Theory and Methodology

A simple duality proof in convex quadratic programming with a quadratic constraint, and some applications

Mustafa Ç. Pınar [*]

Department of Industrial Engineering, Faculty of Engineering, Bilkent University, 06533 Bilkent, Ankara, Turkey

Abstract

In this paper a simple derivation of duality is presented for convex quadratic programs with a convex quadratic constraint. This problem arises in a number of applications including trust region subproblems of nonlinear programming, regularized solution of ill-posed least squares problems, and ridge regression problems in statistical analysis. In general, the dual problem is a concave maximization problem with a linear equality constraint. We apply the duality result to: (1) the trust region subproblem, (2) the smoothing of empirical functions, and (3) to piecewise quadratic trust region subproblems arising in nonlinear robust Huber M-estimation problems in statistics. The results are obtained from a straightforward application of Lagrange duality. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Lagrange duality; Convex quadratic programming with a convex quadratic constraint; Ill-posed least squares problems; Trust region subproblems

1. Convex quadratic programs with an ellipsoidal constraint

Consider the problem (P)

$$\min_{y} \quad -d^{\mathrm{T}}y + \frac{1}{2}y^{\mathrm{T}}Qy$$

$$\text{subject to} \quad y^{\mathrm{T}}Py \leqslant \delta,$$

where Q is a symmetric, positive semidefinite $n \times n$ matrix, d an n vector not identically zero, P an $n \times n$ symmetric positive semidefinite matrix, y an n

vector, and δ a positive scalar. This problem arises in many applications including trust region subproblems of nonlinear programming [9,4], and regularization of ill-posed least squares problems [7]. It is also related to the technique of ridge regression in statistical estimation [7]. Recently, the problem has received renewed interest due to its relation to semidefinite programming; see Ref. [15]. The last reference derives a semidefinite dual problem to (P) for the case where Q is a symmetric, possibly indefinite matrix. The dual problem derived in Ref. [15] has a single variable and also applies to the convex case while it involves the pseudo-inverse of a certain symmetric matrix. It is a maximization problem over a positive semidefiniteness constraint on the matrix $Q - \lambda I$ where λ is

---
[*] Tel.: +90-312-290-1514; fax: +90-312-266-4126.

E-mail address: mustafap@bilkent.edu.tr (M.Ç. Pınar).

a scalar. Then, a semidefinite dual to this problem is given, and this primal–dual pair is used to motivate an algorithm for the trust region problem. Other related references that deal with the nonconvex case include Refs. [6,2] where dual problems to the nonconvex quadratic program with an ellipsoidal (or, spherical) constraint are derived. In particular, in Ref. [2] the problem is shown to be equivalent to a convex program through duality.

Our purposes in the present note are more modest. We wish to provide the interested reader with a compact and accessible reference on duality pertinent to convex quadratic programs with a single quadratic constraint. We also present a catalogue of three applications from the literature including the trust region subproblems. It is hoped that the present paper will serve to generate more insight to the designers of algorithms for the aforementioned problem class. Although the optimality conditions for the trust region subproblem (with $P = I$) (or, the regularization of ill-posed least squares problems) are well studied, resulting in efficient algorithms [9,4,7], to the best of our knowledge, derivation of duality for the convex trust region problem has not been exposed before in the simple form given below. In the present note we derive a dual problem to (P) using Lagrange duality [14]. Our dual problem is a concave maximization problem over linear constraints. In particular, in all cases the dual simplifies to a concave maximization problem with a quadratic term and a nondifferentiable two-norm term in the objective function. Our approach is essentially inspired from Ref. [17] where a Lagrange dual for entropy minimization problems is given. The main duality result of the present paper can be seen to be similar to the results of Refs. [11–13]. However, we use a more direct and simpler derivation technique from Lagrange duality. Baron [1] derives a Wolfe dual for the problem, which contains a large number of variables despite the simplicity of the derivation. Lagrange duality for such problems is also discussed in Ref. [18] using the theory of ℓ_p programming. This last reference discusses weak and strong duality, and uniqueness of solutions as well as regularity of ℓ_p programming problems. It is shown that these problems are solvable in polynomial time in Ref. [3]. A specialized interior-point method applied to truss topology design problems was implemented with success in Ref. [10].

In Section 2.1 we apply our duality result to quadratic trust region subproblems of nonlinear programming. In Section 2.2 we discuss the smoothing of empirical functions [19] by quadratic programming. Another contribution of the paper is to show in Section 2.3 that our derivation technique is also extended easily to minimization of piecewise quadratic objective functions over a quadratic (ellipsoidal) constraint. We illustrate this on an important problem from robust statistics.

The main result of the paper can be summarized in the following.

Proposition 1. (1) *The Lagrange dual of* (P) *is the following concave program* (D)

$$\max_{x \in \mathfrak{R}^m, \, z \in \mathfrak{R}^m, \, \mu \in \mathfrak{R}} \phi_1(x, \mu) - \frac{1}{2} z^\mathrm{T} z - \mu \delta$$

$$\text{subject to} \qquad A^\mathrm{T} z + \phi_2(x, \mu) = d,$$

$$\mu \geqslant 0,$$

where $Q = A^\mathrm{T} A$ *and* $P = E^\mathrm{T} E$.

$$\phi_1(x, \mu) = \begin{cases} -(1/4\mu)x^\mathrm{T} x & \text{if } \mu > 0, \\ 0 & \text{if } \mu = 0, \end{cases}$$

$$\phi_2(x, \mu) = \begin{cases} E^\mathrm{T} x & \text{if } \mu > 0, \\ 0 & \text{if } \mu = 0, \end{cases}$$

under the condition that $\mu = 0$ *implies* $x = 0$.

(2) *The optimal solution of the dual problem* (z^*, x^*, μ^*) *for* $\mu^* > 0$ *and a primal optimal solution* y^* *are related by the identities*

$$Ey^* = \frac{x^*}{2\mu^*} \tag{1}$$

and

$$z^* = Ay^*. \tag{2}$$

Proof. Since Q is symmetric positive semidefinite, there exist full row rank matrices $A \in \mathfrak{R}^{m \times n}$ such that $Q = A^\mathrm{T} A$, and $E \in \mathfrak{R}^{m \times n}$ such that $P = E^\mathrm{T} E$. Now we can rewrite (P) as follows:

$$\min_{y,u,w} \quad -d^{\mathrm{T}}y + \frac{1}{2}u^{\mathrm{T}}u$$

subject to $\quad w^{\mathrm{T}}w \leqslant \delta,$

$$Ay - u = 0,$$

$$Ey - w = 0.$$

We associate the multipliers $z \in \mathfrak{R}^m$ with the equality constraints $Ay - u = 0$, and $x \in \mathfrak{R}^m$ with $Ey - w = 0$. Adding a nonnegative slack variable λ to the quadratic constraint $w^{\mathrm{T}}w \leqslant \delta$ and associating a multiplier μ we form the following Lagrangean problem:

$$\max_{z,x,\mu} \min_{y,u,w,\lambda \geqslant 0} \left\{ \frac{1}{2}u^{\mathrm{T}}u - d^{\mathrm{T}}y + \mu(w^{\mathrm{T}}w + \lambda - \delta) \right.$$
$$\left. + z^{\mathrm{T}}(Ay - u) + x^{\mathrm{T}}(E^{\mathrm{T}}y - w) \right\}. \quad (3)$$

This is equivalent to

$$\max_{z,x,\mu} - \delta\mu + \left\{ \min_u \left\{ \frac{1}{2}u^{\mathrm{T}}u - z^{\mathrm{T}}u \right\} \right.$$
$$+ \min_y \left\{ -d^{\mathrm{T}}y + z^{\mathrm{T}}Ay + x^{\mathrm{T}}E^{\mathrm{T}}y \right\}$$
$$\left. + \min_{\lambda \geqslant 0} \{\mu\lambda\} + \min_w \{\mu w^{\mathrm{T}}w - x^{\mathrm{T}}w\} \right\}. \quad (4)$$

The minimization over $\lambda \geqslant 0$ yields the requirement

$$\mu \geqslant 0. \quad (5)$$

The minimization over u yields $u = z$ which in turn gives the term $-\frac{1}{2}z^{\mathrm{T}}z$. The minimization over y gives the identity

$$A^{\mathrm{T}}z + E^{\mathrm{T}}x = d. \quad (6)$$

The minimization over w yields

$$w = \frac{x}{2\mu}, \quad (7)$$

if μ is non-zero. If $\mu = 0$ and $x_i \neq 0$ for some i, then the minimization over w yields $-\infty$. Hence, in this case we let $x = 0$. Substituting these expressions back into Lagrange function and rearranging terms we obtain (D). Note that in the case where $\mu = 0$, we obtain the dual problem

$$\max_{z \in \mathfrak{R}^n} \quad -\frac{1}{2}z^{\mathrm{T}}z$$

subject to $\quad A^{\mathrm{T}}z = d.$

For illustration we do the converse now, i.e., we start from (D) and obtain (P) as a dual assuming $\mu > 0$ at the optimal solution (the alternative case is much simpler and uninteresting). Associating multipliers $y \in \mathfrak{R}^n$ with the equality constraint in (D), we get the following Lagrangean problem:

$$\min_y \max_{z,x,\mu \geqslant 0} \frac{1}{4\mu}x^{\mathrm{T}}x - \frac{1}{2}z^{\mathrm{T}}z - \delta\mu$$
$$+ y^{\mathrm{T}}(A^{\mathrm{T}}z + E^{\mathrm{T}}x - d). \quad (8)$$

Rewrite this as

$$\min_y \left\{ -y^{\mathrm{T}}d + \max_z \left\{ -\frac{1}{2}z^{\mathrm{T}}z + y^{\mathrm{T}}A^{\mathrm{T}}z \right\} \right.$$
$$\left. + \max_{x,\mu \geqslant 0} \left\{ -\frac{1}{4\mu}x^{\mathrm{T}}x - \delta\mu + y^{\mathrm{T}}E^{\mathrm{T}}x \right\} \right\}. \quad (9)$$

Now, fix $\mu > 0$. The maximization over x yields the identity $x = 2\mu Ey$. Substituting this back, and after some algebraic simplification we obtain the term $\mu y^{\mathrm{T}}E^{\mathrm{T}}Ey - \delta\mu$ to be maximized over $\mu > 0$. This yields the equality $y^{\mathrm{T}}Py = \delta$. The maximization over z yields the identity $z = Ay$, which yields the term $\frac{1}{2}y^{\mathrm{T}}A^{\mathrm{T}}Ay$. But, this is precisely the problem (P) with the stipulation that at optimal (y^*, x^*, z^*, μ^*) strong duality between (P) and (D) is equivalent to the fact that $\mu^* > 0$ and $y^{*\mathrm{T}}Py^* = \delta$.

The concavity of the dual objective function for $\mu > 0$, $x \in \mathfrak{R}^m$ and $z \in \mathfrak{R}^m$ can be verified by simply forming the second derivative matrix from the objective function. This yields the matrix $H(z, x, \mu)$:

$$H(z,x,\mu) = \begin{pmatrix} -\frac{1}{\mu}I & 0 & 0 \\ 0 & -\frac{1}{2\mu}I & 0 \\ 0 & 0 & -\frac{1}{2\mu^3}x^{\mathrm{T}}x \end{pmatrix},$$

which is negative semidefinite for any positive μ, $x \in \mathfrak{R}^m$ and $z \in \mathfrak{R}^m$. Since the constraints are linear, the concavity follows. $\quad\square$

Note that in the case where $\mu > 0$ and the primal constraint is active (strict complementarity),

i.e., $y^{\mathrm{T}}Py = \delta$ at an optimal pair (y, μ) we can obtain a simplified dual problem. Since $x = 2\mu Ey$ we obtain $x^{\mathrm{T}}x/4 - \delta\mu^2 = 0$. Therefore in the case where strict complementarity holds the simplified dual is

$$\max_{x\in\Re^n, z\in\Re^n} \quad -\frac{1}{2}z^{\mathrm{T}}z - \sqrt{\delta}\|x\|_2$$

$$\text{subject to} \quad A^{\mathrm{T}}z + E^{\mathrm{T}}x = d.$$

Notice that the objective function has a quadratic term in z, and a nondifferentiable two-norm term in x.

2. Applications

2.1. The quadratic trust region subproblem

We consider the case where $P = I$, i.e., the trust region subproblem. This leads to the following corollary.

Corollary 1. (1) *The Lagrange dual of* (P) *(with* $P = I$) *is the following concave program* (D2):

$$\max_{z\in\Re^m, \mu\geqslant 0} \phi_3(z, \mu) - \frac{1}{2}z^{\mathrm{T}}z - \mu\delta,$$

where

$$\phi_3(z, \mu) = \begin{cases} \frac{1}{2\mu}\left(-\frac{1}{2}d^{\mathrm{T}}d + d^{\mathrm{T}}A^{\mathrm{T}}z - \frac{1}{2}z^{\mathrm{T}}AA^{\mathrm{T}}z\right) & \text{if } \mu > 0, \\ 0 & \text{if } \mu = 0, \end{cases}$$

under the condition that $\mu = 0$ *implies* $A^{\mathrm{T}}z = d$.

(2) *For the optimal solution of the dual problem* (z^*, μ^*) *with* $\mu^* > 0$ *the point*

$$y^* = \frac{d - A^{\mathrm{T}}z^*}{2\mu^*} \tag{10}$$

is an optimal solution to (P). *Furthermore, an optimal solution* y^* *to* (P) *and the optimal* z^* *to* (D2) *are also related by*

$$z^* = Ay^*. \tag{11}$$

This result is obtained by taking $E = I$, and substituting $d - A^{\mathrm{T}}z$ for x. For the case where strict

complementarity holds we have the following simple dual:

$$\max_z \quad -\frac{1}{2}z^{\mathrm{T}}z - \sqrt{\delta}\|d - A^{\mathrm{T}}z\|_2.$$

The concavity of the dual problem for $\mu > 0$ is again verified by forming the second derivative matrix from the dual objective function, which gives

$$H(z, \mu)$$
$$= \begin{pmatrix} -\frac{1}{2\mu}AA^{\mathrm{T}} - I & -\frac{1}{2\mu^2}(Ad - AA^{\mathrm{T}}z) \\ -\frac{1}{2\mu^2}(Ad - AA^{\mathrm{T}}z)^{\mathrm{T}} & \frac{1}{\mu^3}\left(-\frac{1}{2}d^{\mathrm{T}}d + d^{\mathrm{T}}Az - \frac{1}{2}z^{\mathrm{T}}AA^{\mathrm{T}}z\right) \end{pmatrix}.$$

The product

$$(z\ \mu)H(z, \mu)\begin{pmatrix} z \\ \mu \end{pmatrix}$$

yields $-z^{\mathrm{T}}z - \frac{1}{2\mu}d^{\mathrm{T}}d$ which is strictly negative for any z, and $\mu > 0$.

Notice that substituting (11) into (10) we obtain the well-known optimality condition for the trust region problem: namely that

$$y^* = \frac{d - A^{\mathrm{T}}Ay^*}{2\mu^*},$$

or equivalently,

$$(Q + 2\mu^*I)y^* = d$$

with $y^{*\mathrm{T}}y^* = \delta$, cf Lemma 3.5 of [9]. The above equation is also known as the *secular* equation.

2.2. An application to smoothing empirical functions

In [19] Terlaky treats the smoothing of empirical functions by means of mathematical programming. He develops duality results for such problems using the theory of ℓ_p programming. Here we will derive dual problems using our simple machinery of the previous section.

The problem of smoothing empirical functions is as follows. Let $\gamma_1, \ldots, \gamma_n$ be the observed (measured) values of a function f at equidistant points. Denote by $y_1, \ldots, y_n$ the unknown values of f at these points. Then the kth differences $\Delta^k y_1, \ldots, \Delta^k y_{n-k}$ where

$$\Delta^k y_i = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} y_{i+j}$$

are also unknown. One makes another observation for these kth differences. Let us denote the result by $\epsilon_1, \ldots, \epsilon_{n-k}$. The problem is to find $y_1, \ldots, y_n$ values that are not far from the $\gamma_1, \ldots, \gamma_n$ values such that the kth differences $\Delta^k y_1, \ldots, \Delta^k y_{n-k}$ are also good approximations for $\epsilon_1, \ldots, \epsilon_{n-k}$ values.

One way to find such y_i values is to solve the problem

$$\max_{y \in \Re^n} \quad \sum_{i=1}^{n-k} (\Delta^k y_i - \epsilon_i)^2$$

$$\text{subject to} \quad \sum_{i=1}^{n} (y_i - \gamma_i)^2 \leqslant \delta^2.$$

This model aims at minimizing the error in kth differences under the assumption that the Euclidean distance between $(\gamma_1, \ldots, \gamma_n)$ and $(y_1, \ldots, y_n)$ is at most δ. This problem can be rewritten as

$$\max_{y \in \Re^n} \quad \frac{1}{2} (Ay - e)^T (Ay - e)$$

$$\text{subject to } (y - c)^T (y - c) \leqslant \delta^2.$$

We can pose this model as

$$\min_{y, u, w} \quad \frac{1}{2} u^T u$$

$$\text{subject to} \quad w^T w \leqslant \delta,$$
$$Ay - e = u,$$
$$y - c = w.$$

From this point on we can carry out the derivation exactly as in the previous section. This yields the following dual:

$$\max_{z, x, \mu} \quad -\mu \delta^2 + x^T e - \frac{1}{2} x^T x + \phi_4(z, \mu)$$

$$\text{subject to} \quad A^T x + \phi_5(z) = 0,$$
$$\mu \geqslant 0,$$

where

$$\phi_4(z, \mu) = \begin{cases} -\frac{1}{4\mu} z^T z + z^T c & \text{if } \mu > 0, \\ 0 & \text{if } \mu = 0, \end{cases}$$

$$\phi_5(z, \mu) = \begin{cases} z & \text{if } \mu > 0, \\ 0 & \text{if } \mu = 0, \end{cases}$$

under the condition that $\mu = 0$ implies $z = 0$. Simplifying this model for the case where strict complementarity holds we obtain the following unconstrained dual:

$$\max_{x} x^T e - xAc - \frac{1}{2} x^T x - \delta \|A^T x\|_2.$$

It is easy to easy to see that the primal and dual optimal solutions y^* and x^*, respectively, are related by the identity

$$y^* = -\delta \frac{A^T x}{\|A^T x\|_2} + c.$$

A second model treated by Terlaky [19] assumes that the $\epsilon_1, \ldots, \epsilon_{n-k}$ values are good approximations to $\Delta^k y_1, \ldots, \Delta^k y_{n-k}$ values. That is, the Euclidean distance between the vectors $(\epsilon_1, \ldots, \epsilon_{n-k})$ and $(\Delta^k y_1, \ldots, \Delta^k y_{n-k})$ is at most δ. Here the optimization model is

$$\max_{y \in \Re^n} \quad \sum_{i=1}^{n} (y_i - \gamma_i)^2$$

$$\text{subject to} \quad \sum_{i=1}^{n-k} (\Delta^k y_i - \epsilon_i)^2 \leqslant \delta^2.$$

This problem can be rewritten as

$$\max_{y \in \Re^n} \quad \frac{1}{2} (y - c)^T (y - c)$$

$$\text{subject to} \quad (Ay - e)^T (Ay - e) \leqslant \delta^2.$$

This application is also straightforward using the same machinery as above, and results in the dual

$$\max_{z, x, \mu} \quad -\mu \delta^2 + x^T c - \frac{1}{2} x^T x + \phi_6(z, \mu)$$

$$\text{subject to} \quad x + \phi_7(z) = 0,$$
$$\mu \geqslant 0,$$

where

$$\phi_6(z, \mu) = \begin{cases} -\frac{1}{4\mu} z^T z + z^T e & \text{if } \mu > 0, \\ 0 & \text{if } \mu = 0, \end{cases}$$

$$\phi_7(z, \mu) = \begin{cases} A^T z & \text{if } \mu > 0, \\ 0 & \text{if } \mu = 0, \end{cases}$$

under the condition that $\mu = 0$ implies $z = 0$. Simplified for the strictly complementary case, this yields the dual

$$\max_z - z^{\mathrm{T}} Ac - \delta \|z\|_2 + z^{\mathrm{T}} e - \frac{1}{2} \|A^{\mathrm{T}} z\|_2^2.$$

It is easy to verify that dual optimal z^* and primal optimal y^* are related by

$$Ay^* = -\delta \frac{z}{\|z\|_2} + e.$$

2.3. An application to robust M-estimation

There has been considerable interest in the theory and algorithms for robust estimation in the past two decades. In particular, Huber's M-estimator [8] has received a great deal of attention from both theoretical and computational points of view. Robust estimation is concerned with identifying "outliers" among data points and giving them less weight. Huber's M-estimator is essentially the least squares estimator, which uses the ℓ_1-norm for points that are considered outliers with respect to a certain threshold. Hence, the Huber criterion is less sensitive to the presence of outliers.

More precisely, the Huber's M-estimate is a minimizer $x^* \in \mathfrak{R}^n$ of the function

$$F(x) = \sum_{i=1}^{m} \rho(r_i(x)/\sigma), \tag{12}$$

where

$$\rho(t) = \begin{cases} \frac{1}{2\gamma} t^2 & \text{if } |t| < \gamma, \\ |t| - \frac{1}{2} \gamma & \text{if } |t| \geqslant \gamma \end{cases} \tag{13}$$

with a tuning constant $\gamma > 0$, and a scaling factor σ that depends on the data to be estimated. The residual $r_i(x)$ is defined as

$$r_i(x) = a_i^{\mathrm{T}} x - b_i \tag{14}$$

for all $i = 1, \ldots, m$ with $r = A^{\mathrm{T}} x - b$. To view this minimization problem in a different format, define a "sign vector"

$$s(x) = [s_1(x), \ldots, s_m(x)] \tag{15}$$

with

$$s_i(x) = \begin{cases} -1 & \text{if } r_i(x) < -\gamma, \\ 0 & \text{if } |r_i(x)| \leqslant \gamma, \\ 1 & \text{if } r_i(x) > \gamma, \end{cases} \tag{16}$$

and

$$W = \operatorname{diag}(w_1, \ldots, w_m), \tag{17}$$

where

$$w_i = 1 - s_i^2. \tag{18}$$

Now, assuming a unit σ, the Huber's M-estimation problem can be expressed as the following minimization problem:

$$\text{minimize} \quad F(x) \equiv \frac{1}{2\gamma} r^T W r + s^T \left[r - \frac{1}{2} \gamma s \right], \tag{19}$$

where the argument x of r is dropped for notational convenience. Clearly, F measures the "small" residuals ($|r_i(x)| \leqslant \gamma$) by their squares while the "large" residuals are measured by the ℓ_1 function. Thus, F is a piecewise quadratic function, and it is once continuously differentiable in $\mathfrak{R}^n$.

In [5], the trust region approach was extended to nonlinear Huber M-estimation problems where the residual functions r_i are nonlinear. By linearizing the functions r_i at the current iterate, one obtains the following trust region subproblem:

$$\min_{r,x} \quad \frac{1}{2\gamma} r^{\mathrm{T}} W r + s^{\mathrm{T}} \left(r - \frac{1}{2} \gamma s \right)$$

$$\text{subject to} \quad r = A^{\mathrm{T}} x - b,$$
$$x^{\mathrm{T}} x \leqslant \delta.$$

Rewrite this problem as

$$\min_{r,x,\lambda \geqslant 0} \quad \frac{1}{2\gamma} r^{\mathrm{T}} W r + s^{\mathrm{T}} \left(r - \frac{1}{2} \gamma s \right)$$

$$\text{subject to} \quad r = A^{\mathrm{T}} x - b,$$
$$x^{\mathrm{T}} x + \lambda = \delta.$$

Attaching multipliers $y \in \mathfrak{R}^m$ and $\mu \in \mathfrak{R}$ to the two sets of constraints, respectively, we form the Lagrangean problem

$$\max_{y,\mu} \min_{r,x,\lambda \geqslant 0} \frac{1}{2\gamma} r^{\mathrm{T}} W r + s^{\mathrm{T}} \left(r - \frac{1}{2} \gamma s \right)$$
$$+ y^{\mathrm{T}} (A^{\mathrm{T}} x - b - r) + \mu (x^{\mathrm{T}} x + \lambda - \delta).$$

This separates into the minimization problems over $\lambda \geqslant 0$, x and r, respectively, after pulling out the constant terms $-b^{\mathrm{T}} y - \mu\delta$. The minimization over λ yields the constraint $\mu \geqslant 0$. The terms with x give the expression

$$x = -\frac{Ay}{2\mu} \tag{20}$$

with the objective function term $-\frac{1}{4\mu} y^{\mathrm{T}} A^{\mathrm{T}} A y$. The minimization over r requires a bit more attention since this is a piecewise quadratic term. The simple trick here is to work with the scalar term $\frac{1}{2\gamma} r_i^2 - y_i r_i$ which is valid only if $|r_i| \leqslant \gamma$. But, the minimization over r_i yields $r_i = \gamma y_i$ which is equivalent to saying that

$$-1 \leqslant y_i \leqslant 1$$

for all i. For the linear segment we obtain the condition $y_i = s_i$ for the minimization over r to yield a bounded optimal value. Plugging the expression $r = \gamma y$ into $\frac{1}{2\gamma} r^{\mathrm{T}} r - y^{\mathrm{T}} r$ we obtain the term $-\frac{1}{2} \gamma y^{\mathrm{T}} y$. So, we have the dual problem

$$\max_{y,\mu \geqslant 0} \quad -\frac{1}{2} \gamma y^{\mathrm{T}} y - \frac{1}{4\mu} y^{\mathrm{T}} A^{\mathrm{T}} A y - b^{\mathrm{T}} y - \delta\mu$$

subject to $\quad -1 \leqslant y \leqslant 1$,

where strong duality holds for optimal $y^*, x^*, \mu^* > 0$ as in Corollary 1. Note also that the dual solution is related to the primal solution by the identity (20) and the following:

$$y^* = \frac{1}{\gamma} W r(x^*) + s$$

with $s = s(x^*)$ and W is derived from s.

Notice that when $\mu = 0$ from the term $\min_x y^{\mathrm{T}} A^{\mathrm{T}} x + \mu x^{\mathrm{T}} x$ one obtains the requirement of $Ay = 0$. Therefore, in the case where the primal is essentially unconstrained we have the dual

$$\max_{y} \quad -\frac{1}{2} \gamma y^{\mathrm{T}} y - b^{\mathrm{T}} y$$

subject to $\quad Ay = 0$,
$$-1 \leqslant y \leqslant 1.$$

When strict complementarity holds a simplification of the dual as in Section 1 is possible. After straightforward calculation we get

$$\max_{y} \quad -\frac{1}{2} \gamma y^{\mathrm{T}} y - \sqrt{\delta} \|Ay\|_2 - b^{\mathrm{T}} y$$

subject to $\quad -1 \leqslant y \leqslant 1.$

Finally, we note that optimality conditions for nonconvex piecewise quadratic trust region subproblems are investigated in [16].

Acknowledgements

References

[1] D.P. Baron, Quadratic programming with a quadratic constraint, Naval Research Logistics Quarterly 19 (1972) 105–119.

[2] A. Ben-Tal, M. Teboulle, Hidden convexity in some nonconvex quadratically constrained quadratic programming, Mathematical Programming 72 (1996) 51–63.

[3] D. den Hertog, F. Jarre, C. Roos, T. Terlaky, A sufficient condition for self-concordance with application to some classes of structured convex programming problems, Mathematical Programming 69 (1995) 75–88.

[4] J.E. Dennis, R.E. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, second printing, SIAM Classics in Applied Mathematics, Philadelphia, 1996.

[5] O. Edlund, Linear M-estimation with bounded variables, BIT 37 (1997) 13–23.

[6] O.E. Flippo, B. Jansen, Duality and sensitivity in nonconvex quadratic optimization over an ellipsoid, European Journal of Operational Research 94 (1996) 167–178.

[7] G.H. Golub, C. Van Loan, Matrix computations, The Johns Hopkins University Press, Baltimore, MD, 1989.

[8] P.J. Huber, Robust Statistics, Wiley, New York, 1980.

[9] J.J. Moré, D.C. Sorensen, Newton's method, in: G.H. Golub (Ed.), Studies in Numerical Analysis, 1984, pp. 29–82.

[10] F. Jarre, M. Kocvara, J. Zowe, Truss Topology Design by Interior-Point Methods, Technical Report 173, Institut für Angewandte Mathematik, Universität Erlangen-Nürnberg, Erlangen, Germany, 1996.

[11] E.L. Peterson, J.G. Ecker, Geometric Programming: Duality and ℓ_p Approximation I, in: H.W. Kuhn, A.W.

Tucker (Eds.), Proceedings of the International Symposium on Mathematical Programming, Princeton, 1970.

[12] E.L. Peterson, J.G. Ecker, Geometric programming: Duality in quadratic programming and ℓ_p approximation II, SIAM Journal on Applied Mathematics 17 (1969) 317–340.

[13] E.L. Peterson, J.G. Ecker, Geometric programming: Duality in degenerate programs quadratic programming and ℓ_p approximation III (Degenerate Programs), Journal of Mathematical Analysis and Applications 29 (1970) 365–383.

[14] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.

[15] F. Rendl, H. Wolkowicz, A semidefinite framework for trust region subproblems with applications to large scale minimization, Mathematical Programming B 77 (1997) 273–300.

[16] J. Sun, On piecewise quadratic newton and trust region problems, Mathematical Programming B 76 (1997) 451–468.

[17] M. Teboulle, A simple duality proof for quadratically constrained entropy functionals and extension to convex constraints, SIAM Journal on Applied Mathematics 49 (1989) 1845–1850.

[18] T. Terlaky, On ℓ_p programming, European Journal of Operational Research 22 (1985) 70–100.

[19] T. Terlaky, Smoothing empirical functions by ℓ_p programming, European Journal of Operational Research 27 (1986) 343–363.