functions p and v ,

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

4 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 (1998). We adopt a geometric viewpoint that we have found helpful for understanding the connection between duality and the level-set method, and lays out the objects needed for the analysis in subsequent sections.

For this section only, consider the generic convex optimization problem

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad h(x),$$

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

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

Thus, the particular choice of F determines the perturbation function

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

which describes how the optimal value of h 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 τ_p^* , i.e., $p(0) = \tau_p^*$.

The convex conjugate of the function p is

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

defines the affine function $\mu \mapsto \langle \mu, u \rangle - p^*(\mu)$ that minorizes p and supports the epigraph of p ; see Figure 4.1. The biconjugate p^{**} provides a convex and closed function that is a global lower envelope for p , i.e., $p^{**}(u) \leq p(u)$ for all u . This last inequality is tight at a point u , i.e., $p^{**}(u) = p(u)$, if and only if p is lower-semicontinuous at u (Rockafellar 1970, Theorem 7.1). Because of the connection between lower semicontinuity and the closure of the epigraph, we say that p is *closed* at such points u .

As described by Rockafellar and Wets (1998, Lemma 11.38), the function p and its biconjugate p^{**} define dual pairs of optimization problems given by

$$p(0) = \inf_x F(x, 0) \quad \text{and} \quad p^{**}(0) = \sup_y -F^*(0, y), \quad (4.1)$$

which define the primal and dual optimal values

$$\tau_d^* := p^{**}(0) \leq p(0) =: \tau_p^*. \quad (4.2)$$

Strong duality holds when $\tau_p^* = \tau_d^*$, which indicates the closure of p at the origin. As we show in Section 5, the optimal dual value τ_d^* coincides with the value of the infimal value defined in (1.3).

The following well-known result establishes a constraint qualification for (P) that ensures strong duality holds. See Rockafellar and Wets (1998, Theorem 11.39) for a more comprehensive version of this result.

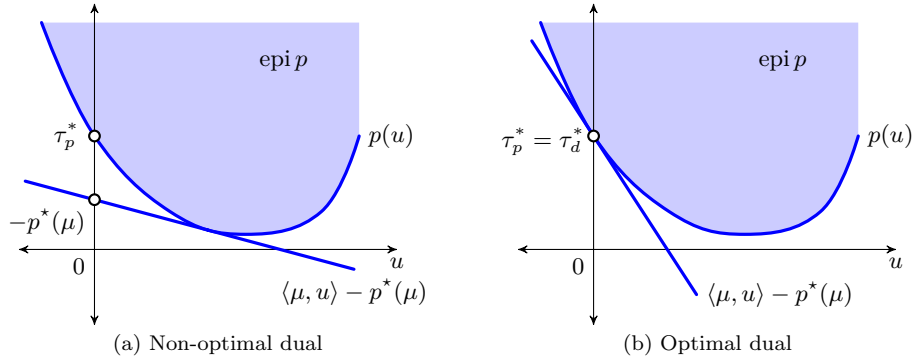


Fig. 4.1 The relationship between the primal perturbation value $p(u)$ and a single instance (with slope μ and intercept q_μ) of the uncountably many minorizing affine functions that define the dual problem. The panel on the left depicts a non-optimal supporting hyperplane that crosses the vertical axis at $-p^*(\mu) < \tau_p^*$; the panel on the right depicts an optimal supporting hyperplane that generates a slope μ and intercept $-p^*(\mu) = \tau_p^*$.

Theorem 4.1 (Weak and strong duality) Consider the primal-dual pair (4.1).

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

To establish the connection between the pair of value functions (3.1) for (P) and this duality framework, we observe that

$$p(u) = \inf_{x \in \mathcal{X}} \{ f(x) \mid g(x) \leq u \} = \inf_x F(x, u),$$

where

$$F(x, u) := f(x) + \delta_{\mathcal{X}}(x) + \delta_{\text{epi } g}(x, u), \quad (4.3)$$

and the indicator function $\delta_{\mathcal{C}}$ vanishes on the set \mathcal{C} and is $+\infty$ otherwise. The dual problem $p^{**}(0)$ defined in (4.1) is derived as follows:

$$\begin{aligned} p^{**}(0) &= \sup_{\lambda} -F^*(0, \lambda) \\ &= \sup_{\lambda} \inf_{x, u} \{ f(x) + \delta_{\mathcal{X}}(x) - \lambda u + \delta_{\text{epi } g}(x, u) \} \\ &= \sup_{\lambda \leq 0} \inf_{x \in \mathcal{X}} \{ f(x) - \lambda g(x) \}. \end{aligned} \quad (4.4)$$

We recognize this last expression as the familiar Lagrangian-dual for the optimization problem (P).

5 Duality of the value function root

We now provide a formal statement and proof our main result concerning problem (P) and the inequality shown in (1.3). In the latter part of this section we also provide a straight-forward extension of the main result that allows for multiple constraints, and not just a single constraint function, as specified by (P).

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

Theorem 5.1 (Duality of the value function root) *For problem (P) and the pair of value function v and p , defined by (3.1),*

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

where $\tau_d^* := p^{**}(0)$ is the optimal value of the Lagrange-dual problem (4.4).

Before giving the proof, below, we provide an intuitive argument for Theorem 5.1. Suppose that strong duality holds for (P). Hence, $\tau_p^* = p(0) = p^{**}(0) = \tau_d^*$, which means that the perturbation function p is closed at the origin. We sketch in the top row of Figure 5.1 example pairs of value functions p and v that exhibit this behavior. To understand this picture, first consider the value $\tau_1 < \tau_p^*$, shown in the top row. It is evident that $v(\tau_1)$ is positive, because otherwise there must exist a vector $x \in \mathcal{X}$ that is super-optimal and feasible, i.e.,

$$f(x) \leq \tau_1 < \tau_p^* \quad \text{and} \quad g(x) \leq 0,$$

which contradicts the definition of τ_p^* . It then follows that the value $u := v(\tau_1)$ yields $p(u) = \tau_1$. For $\tau_2 > \tau_p^*$, 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 τ_1 , where u_1 is the smallest root of $p(u) = \tau_1$.

Next consider the second row of Figure 5.1. 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_τ) to have zero value. But for $\tau_d^* < \tau_2 < \tau_p^*$, we see that $v(\tau_2) = 0$, because for any positive ϵ 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 ϵ there exists an ϵ -infeasible point that achieves that objective value.

Proof (Theorem 5.1)

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 τ_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 two conditions $g(x) \leq u$ and $f(x) \leq \tau$ are equivalent to the single condition $F(x, u) \leq \tau$, where F is defined by (4.3). We will therefore prove that

$$\forall \epsilon > 0, \exists x \in \mathcal{X} \text{ such that } F(x, u) \leq \tau, \quad u \leq \epsilon, \quad (5.1)$$

which is equivalent to the required condition $v(\tau) \leq 0$. It follows from the convexity of $\text{epi } p$ and from (4.2) 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

$$\begin{aligned} \liminf_{\epsilon \downarrow 0} \{ p(u) \mid |u| \leq \epsilon \} &\stackrel{(i)}{=} \liminf_{\epsilon \downarrow 0} \{ p^{**}(u) \mid |u| \leq \epsilon \} \\ &\stackrel{(ii)}{=} p^{**}(0) \stackrel{(iii)}{=} \tau_d^*, \end{aligned} \quad (5.2)$$

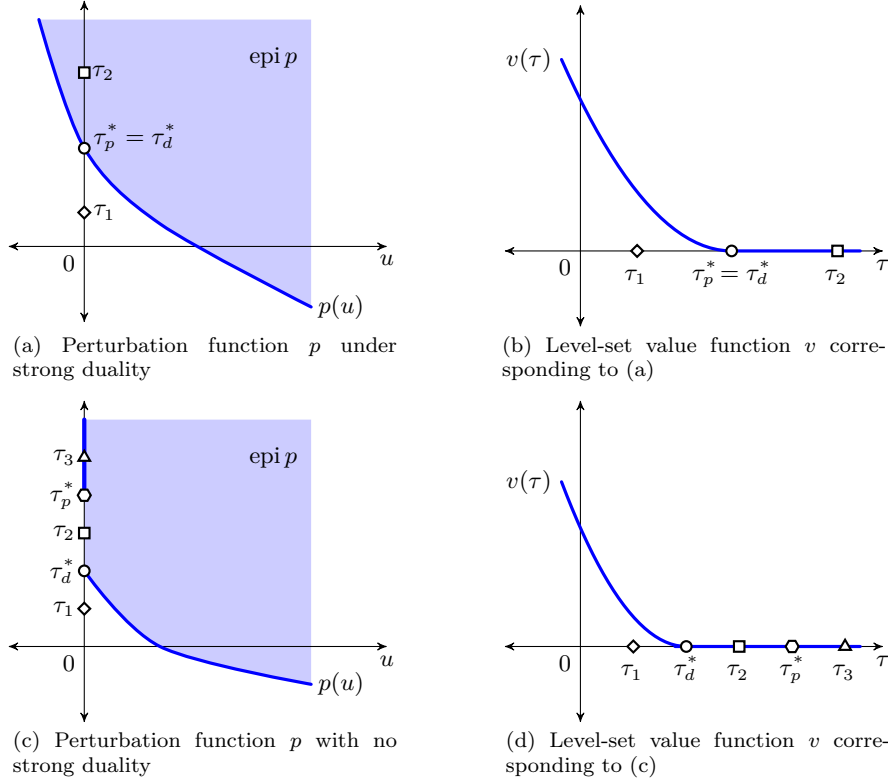


Fig. 5.1 The perturbation function $p(u)$ and corresponding level-set value function $v(\tau)$ for problems with strong duality (top row) and no strong duality (bottom row). Panel (c) illustrates the case when strong duality fails and the graph of p is open at the origin, which implies that $\tau_d^* < \tau_p^* \equiv p(0)$.

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 (4.2). 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 ϵ 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,

$$p(\hat{u}) < \tau_d^* + \mu. \quad (5.3)$$

Moreover, it follows from the definition of $p(\hat{u})$, cf. (3.1a), 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 (5.3), 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 (5.1), 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

$$\begin{aligned} 0 &< \inf_u \{ u \mid (u, \tau) \in \text{cl epi } p \} \\ &= \inf_u \{ u \mid (u, \tau) \in \text{epi } p \} \\ &= \inf_u \{ u \mid \exists x \in \mathcal{X} \text{ such that } F(x, u) \leq \tau \} = v(\tau), \end{aligned} \tag{5.4}$$

which completes the proof. \square

The proof of Theorem 5.1 reveals that the behavior exhibited by Examples 2.1 and 2.2 stems from the failure of strong duality with respect to perturbations in the linear constraints.

5.1 General perturbation framework

We now generalize Theorem 5.1 to include arbitrary perturbations to (P), and thus more general notions of duality. In this case we are interested in the value function pair

$$p(u) = \inf_{x \in \mathcal{X}} F(x, u), \tag{5.5a}$$

$$v(\tau) = \inf_{x \in \mathcal{X}} \{ \|u\| \mid F(x, u) \leq \tau \}, \tag{5.5b}$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{\infty\}$ is an arbitrary convex function with the property that $F(x, 0) = f(x)$ (cf. Section 4), and $\|\cdot\|$ is any norm. Because p is parameterized by an m -vector u and not just a scalar as previously considered, we must consider the norm of the perturbation. 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 5.1.

Example 5.1 (Multiple constraints) Consider the convex optimization problem

$$\underset{x}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad c(x) \leq 0, \quad Ax = b, \tag{5.6}$$

where $c = (c_i)_{i=1}^m$ 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 (5.6) 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 τ . \square

Theorem 5.2 For the functions p and v defined by (5.5),

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

The proof is almost identical to that of Theorem 5.1, except that we treat u as a vector, and replace u by $\|u\|$ in (5.1), (5.2), and (5.4).

Theorems 5.1 and 5.2 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 being solved. This means that the level-set method may provide a point 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.

Example 5.2 (Basis pursuit denoising (Chen et al. 1998, 2001)) The level-set method implemented in the SPGL1 software package solves the 1-norm regularized least-squares problem

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

for any value of $u \geq 0$, assuming that the problem remains feasible. (The case $u = 0$ 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 τ_k is chosen so that the corresponding solution x_k satisfies $\|Ax_k - b\|_2 \approx u$. Strong duality holds because the domains of the nonlinear functions (i.e., the 1- and 2-norms) cover the whole space. Thus, the level-set method succeeds on this problem. \square

6 Sufficient conditions for strong duality

The condition that $0 \in \text{dom } p$ may be interpreted as Slater's constraint qualification (Borwein and Lewis 2010, §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 5.1 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 6.1 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_τ) and its corresponding optimal-value function $v(\tau)$ given by (1.1). In either case (a) or (b), the feasible set

$$\{x \in \mathcal{X} \mid f(x) \leq \tau\}$$

of (1.1) is compact because either \mathcal{X} is compact or the level sets of f are compact. Therefore, (Q_τ) always attains its minimum for all $\tau \geq \inf \{f(x) \mid x \in \mathcal{X}\}$.

Suppose strong duality does not hold. Theorem 5.1 then confirms that there exists a parameter $\tau \in (\tau_d^*, \tau_p^*)$ such that $v(\tau) = 0$. However, because (Q_τ) always attains its minimum, there must exist a point $\hat{x} \in \mathcal{X}$ such that $f(\hat{x}) \leq \tau < \tau_p^*$ and $g(x) \leq 0$, which contradicts the fact that τ_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 6.1 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

$$\underset{x}{\text{minimize}} \langle c, x \rangle \quad \text{subject to} \quad \mathcal{A}x = b, \quad x \in \mathcal{K}, \quad (6.1)$$

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 (Ben-Tal and Nemirovski 2001). If c is 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, (6.1) 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 6.1 applies, and strong duality holds.

A concrete application of this model problem is the SDP relaxation of the celebrated phase-retrieval problem (Candès et al. 2013; Waldspurger et al. 2015)

$$\underset{X}{\text{minimize}} \quad \text{tr}(X) \quad \text{subject to} \quad \mathcal{A}X = b, \quad X \succeq 0, \quad (6.2)$$

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, Candès et al. (2013) prove that with high probability, the feasible set of (6.1) 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 (Boyd and Vandenberghe 2004, Example 2.24), clearly $c \in \text{int } \mathcal{K}$, and by the discussion above, we can use Proposition 6.1 to establish that strong duality holds (6.2).

A consequence of Proposition 6.1 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 μ 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 6.1 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 2.1 and 2.2, 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.

Acknowledgements 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_r) 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.

References

- Aravkin AY, Burke J, Friedlander MP (2013) Variational properties of value functions. *SIAM J Optim* 23(3):1689–1717, DOI 10.1137/120899157, <http://epubs.siam.org/doi/pdf/10.1137/120899157>
- Aravkin AY, Burke JV, Drusvyatskiy D, Friedlander MP, Roy S (2018) Level-set methods for convex optimization. *Math Program, Ser B* 174(1-2):359–390, DOI 10.1007/s10107-018-1351-8
- Ben-Tal A, Nemirovski A (2001) *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, MPS/SIAM Series on Optimization, vol 2. Society of Industrial and Applied Mathematics, Philadelphia
- van den Berg E, Friedlander MP (2007) SPGL1: A solver for large-scale sparse reconstruction. [Http://www.cs.ubc.ca/labs/scl/spgl1](http://www.cs.ubc.ca/labs/scl/spgl1)
- van den Berg E, Friedlander MP (2008a) Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing* 31(2):890–912, DOI 10.1137/080714488, URL <http://link.aip.org/link/?SCE/31/890>
- van den Berg E, Friedlander MP (2008b) Probing the Pareto frontier for basis pursuit solutions. *SIAM J Sci Comput* 31(2):890–912, DOI 10.1137/080714488
- van den Berg E, Friedlander MP (2011) Sparse optimization with least-squares constraints. *SIAM J Optim* 21(4):1201–1229, DOI 10.1137/100785028
- van den Berg E, Friedlander MP (2013) SPGL1: A solver for large-scale sparse reconstruction. URL <https://www.cs.ubc.ca/~mpf/spgl1/>
- Borwein J, Lewis AS (2010) *Convex analysis and nonlinear optimization: theory and examples*. Springer Science & Business Media
- Boyd S, Vandenberghe L (2004) *Convex optimization*. Cambridge University Press, Cambridge, DOI 10.1017/CBO9780511804441
- Candès EJ (2006) Compressive sampling. In: *Proceedings of the International Congress of Mathematicians*
- Candès EJ, Strohmer T, Voroninski V (2013) PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming. *Comm Pure Appl Math* 66(8):1241–1274, DOI 10.1002/cpa.21432, URL <http://dx.doi.org/10.1002/cpa.21432>
- Chen SS, Donoho DL, Saunders MA (1998) Atomic decomposition by basis pursuit. *SIAM J Sci Comput* 20(1):33–61
- Chen SS, Donoho DL, Saunders MA (2001) Atomic decomposition by basis pursuit. *SIAM Rev* 43(1):129–159
- Conn AR, Gould NIM, Toint PhL (2000) *Trust-Region Methods*. MPS-SIAM Series on Optimization, Society of Industrial and Applied Mathematics, Philadelphia
- Donoho DL (2006) For most large underdetermined systems of linear equations, the minimal l_1 -norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics* 59
- Higham NJ (2002) *Accuracy and stability of numerical algorithms*, vol 80. Society for Industrial and Applied Mathematics
- Lemaréchal C, Nemirovskii A, Nesterov Y (1995) New variants of bundle methods. *Math Program* 69:111–147
- Markowitz HM (1987) *Mean-Variance Analysis in Portfolio Choice and Capital Markets*. Frank J. Fabozzi Associates, New Hope, Pennsylvania
- Marquardt D (1963) An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* 11:431–441
- Morrison DD (1960) *Methods for nonlinear least squares problems and convergence proofs*. In: Lorell J, Yagi F (eds) *Proceedings of the Seminar on Tracking Programs and Orbit Determination*, Jet Propulsion Laboratory, Pasadena, USA, pp 1–9
- Rockafellar RT (1970) *Convex Analysis*. Princeton University Press, Princeton
- Rockafellar RT, Wets RJB (1998) *Variational Analysis*, vol 317. Springer, 3rd printing

-
- WalDSPURGER I, d'ASPREMONT A, MALLAT S (2015) Phase recovery, maxcut and complex semidefinite programming. *Math Program* 149(1-2):47-81
- 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