J Optim Theory Appl (2007) 135: 429–443
DOI 10.1007/s10957-007-9281-2

Approximate Duality

C. Scovel · D. Hush · I. Steinwart

Published online: 31 October 2007

Abstract We extend the Lagrangian duality theory for convex optimization problems to incorporate approximate solutions. In particular, we generalize well-known relationships between minimizers of a convex optimization problem, maximizers of its Lagrangian dual, saddle points of the Lagrangian, Kuhn–Tucker vectors, and Kuhn–Tucker conditions to incorporate approximate versions. As an application, we show how the theory can be used for convex quadratic programming and then apply the results to support vector machines from learning theory.

Keywords Lagrangian duality · Approximations · Saddle points · Kuhn–Tucker conditions · Support vector machines

1 Introduction

Duality theory provides a rich framework for the development of solution methods for convex optimization problems. Key components of this theory include a primal optimization problem, a Lagrangian defined on the space of primal and dual variables, a dual optimization problem defined on the space of dual variables (i.e. Lagrange multipliers), and Kuhn–Tucker conditions defined on the space of primal and dual variables. Solution methods include *primal methods* which work directly in the space of primal variables; *dual methods* which solve a dual optimization problem and then construct a primal solution from a dual solution; and *primal–dual methods* which solve for the primal and dual variables simultaneously. Dual and primal–dual methods are designed to find points that either satisfy the Kuhn–Tucker conditions or

Communicated by M. Fukushima.

C. Scovel (✉) · D. Hush · I. Steinwart
Modeling, Algorithms, and Informatics Group, Los Alamos National Laboratory,
Los Alamos, NM, USA
e-mail: jcs@lanl.gov

correspond to a saddle of the Lagrangian. Although much of existing duality theory assumes that these methods produce exact solutions, in practice it is more common to produce approximate solutions. Indeed, many practical algorithms do not converge to an exact solution in a finite number of iterations and therefore produce approximate solutions. Moreover, a simpler algorithm which produces an approximate solution is often utilized instead of a more complex algorithm which, in principle, produces an exact solution. More generally, practical algorithms operate with finite precision arithmetic and therefore the accuracy of their solutions is limited. Thus there is a need for a duality theory for approximate solutions. In particular when a dual method is used there is a need to know how best to construct an approximate primal solution from an approximate dual solution, and when a primal–dual method is used there is a need to relate the accuracy with which the Kuhn–Tucker conditions or the saddle point problem are solved to the accuracy of the corresponding approximate primal solution. Although there has been some work in this direction (see e.g. [1] for an approximate Kuhn–Tucker Theorem, [2] for a variational principle for approximate minimizers, and [3] and the references therein) we feel there is a need for a general treatment of duality theory for approximate solutions. Consequently, the main part of this paper is written as an approximate version of Chap. 28 titled Ordinary Convex Programs and Lagrange Multipliers in [4] by Rockafellar. As an application, in Sect. 3 we show how this framework can be used for quadratic convex programming problems and apply the results to support vector machines from learning theory. Although we extend to optimization in Hausdorff locally convex topological vector spaces we did not extend to an infinite number of constraints. We suspect that such an extension is straightforward along the lines of [5].

2 Main Results

We will state and prove an approximate version of the Kuhn–Tucker theorem and other related results as presented in [4, Chap. 28]. However, first we need to define terminology. Let X be a Hausdorff locally convex topological vector space, and consider a nonempty closed convex set $C \subset X$, a set of lower semi-continuous convex functions $f_i : X \to \mathbb{R}$, $i = 0, \dots, r$, and a set of continuous affine functions $f_i : X \to \mathbb{R}$, $i = r + 1, \dots, m$. Throughout we make the following assumption:

There exists a point $x \in C$ where all of the functions $f_1, \dots, f_r$ are continuous. (1)

We define the convex programming problem (P) as follows:

$$\begin{aligned} \text{(P)} \quad \min \quad & f_0(x), \\ \text{s.t.} \quad & x \in C, \\ & f_1(x) \le 0, \dots, f_r(x) \le 0, \\ & f_{r+1}(x) = 0, \dots, f_m(x) = 0. \end{aligned} \tag{2}$$

Let us define

$$C_i = \{x \in X : f_i(x) \le 0\}, \quad i = 1, \dots, r,$$
$$C_i = \{x \in X : f_i(x) = 0\}, \quad i = r + 1, \dots, m,$$

and

$$C_0 = C \cap C_1 \cap \cdots \cap C_m.$$

Throughout, we assume that C_0 is nonempty. We define the optimal value

$$\nu := \inf_{x \in C_0} f_0(x)$$

for the convex programming problem (P). For $\epsilon \geq 0$, we define the set of ϵ-minimizers of (P) as

$$\mathcal{O}_\epsilon(\mathrm{P}) = \left\{ x \in C_0 : f_0(x) \leq \inf_{x' \in C_0} f_0(x') + \epsilon \right\}. \tag{3}$$

Let $E_r = \{\lambda \in \mathbb{R}^m : \lambda_i \geq 0, i = 1, \ldots, r\}$ and define the shorthand

$$h_\lambda(x) := f_0(x) + \lambda_1 f_1(x) + \cdots + \lambda_m f_m(x).$$

We define the set of ϵ-Kuhn–Tucker vectors for (P) by

$$\mathrm{KT}_\epsilon := \left\{ \lambda \in E_r : \inf_C h_\lambda \geq \nu - \epsilon > -\infty \right\}. \tag{4}$$

Let X^* denote the topological dual to X and consider the ϵ-subdifferential of a convex function h at x defined by

$$\partial_\epsilon h(x) := \{x^* \in X^* : h(y) \geq h(x) + x^*(y - x) - \epsilon, \ \forall y \in X\},$$

where $\partial_\epsilon h(x) := \emptyset$ when $h(x)$ is not finite. We note the following important facts concerning the approximate subdifferential of sums of functions: If h_1 and h_2 are lower semicontinuous proper convex functions, it is well known that

$$\partial_\epsilon (h_1 + h_2) \supset \bigcup_{\epsilon_1, \epsilon_2 \geq 0, \ \epsilon_1 + \epsilon_2 = \epsilon} \{\partial_{\epsilon_1} h_1 + \partial_{\epsilon_2} h_2\}. \tag{5}$$

However, if in addition there exists a point $x \in X$ where both h_1 and h_2 are finite and one of them is continuous, then Theorem 2.8.3 of [6] shows that

$$\partial_\epsilon (h_1 + h_2) = \bigcup_{\epsilon_1, \epsilon_2 \geq 0, \ \epsilon_1 + \epsilon_2 = \epsilon} \{\partial_{\epsilon_1} h_1 + \partial_{\epsilon_2} h_2\}. \tag{6}$$

Let $\partial := \partial_0$ denote the usual subdifferential. Also, let dh denote the differential of a function h. For affine continuous functions h, we know that we have

$$h(x) = x^*(x) + h(0),$$

for some $x^* \in X^*$, and that

$$\partial h(x) = dh(x) = x^*, \quad \forall x \in X.$$

For a subset $S \subset X$, we let $\delta_S(x)$ denote the indicator function of the set S, i.e.

$$\delta_S(x) := \begin{cases} 0, & x \in S, \\ \infty, & x \notin S. \end{cases}$$

We define the set of points which satisfy the ϵ-Kuhn–Tucker conditions KTC_ϵ to be the set of $(x, \lambda) \in C_0 \times E_r$ which satisfy

$$\left\{ \begin{array}{l} -\epsilon \le \sum_{i=1}^m \lambda_i f_i(x), \\ \exists \epsilon_C \ge 0,\ \epsilon_i \ge 0, i = 0, \dots, r, \text{ such that } \sum_{i=0}^r \epsilon_i + \epsilon_C \le \epsilon \text{ and} \\ 0 \in \partial_{\epsilon_0} f_0(x) + \sum_{i=1}^r \partial_{\epsilon_i}(\lambda_i f_i)(x) + \sum_{i=r+1}^m \lambda_i d f_i(x) + \partial_{\epsilon_C} \delta_C(x) \end{array} \right\} \tag{7}$$

We define the Lagrangian on $X \times \mathbb{R}^m$ to be

$$L(x, \lambda) := \begin{cases} h_\lambda(x), & \lambda \in E_r, x \in C, \\ -\infty, & \lambda \notin E_r, x \in C, \\ \infty, & x \notin C. \end{cases} \tag{8}$$

Since C is nonempty, it follows that

$$\inf_{x \in X} L(x, \lambda) = \begin{cases} \inf_C h_\lambda, & \lambda \in E_r, \\ -\infty, & \lambda \notin E_r. \end{cases} \tag{9}$$

Moreover, we note the useful identity

$$\sup_{\lambda \in \mathbb{R}^m} L(x, \lambda) = f_0(x) + \delta_{C_0}(x), \tag{10}$$

from which it follows that

$$\nu = \inf_{x \in X} \sup_{\lambda \in \mathbb{R}^m} L(x, \lambda). \tag{11}$$

We define the set of ϵ-saddle points of L to be

$$\mathrm{Sad}_\epsilon := \{(x, \lambda) \in X \times \mathbb{R}^m\colon\ L(x, \lambda') - \epsilon \le L(x, \lambda) \le L(x', \lambda) + \epsilon,\ \forall (x', \lambda') \in X \times \mathbb{R}^m\}. \tag{12}$$

We now prove an approximate version of Theorem 28.3, [4] which provides the basic connections between the sets $\mathcal{O}_\epsilon(\mathrm{P})$, KT_ϵ, KTC_ϵ, Sad_ϵ.

Theorem 2.1 *For all* $0 \le \epsilon < \infty$, *we have*:

(i) $\mathrm{Sad}_\epsilon \subset \mathcal{O}_{2\epsilon}(\mathrm{P}) \times \mathrm{KT}_{2\epsilon}$ *and* $\mathcal{O}_\epsilon(\mathrm{P}) \times \mathrm{KT}_\epsilon \subset \mathrm{Sad}_{2\epsilon}$.
(ii) $\mathrm{Sad}_\epsilon \subset \mathrm{KTC}_{2\epsilon}$ *and* $\mathrm{KTC}_\epsilon \subset \mathrm{Sad}_{2\epsilon}$.
(iii) $\mathcal{O}_\epsilon(\mathrm{P}) \times \mathrm{KT}_\epsilon \subset \mathrm{KTC}_{2\epsilon}$ *and* $\mathrm{KTC}_\epsilon \subset \mathcal{O}_{2\epsilon}(\mathrm{P}) \times \mathrm{KT}_{2\epsilon}$.

Remark 2.1 We note that Theorem 2.1 establishes a connection between approximate primal-dual solutions and approximate primal solutions. That is, if the primal-dual method produces a $(x, \lambda) \in \mathrm{Sad}_\epsilon$ or $(x, \lambda) \in \mathrm{KTC}_\epsilon$, then $x \in \mathcal{O}_{2\epsilon}(\mathrm{P})$.

Proof We proceed by developing an intermediate set I_ϵ and proving the intermediate assertions $\mathrm{I}_\epsilon \subset \mathrm{Sad}_\epsilon \subset \mathrm{I}_{2\epsilon}$, $\mathrm{I}_\epsilon \subset \mathcal{O}_\epsilon(\mathrm{P}) \times \mathrm{KT}_\epsilon \subset \mathrm{I}_{2\epsilon}$, and $\mathrm{I}_\epsilon \subset \mathrm{KTC}_\epsilon \subset \mathrm{I}_{2\epsilon}$. The theorem then follows directly.

Define the intermediate set

$$\mathrm{I}_\epsilon := \left\{(x,\lambda) \in C_0 \times E_r : f_0(x) \le \inf_C h_\lambda + \epsilon\right\} \tag{13}$$

and suppose that $(x,\lambda) \in \mathrm{I}_\epsilon$. Then $x \in C_0$ and the identity (10) implies that $L(x,\lambda') \le f_0(x)$ for all $\lambda' \in \mathbb{R}^m$. Consequently, the definition (13) of I_ϵ yields

$$L(x,\lambda') - \epsilon \le f_0(x) - \epsilon \le \inf_C h_\lambda.$$

In addition, since $\lambda \in E_r$, the identity (9) implies that $\inf_C h_\lambda \le L(x,\lambda)$, so that we obtain

$$L(x,\lambda') - \epsilon \le L(x,\lambda).$$

On the other hand, since $\lambda \in E_r$, the identity (9) also implies that

$$\inf_C h_\lambda \le L(x',\lambda), \quad \forall x' \in X,$$

and since $x \in C_0$, the definition (13) of I_ϵ yields

$$L(x,\lambda) \le f_0(x) \le \inf_C h_\lambda + \epsilon \le L(x',\lambda) + \epsilon.$$

Therefore, $\mathrm{I}_\epsilon \subset \mathrm{Sad}_\epsilon$. Moreover, $\lambda \in E_r$ implies that, for all $x' \in C_0$, we have $h_\lambda(x') \le f_0(x')$ so that

$$\inf_C h_\lambda \le \inf_{C_0} h_\lambda \le \inf_{C_0} f_0 = \nu \le f_0(x) \le \inf_C h_\lambda + \epsilon,$$

from which we conclude that

$$-\infty < f_0(x) \le \nu + \epsilon;$$

therefore,

$$\inf_C h_\lambda \ge \nu - \epsilon > -\infty.$$

Consequently, $\mathrm{I}_\epsilon \subset \mathcal{O}_\epsilon(\mathrm{P}) \times \mathrm{KT}_\epsilon$. In addition,

$$f_0(x) \le \inf_C h_\lambda + \epsilon \le h_\lambda(x) + \epsilon = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \epsilon \le f_0(x) + \epsilon,$$

so that

$$-\epsilon \le \sum_{i=1}^m \lambda_i f_i(x) \quad \text{and} \quad h_\lambda(x) \le \inf_C h_\lambda + \epsilon.$$

The latter inequality is equivalent to

$$0 \in \partial_\epsilon (h_\lambda + \delta_C)(x). \tag{14}$$

Using the assumption (1) and the sum formula (6), we obtain the existence of $\epsilon_i \geq 0, i = \overline{0, m}, \epsilon_C \geq 0, \sum_{i=0}^m \epsilon_i + \epsilon_C = \epsilon$ such that

$$0 \in \partial_{\epsilon_0} f_0(x) + \partial_{\epsilon_1}(\lambda_1 f_1)(x) + \cdots + \partial_{\epsilon_m}(\lambda_m f_m)(x) + \partial_{\epsilon_C} \delta_C(x).$$

For the affine functions $f_i, i = \overline{r+1, m}$, $\partial_{\epsilon_i}(\lambda_i f_i) = \partial(\lambda_i f_i) = \lambda_i df_i$, so that we conclude that $\mathrm{I}_\epsilon \subset \mathrm{KTC}_\epsilon$.

Now suppose that $(x, \lambda) \in \mathrm{Sad}_\epsilon$. The inequality (12) implies that

$$\sup_{\lambda' \in \mathbb{R}^m} L(x, \lambda') - \epsilon \leq L(x, \lambda) \leq \epsilon + \inf_{x' \in X} L(x', \lambda);$$

therefore, the identities (10) and (9) and the fact that $f_0(x) \in \mathbb{R}$ imply that

$$-\infty < f_0(x) + \delta_{C_0}(x) - \epsilon \leq \epsilon + \begin{cases} \inf_C h_\lambda, & \lambda \in E_r, \\ -\infty, & \lambda \notin E_r, \end{cases} < \infty.$$

Therefore, we conclude that $x \in C_0$, $\lambda \in E_r$, and

$$f_0(x) \leq \inf_C h_\lambda + 2\epsilon.$$

That is, $\mathrm{Sad}_\epsilon \subset \mathrm{I}_{2\epsilon}$ and we have established $\mathrm{I}_\epsilon \subset \mathrm{Sad}_\epsilon \subset \mathrm{I}_{2\epsilon}$.

Now, suppose that $(x, \lambda) \in \mathcal{O}_\epsilon(\mathrm{P}) \times \mathrm{KT}_\epsilon$. Then, $x \in C_0$, $\lambda \in E_r$, and

$$f_0(x) \leq \nu + \epsilon \quad \text{and} \quad \inf_C h_\lambda \geq \nu - \epsilon;$$

therefore,

$$f_0(x) \leq \inf_C h_\lambda + 2\epsilon$$

and so we conclude that $\mathcal{O}_\epsilon(\mathrm{P}) \times \mathrm{KT}_\epsilon \subset \mathrm{I}_{2\epsilon}$, thus establishing that $\mathrm{I}_\epsilon \subset \mathcal{O}_\epsilon(\mathrm{P}) \times \mathrm{KT}_\epsilon \subset \mathrm{I}_{2\epsilon}$.

Now suppose that $(x, \lambda) \in \mathrm{KTC}_\epsilon$. It is well known that since C is closed and convex δ_C is a lower semicontinuous proper convex function. Moreover, the relation (5) applied to the subdifferential relation of KTC_ϵ implies that $0 \in \partial_\epsilon(h_\lambda + \delta_C)(x)$ which in turn implies that $h_\lambda(x) \leq \inf_C h_\lambda + \epsilon$. Moreover, $-\epsilon \leq \sum_{i=1}^m \lambda_i f_i(x)$ implies that

$$h_\lambda(x) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \geq f_0(x) - \epsilon,$$

so that we obtain

$$f_0(x) \leq \inf_C h_\lambda + 2\epsilon$$

and conclude that $\mathrm{KTC}_\epsilon \subset \mathrm{I}_{2\epsilon}$, thus establishing that $\mathrm{I}_\epsilon \subset \mathrm{KTC}_\epsilon \subset \mathrm{I}_{2\epsilon}$. □

We obtain as a corollary the following approximate version of the Kuhn–Tucker Theorem. We note that Strodiot et al. [1] prove a stronger version of (ii) in $\mathbb{R}^n$ and that Yokoyama and Shiraishi [7] and have thoroughly explored the possibility of the removal of the Slater constraint qualifications in the results of [1].

Corollary 2.1 *For all* $0 \le \epsilon < \infty$, *we have*:

(i) *If* KT_ϵ *is not empty, then for all* $x \in \mathcal{O}_\epsilon(\mathrm{P})$ *there exists a* λ *such that* $(x, \lambda) \in \mathrm{Sad}_{2\epsilon}$. *Conversely, for a fixed* x, *if there exists a* λ *such that* $(x, \lambda) \in \mathrm{Sad}_\epsilon$ *then* $x \in \mathcal{O}_{2\epsilon}(\mathrm{P})$.

(ii) *If* KT_ϵ *is not empty, then for all* $x \in \mathcal{O}_\epsilon(\mathrm{P})$ *there exists a* λ *such that* $(x, \lambda) \in \mathrm{KTC}_{2\epsilon}$. *Conversely, for a fixed* x, *if there exists a* λ *such that* $(x, \lambda) \in \mathrm{KTC}_\epsilon$ *then* $x \in \mathcal{O}_{2\epsilon}(\mathrm{P})$.

We now prove an approximate version of Theorem 28.4, [4] which shows how the optimal value ν relates to the value of the Lagrangian at approximate minimizers and approximate Kuhn–Tucker vectors.

Theorem 2.2 *For all* $0 \le \epsilon < \infty$, *we have*:

(i) $(x, \lambda) \in \mathcal{O}_\epsilon(\mathrm{P}) \times \mathrm{KT}_\epsilon$ *implies that*

$$\nu - \epsilon \le L(x, \lambda) \le \nu + \epsilon.$$

(ii) $\lambda \in \mathrm{KT}_\epsilon$ *if and only if* $\inf_{x \in X} L(x, \lambda) \ge \nu - \epsilon > -\infty$ *and in this case*

$$\inf_X \sup_{\mathbb{R}^m} L - \sup_{\mathbb{R}^m} \inf_X L \le \epsilon.$$

Proof Suppose that $(x, \lambda) \in \mathcal{O}_\epsilon(\mathrm{P}) \times \mathrm{KT}_\epsilon$. Then,

$$L(x, \lambda) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \le f_0(x) \le \nu + \epsilon$$

and

$$L(x, \lambda) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \ge \nu - \epsilon$$

completes the proof of the assertion (i). Now, suppose that $\lambda \in \mathrm{KT}_\epsilon$. The identity (9) implies

$$\inf_{x \in X} L(x, \lambda) = \inf_C h_\lambda \ge \nu - \epsilon > -\infty$$

and the identity (11) then implies that

$$\sup_{\mathbb{R}^m} \inf_X L \ge \nu - \epsilon = \inf_X \sup_{\mathbb{R}^m} L - \epsilon,$$

proving the forward part of assertion (ii).

Conversely,

$$-\infty < \nu - \epsilon \le \inf_{x \in X} L(x, \lambda)$$

and the identity (9) implies that $\lambda \in E_r$ so that

$$\inf_C h_\lambda = \inf_{x \in X} L(x, \lambda) \ge \nu - \epsilon > -\infty.$$

Consequently $\lambda \in \mathrm{KT}_\epsilon$ and assertion (ii) is proved. □

Let us now consider the Lagrange dual problem

$$\begin{array}{lll} \text{(D)} & \max & g(\lambda), \\ & \text{s.t.} & \lambda \in \mathbb{R}^m, \end{array} \tag{15}$$

with criterion function

$$g(\lambda) := \inf_{x \in X} L(x, \lambda). \tag{16}$$

The dual optimal value is defined as

$$\nu^* := \sup_{\lambda \in \mathbb{R}^m} g(\lambda)$$

and the approximate maximizers for the dual problem are defined by

$$\mathcal{O}_\epsilon(\mathrm{D}) := \left\{ \lambda \in \mathbb{R}^m \colon\ g(\lambda) \ge \sup_{\lambda' \in \mathbb{R}^m} g(\lambda') - \epsilon \right\}. \tag{17}$$

We note that the minmax inequality implies that

$$\nu^* \le \nu.$$

We can now state the following important corollary to Theorem 2.2.

Corollary 2.2 *Consider the Lagrangian dual maximization problem* (15) *with concave criterion function defined in* (16). *Then,* $\mathcal{O}_\epsilon(\mathrm{D}) \ne \emptyset$ *for all* $0 < \epsilon < \infty$ *and, for all* $0 \le \epsilon < \infty$, *we have*:

(i) $\mathrm{KT}_\epsilon \subset \mathcal{O}_\epsilon(\mathrm{D})$.
(ii) $\mathrm{KT}_\epsilon \ne \emptyset$ *implies that* $\nu > -\infty$ *and* $\inf_X \sup_{\mathbb{R}^m} L - \sup_{\mathbb{R}^m} \inf_X L \le \epsilon$.
(iii) $\nu > -\infty$ *and* $\inf_X \sup_{\mathbb{R}^m} L - \sup_{\mathbb{R}^m} \inf_X L \le \epsilon_1$ *for some* $0 \le \epsilon_1 < \infty$ *implies that* $\mathcal{O}_\epsilon(\mathrm{D}) \subset \mathrm{KT}_{\epsilon+\epsilon_1}$.

Remark 2.2 We note that the assumptions $\nu > -\infty$ and $\inf_X \sup_{\mathbb{R}^m} L - \sup_{\mathbb{R}^m} \inf_X L \le \epsilon$ are equivalent to the assumptions $\nu^* > -\infty$ and $\inf_X \sup_{\mathbb{R}^m} L - \sup_{\mathbb{R}^m} \inf_X L \le \epsilon$.

Remark 2.3 The duality gap $\inf_X \sup_{\mathbb{R}^m} L - \sup_{\mathbb{R}^m} \inf_X L$ can often be proven to be zero (see e.g. Sect. 2.3.3, [8]).

Proof of Corollary 2.2 It follows from the min-max inequality that $\sup_{\mathbb{R}^m} g = \sup_{\mathbb{R}^m} \inf_X L \leq \inf_X \sup_{\mathbb{R}^m} L = \nu$ and since C_0 is nonempty it follows that the right-hand side is less than ∞. Consequently, $\mathcal{O}_\epsilon(\mathrm{D}) \neq \emptyset$ for all $0 < \epsilon < \infty$. Now let $\lambda \in \mathrm{KT}_\epsilon$ for $0 \leq \epsilon < \infty$. Then, Theorem 2.2 and the min-max inequality imply that

$$g(\lambda) = \inf_{x \in X} L(x, \lambda) \geq \nu - \epsilon = \inf_X \sup_{\mathbb{R}^m} L - \epsilon \geq \sup_{\mathbb{R}^m} \inf_X L - \epsilon = \sup_{\mathbb{R}^m} g - \epsilon$$

proving assertion (i). Assertion (ii) follows directly from assertion (ii) of Theorem 2.2. For assertion (iii), consider $\lambda \in \mathcal{O}_\epsilon(\mathrm{D})$ for $0 \leq \epsilon < \infty$. The assumptions $\nu > -\infty$ and $\inf_X \sup_{\mathbb{R}^m} L - \sup_{\mathbb{R}^m} \inf_X L \leq \epsilon_1$ implies that $\sup_{\mathbb{R}^m} \inf_X L > -\infty$. Therefore, since $\lambda \in \mathcal{O}_\epsilon(\mathrm{D})$, we obtain

$$\inf_{x \in X} L(x, \lambda) = g(\lambda) \geq \sup_{\mathbb{R}^m} \inf_X L - \epsilon > -\infty.$$

Consequently, we conclude from the identity (9) that $\lambda \in E_r$ and

$$\inf_C h_\lambda = g(\lambda) \geq \sup_{\mathbb{R}^m} \inf_X L - \epsilon \geq \inf_X \sup_{\mathbb{R}^m} L - \epsilon_1 - \epsilon = \nu - \epsilon_1 - \epsilon$$

and the proof is finished. □

The following corollary is important to produce approximate primal solutions from approximate dual solutions.

Corollary 2.3 *Suppose that* $0 \leq \epsilon < \infty$. *Then, the following statements hold*:

(i) *Suppose that we have* $\lambda \in \mathcal{O}_\epsilon(\mathrm{D})$. *Also suppose that* $\nu > -\infty$ *and that* $\inf_X \sup_{\mathbb{R}^m} L - \sup_{\mathbb{R}^m} \inf_X L \leq \epsilon_1$ *for some* $0 \leq \epsilon_1 < \infty$. *When* $\epsilon = 0$ *suppose further that* $\mathcal{O}_0(\mathrm{P})$ *is nonempty. Then the set of* x *for which* $(x, \lambda) \in \mathrm{KTC}_\tau$ *is nonempty for all* $\tau \geq 2\epsilon + 2\epsilon_1$.
(ii) *Given a fixed* $\lambda \in \mathbb{R}^m$, *for every* x *such that* $(x, \lambda) \in \mathrm{KTC}_\epsilon$, *we have* $x \in \mathcal{O}_{2\epsilon}(\mathrm{P})$.

Proof For the first claim, Corollary 2.2 implies that $\lambda \in \mathrm{KT}_{\epsilon+\epsilon_1}$. It follows from $\nu > -\infty$ when $\epsilon > 0$ and from the monotonicity of $\mathcal{O}_\tau(\mathrm{P})$ in τ when $\epsilon = 0$ that $\mathcal{O}_{\epsilon+\epsilon_1}(\mathrm{P})$ is nonempty. From Theorem 2.1(iii) we can conclude that the set of x such that $(x, \lambda) \in \mathrm{KTC}_{2\epsilon+2\epsilon_1}$ is not empty. The monotonicity of KTC_τ in τ proves the first claim. For the second, observe that for any $(x, \lambda) \in \mathrm{KTC}_\epsilon$ it follows also from Theorem 2.1(iii) that $x \in \mathcal{O}_{2\epsilon}(\mathrm{P})$. □

3 Applications

Corollary 2.3 provides a mechanism for generating approximate solutions to the primal problem (P) from approximate solutions to its dual (D) in the following way: Suppose the duality gap is zero and $\nu > -\infty$. Then given an ϵ-maximizer λ^* of the dual problem, Corollary 2.3(i) states that the set of solutions of the approximate

Kuhn–Tucker equations $KTC_{2\epsilon}$ of the form (x, λ^*) is nonempty. Find such a solution $\hat{x}$. Then Corollary 2.3(ii) shows that $\hat{x}$ is a 4ϵ-minimizer of the primal problem (P). Since this work involves the study of approximate solutions we note here that Corollary 2.3 also provides a method for utilizing *approximate* solutions to the approximate Kuhn–Tucker equations in the following way: Choose a solution method such that the approximate solution $\hat{x}$ to $KTC_{2\epsilon}$ is an exact solutions to $KTC_{2\epsilon'}$ for some $\epsilon' \geq \epsilon$. Then $\hat{x}$ will be a $4\epsilon'$-minimizer of the primal problem (P). As an illustration we now show how this observation aids in the design of efficient algorithms for convex quadratic programming problems. In particular, we will show, under a modulus condition on the constraint matrix, that an approximate solution to the primal can be obtained by solving the dual approximately and then solving a linear programming problem approximately.

Before we state the main result of this section we need some preparations. Consider the convex quadratic programming problem (P) with linear constraints:

$$\min_{\substack{x\in\mathcal{H}\\ Ax\leq b}} \frac{1}{2}\langle Qx, x\rangle + \langle q, x\rangle, \tag{18}$$

where $\mathcal{H}$ is a Hilbert space, $q \in \mathcal{H}$, $b \in \mathbb{R}^N$, $A : \mathcal{H} \to \mathbb{R}^N$ is linear, and $Q : \mathcal{H} \to \mathcal{H}$ is a continuous nonnegative self adjoint operator such that $Q : \mathrm{Ker}(Q)^\perp \to \mathrm{Ker}(Q)^\perp$ has a continuous inverse. Let Q^{-1} denote the continuous inverse of $Q : \mathrm{Ker}(Q)^\perp \to \mathrm{Ker}(Q)^\perp$. Then, the Lagrangian dual programming problem (D) is

$$\max_{\substack{\lambda\geq 0\\ A^t\lambda+q\in\mathrm{Ker}(Q)^\perp}} -\frac{1}{2}\langle Q^{-1}(A^t\lambda + q), (A^t\lambda + q)\rangle - \langle\lambda, b\rangle. \tag{19}$$

The following definition describes a useful property of the constraint operator A.

Definition 3.1 Let X be a normed linear space and Y be a partially ordered normed linear space. A linear operator $A : X \to Y$ has *modulus* $\kappa(A)$ if for any $y \in Y$ there exists an $x \in X$ such that

$$Ax \leq y, \quad \text{and} \quad \|x\|_X \leq \kappa(A)\|y\|_Y. \tag{20}$$

To utilize this definition, it is convenient to introduce a norm $\|\cdot\|_Z$ on $\mathrm{Ker}(Q)$ which may differ from $\|\cdot\|_{\mathcal{H}}$. In doing so note that for any $x \in \mathcal{H}$, in addition to its $\mathcal{H}$ norm, we can define its Z^* seminorm by

$$\|x\|_{Z^*} := \sup_{\substack{z\in\mathrm{Ker}(Q)\\ \|z\|_Z\leq 1}} \langle x, z\rangle.$$

The following theorem gives conditions when solving a linear programming problem determined by an approximate solution to (D) determines an approximate solution to (P).

Theorem 3.1 *Consider the quadratic convex programming problem (P). Suppose that (P) has the value $v > -\infty$ and that there is no duality gap. Let ℓ_∞^N denote*

$\mathbb{R}^N$ *equipped with the norm* $\|y\|_\infty = \max_{i=1}^N |y_i|$. *Let* Z *denote the space* $\mathrm{Ker}(Q)$ *equipped with a norm* $\|\cdot\|_Z$. *Suppose that* $A : \mathcal{H} \to \ell_\infty^N$ *is continuous and that* $A : Z \to \ell_\infty^N$ *has finite modulus* $\kappa(A)$. *Consider the dual problem* (D) *and suppose that* $\lambda^* \in \mathcal{O}_\epsilon(\mathrm{D})$ *for some fixed* $\epsilon \geq 0$. *Define*

$$\hat{y} := -Q^{-1}(A^t\lambda^* + q)$$

and for some fixed $\epsilon_0 \geq 0$ *select a* $\hat{z} \in \mathrm{Ker}(Q)$ *that* ϵ_0*-approximately solves the linear programming problem*

$$\min_{\substack{Qz=0 \\ Az \leq b - A\hat{y}}} \langle q, z\rangle.$$

That is, it should satisfy

$$Q\hat{z} = 0, \qquad A\hat{z} \leq b - A\hat{y}, \qquad \langle q, \hat{z}\rangle \leq \min_{\substack{Qz=0 \\ Az \leq b - A\hat{y}}} \langle q, z\rangle + \epsilon_0.$$

Define

$$K := \sqrt{4\|Q\|(\epsilon - \nu + \kappa(A)\|q\|_{Z^*}\|b\|_\infty) + 2\|q\|_{\mathcal{H}}^2}$$

and

$$\acute{\epsilon} := 4\epsilon + 4\|Q^{-1}\|^{\frac{1}{2}}(K + \kappa(A)\|q\|_{Z^*}\|A : \mathrm{Ker}(Q)^\perp \to \ell_\infty^N\|)\sqrt{\epsilon} + 2\epsilon_0.$$

Then, for $\hat{x} := \hat{y} + \hat{z}$, *we have*

$$\hat{x} \in \mathcal{O}_{\acute{\epsilon}}(\mathrm{P}).$$

Proof Since Hiriart-Urruty et al. [9, Example XI.1.2.2, p. 95] show that

$$\partial_{2\epsilon}\left(\frac{1}{2}\langle Qx, x\rangle + \langle q, x\rangle\right) = \{Q(x + \dot{x}) + q : \dot{x} \in \mathcal{H},\ \langle Q\dot{x}, \dot{x}\rangle \leq 4\epsilon\}, \tag{21}$$

we observe that the Kuhn–Tucker equations $\mathrm{KTC}_{2\epsilon}$ at (x, λ) are equivalent to

$$Ax - b \leq 0, \tag{22}$$

$$-2\epsilon \leq \langle \lambda, Ax - b\rangle, \tag{23}$$

$$\langle Q\dot{x}, \dot{x}\rangle \leq 4\epsilon, \tag{24}$$

$$0 = Q(x + \dot{x}) + A^t\lambda + q. \tag{25}$$

Since $\lambda^* \in \mathcal{O}_\epsilon(\mathrm{D})$, Corollary 2.3(i) states that there exists an x^* and $\dot{x}^*$ such that (x^*, λ^*) and $\dot{x}^*$ satisfy (22–25). Decompose $x^* = y^* + z^*$ and $\dot{x}^* = \dot{y}^* + \dot{z}^*$ where the y-components are in $\mathrm{Ker}(Q)^\perp$ and the z-components are in $\mathrm{Ker}(Q)$. Then, we obtain

$$0 = Q(y^* + \dot{y}^*) + A^t\lambda^* + q$$

and so conclude that $\hat{y} = \dot{y}^* + y^*$. Moreover, it follows that for $\hat{x} := \hat{y} + \hat{z}$ we obtain

$$0 = Q(\hat{x}) + A^t\lambda^* + q;$$

consequently, we observe that $(\hat{x}, \lambda^*)$ and $\dot{\hat{x}} := 0$ satisfy three (22), (24), (25) out of the four Kuhn–Tucker conditions of $\mathrm{KTC}_{2\epsilon}$ and since $\frac{\epsilon}{2} \geq 2\epsilon$ it follows that $(\hat{x}, \lambda^*)$ and $\dot{\hat{x}} := 0$ satisfy three (22), (24), (25) out of the four Kuhn–Tucker conditions of $\mathrm{KTC}_{\frac{\epsilon}{2}}$. We now proceed to show that $\langle \lambda^*, A\hat{x} - b\rangle \geq -\frac{\epsilon}{2}$ and so conclude that

$$(\hat{x}, \lambda^*) \in \mathrm{KTC}_{\frac{\epsilon}{2}}.$$

To that end, observe that, since $\hat{x} = \hat{y} + \hat{z} = \dot{y}^* + y^* + \hat{z} = \dot{y}^* + x^* - z^* + \hat{z}$, we have

$$\langle \lambda^*, A\hat{x} - b\rangle = \langle \lambda^*, Ax^* - b\rangle + \langle A^t\lambda^*, \dot{y}^*\rangle + \langle \lambda^*, A\hat{z} - Az^*\rangle. \tag{26}$$

We proceed by bounding the three terms separately. Equation (23) implies that the first term satisfies $\langle \lambda^*, Ax^* - b\rangle \geq -2\epsilon$. For the second term we bound $\|\dot{y}^*\|_{\mathcal{H}}$ and $\|A^t\lambda^*\|_{\mathcal{H}}$. To bound $\|\dot{y}^*\|_{\mathcal{H}}$, observe that $\dot{x}^*$ satisfies (24) so that we obtain

$$\begin{aligned}\|\dot{y}^*\|^2_{\mathcal{H}} &= \|Q^{-\frac{1}{2}}Q^{\frac{1}{2}}\dot{y}^*\|^2_{\mathcal{H}} \leq \|Q^{-\frac{1}{2}}\|^2\|Q^{\frac{1}{2}}\dot{y}^*\|^2_{\mathcal{H}} = \|Q^{-1}\|\langle Q\dot{y}^*, \dot{y}^*\rangle \\ &= \|Q^{-1}\|\langle Q\dot{x}^*, \dot{x}^*\rangle \leq 4\|Q^{-1}\|\epsilon.\end{aligned} \tag{27}$$

To bound $\|A^t\lambda^*\|_{\mathcal{H}}$, observe that the assumption on λ^* implies that $\lambda^* \geq 0$, $A^t\lambda^* + q \perp \mathrm{Ker}(Q)$ and

$$-\frac{1}{2}\langle Q^{-1}(A^t\lambda^* + q), (A^t\lambda^* + q)\rangle - \langle \lambda^*, b\rangle \geq \nu - \epsilon.$$

However, since A has modulus $\kappa(A)$, there exists a $z_0 \in Z$ such that

$$Az_0 \leq b \quad \text{and} \quad \|z_0\|_Z \leq \kappa(A)\|b\|_\infty,$$

and since $A^t\lambda^* + q \perp \mathrm{Ker}(Q)$ and $\lambda^* \geq 0$, we conclude that

$$\langle \lambda^*, b\rangle \geq \langle \lambda^*, Az_0\rangle = \langle A^t\lambda^*, z_0\rangle = -\langle q, z_0\rangle \geq -\|q\|_{Z^*}\|z_0\|_Z \geq -\kappa(A)\|q\|_{Z^*}\|b\|_\infty,$$

so that we obtain

$$\frac{1}{2}\langle Q^{-1}(A^t\lambda^* + q), (A^t\lambda^* + q)\rangle \leq \epsilon - \nu - \langle \lambda^*, b\rangle \leq \epsilon - \nu + \kappa(A)\|q\|_{Z^*}\|b\|_\infty.$$

Since

$$\begin{aligned}\|A^t\lambda^* + q\|^2_{\mathcal{H}} &= \|Q^{\frac{1}{2}}Q^{-\frac{1}{2}}(A^t\lambda^* + q)\|^2_{\mathcal{H}} \leq \|Q^{\frac{1}{2}}\|^2\|Q^{-\frac{1}{2}}(A^t\lambda^* + q)\|^2_{\mathcal{H}} \\ &= \|Q\|\langle Q^{-1}(A^t\lambda^* + q), (A^t\lambda^* + q)\rangle,\end{aligned}$$

and since $\|A^t\lambda^* + q\|^2_{\mathcal{H}} \geq \frac{1}{2}\|A^t\lambda^*\|^2_{\mathcal{H}} - \|q\|^2_{\mathcal{H}}$, we obtain that

$$\|A^t\lambda^*\|^2_{\mathcal{H}} \leq 4\|Q\|(\epsilon - \nu + \kappa(A)\|q\|_{Z^*}\|b\|_\infty) + 2\|q\|^2_{\mathcal{H}} = K^2. \tag{28}$$

For the third term in (26) observe that the definition of the modulus $\kappa(A)$ applied to the vector $-A\dot{y}^*$ implies that there exists a $z_1 \in \text{Ker}(Q)$ such that $Az_1 \le -A\dot{y}^*$ and $\|z_1\|_Z \le \kappa(A)\|A\dot{y}^*\|_\infty$. Moreover, (27) implies that

$$\|z_1\|_Z \le \kappa(A)\|A\dot{y}^*\|_\infty \le \kappa(A)\|\acute{A}\|\|\dot{y}^*\|_{\mathcal{H}} \le 2\kappa(A)\|\acute{A}\|\|Q^{-1}\|^{\frac{1}{2}}\sqrt{\epsilon},$$

where we have used the shorthand notation $\acute{A}$ for the operator $A : \text{Ker}(Q)^\perp \to \ell_\infty^N$. Since $z^* = x^* - \hat{y} + \dot{y}^*$, it follows from (22) and the definition of z_1 that, for $\acute{z} := z^* + z_1$, we have

$$A\acute{z} = Ax^* - A\hat{y} + A\dot{y}^* + Az_1 \le b - A\hat{y} + A\dot{y}^* + Az_1 \le b - A\hat{y}$$

and

$$\begin{aligned}\langle A^t\lambda^*, \acute{z} - z^*\rangle = \langle A^t\lambda^*, z_1\rangle = -\langle q, z_1\rangle &\ge -\|q\|_{Z^*}\|z_1\|_Z \\ &\ge -2\kappa(A)\|q\|_{Z^*}\|\acute{A}\|\|Q^{-1}\|^{\frac{1}{2}}\sqrt{\epsilon}.\end{aligned}$$

Moreover, since $\hat{z}$ is an ϵ_0 -approximate solution and $A\acute{z} \le b - A\hat{y}$, we obtain that

$$\begin{aligned}\langle A^t\lambda^*, \hat{z} - z^*\rangle = -\langle q, \hat{z}\rangle - \langle A^t\lambda^*, z^*\rangle &\ge -\langle q, \acute{z}\rangle - \epsilon_0 - \langle A^t\lambda^*, z^*\rangle \\ &= \langle A^t\lambda^*, \acute{z} - z^*\rangle - \epsilon_0,\end{aligned}$$

and so conclude that

$$\langle A^t\lambda^*, \hat{z} - z^*\rangle \ge -2\kappa(A)\|q\|_{Z^*}\|\acute{A}\|\|Q^{-1}\|^{\frac{1}{2}}\sqrt{\epsilon} - \epsilon_0. \tag{29}$$

Therefore, from (26), and the bounds (27), (28), and (29) we conclude that

$$\langle \lambda^*, A\hat{x} - b\rangle \ge -2\epsilon - 2\|Q^{-1}\|^{\frac{1}{2}}(K + \kappa(A)\|q\|_{Z^*}\|\acute{A}\|)\sqrt{\epsilon} - \epsilon_0$$

thus proving that

$$(\hat{x}, \lambda^*) \in \text{KTC}_{\frac{\acute{\epsilon}}{2}}.$$

Corollary 2.3(ii) then completes the proof. □

Example 3.1 (Support Vector Machines) The support vector machine (SVM), introduced by Vapnik (see e.g. [10]), has become an important method for building a predictive model from ground truth examples. A major component of the SVM method is the computation of an approximate solution to a (potentially large) convex quadratic programming (QP) problem. Because the Lagrangian dual QP problem is often much smaller than the primal QP problem it is common to compute an approximate solution to the dual and then map the approximate dual solution to an approximate primal solution. Recently, Steinwart and Scovel [11] have shown that the predictive performance of the model can be analyzed in terms of the accuracy of the approximate primal solution. In this example we show how to construct an approximate primal solution from an approximate dual solution and quantify the accuracy of the corresponding approximate primal solution. Consider the SVM primal problem

$$\begin{aligned} \min \quad & \gamma\|\psi\|^2 + \langle u, \xi\rangle, \quad \psi \in H, \xi \in \mathbb{R}^n,\ \eta \in \mathbb{R}, \\ \text{s.t.} \quad & 1 - y_i(\langle\psi, z_i\rangle + \eta) - \xi_i \le 0, \quad i = 1, \dots, n, \\ & -\xi_i \le 0, \quad i = 1, \dots, n, \end{aligned} \tag{30}$$

where H is a Hilbert space, $y_i \in \{-1, 1\}$, $z_i \in H$, $u_i \ge 0$, $i = 1, \dots, n$ and $\sum_{i=1}^n u_i = 1$. Note that in the typical case as presented in [10] we have $u_i := \frac{1}{n}$, $i = 1, \dots, n$ so that $\langle u, \xi\rangle = \frac{1}{n}\sum_{i=1}^n \xi_i$. To represent (30) in the form (2), we let $\mathcal{H} := H \oplus \mathbb{R}^n \oplus \mathbb{R}$, $Q := 2\gamma P_H$ where P_H denotes the orthogonal projection from $\mathcal{H}$ to H, and $q := (0, u, 0) \in H \oplus \mathbb{R}^n \oplus \mathbb{R}$. Let $N = 2n$ and define $W : H \to \mathbb{R}^n$ to be the operator defined by $(W\psi)_i = \langle z_i, \psi\rangle$, $i = 1, \dots, n$, Y the diagonal matrix acting on $\mathbb{R}^n$ naturally associated with the vector y. If we define

$$A := \begin{bmatrix} -YW & -I_n & -y \\ 0 & -I_n & 0 \end{bmatrix}, \qquad b := \begin{bmatrix} -\mathbf{1}_n \\ 0_n \end{bmatrix},$$

where I_n is the n-dimensional identity and $\mathbf{1}_n \in \mathbb{R}^n$ is the vector of 1's, we complete the representation. A straightforward calculation shows that the linear programming problem of Theorem 3.1 is

$$\min_{\substack{\xi \ge 0, \eta \in \mathbb{R} \\ \xi_i \ge 1 - y_i(\frac{1}{2\gamma}\sum_{j=1}^n \lambda_j^*\langle z_i, z_j\rangle + \eta),\ i=1,\dots,n}} \langle \xi, u\rangle$$

and we note that, by defining the functions

$$\xi_i(\eta) := \max\left(0, 1 - y_i\left(\frac{1}{2\gamma}\sum_{j=1}^n \lambda_j^*\langle z_i, z_j\rangle + \eta\right)\right), \quad i = 1, \dots, n,$$

this linear program problem can be solved by solving the one-dimensional convex programming problem

$$\min_{\eta\in\mathbb{R}} \langle \xi(\eta), u\rangle.$$

Let us put the ℓ_∞^{n+1} norm on $Z = \ker(Q) = \mathbb{R}^n \oplus \mathbb{R}$ so that Z^* has the ℓ_1^{n+1} norm. Elementary calculations then show that A has modulus $\kappa(A) = 1$ and that $\|q\|_{Z^*} = \|(u, 0)\|_{\ell_1^{n+1}} = 1$, $\|q\|_{\mathcal{H}} = \|u\|_{\ell_2^n}$, $\|b\|_\infty = 1$, $\|Q\| = 2\gamma$, $\|Q^{-1}\| = \frac{1}{2\gamma}$, and $\|A\|_{\mathrm{Ker}(Q)^\perp, \ell_\infty^N} \le R$ where $R := \sup_{i=1}^n \|z_i\|_H$. Moreover, one can easily show that there is no duality gap and $\nu \ge 0$ so that we obtain $K \le \sqrt{8\gamma(\epsilon + 1) + 2\|u\|_{\ell_2^n}^2}$ and, for the $\acute{\epsilon}$ of Theorem 3.1, we have

$$\acute{\epsilon} \le 4\epsilon + 4\left(\sqrt{4\epsilon + 4 + \frac{\|u\|_{\ell_2^n}^2}{\gamma}} + \frac{R}{\sqrt{2\gamma}}\right)\sqrt{\epsilon} + 2\epsilon_0.$$

Note that, for the typical choice of $u_i = \frac{1}{n}$, $i = 1, \dots, n$, used for example in [11], we have $\|u\|_{\ell_2^n} = \frac{1}{\sqrt{n}}$.

References

1. Strodiot, J.-J., Hien Nguyen, V., Huekemes, N.: ϵ-optimal solutions in nondifferentiable convex programming and some related questions. Math. Program. **25**, 307–328 (1983)
2. Ekeland, I.: On the variational principle. J. Math. Anal. Appl. **47**, 324–353 (1974)
3. Mordukhovich, B.S., Wang, B.: Necessary suboptimality and optimality conditions via variational principles. SIAM J. Control Optim. **41**, 623–640 (2002)
4. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
5. Varaiya, P.P.: Nonlinear programming in Banach space. SIAM J. Appl. Math. **15**, 284–293 (1967)
6. Zălinescu, C.: Convex Analysis in General Vector Spaces. World Scientific, River Edge (2002)
7. Yokoyama, K., Shiraishi, S.: An ϵ-optimality condition for convex programming problems without Slater's constraint qualification. Preprint (2006)
8. Barbu, V., Precupanu, Th.: Convexity and Optimization in Banach Spaces. D. Reidel Publishing Company, Dordrecht (1986)
9. Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms II. Springer, New York (1996)
10. Vapnik, V.N.: Statistical Learning Theory. Wiley, New York (1998)
11. Steinwart, I., Scovel, C.: Fast rates for support vector machines using Gaussian kernels. Ann. Stat. **35**(2), 575–607 (2007)