This article is being published in the proceedings of the International Conference on Numerical Optimization and Applications (Xi'an, China, 1986).

A Generalized Approach to Linear-Quadratic Programming

*R.T. Rockafellar**
Department of Mathematics
University of Washington
Seattle, Washington 98195

Abstract. Basic duality theory associates primal and dual problems of optimization with any saddlepoint problem for a convex- concave function on a product of convex sets. When the function is at most quadratic and the sets are polyhedral, a natural class of optimization problems is obtained that appropriately generalizes traditional quadratic programming. This paper discusses the theory of such problems, the kinds of situations where they arise, and the current approaches to solving them.

* This work was supported in part by a grant from the National Science Foundation.

1. Introduction.

The customary description of a quadratic programming problem is that it consists of minimizing a quadratic convex function subject to linear constraints, or in other words, an expression $p \cdot x + \frac{1}{2} x \cdot Px$ (where P is symmetric and positive semidefinite) over a convex polyhedron. There are two flaws in such a definition, however. First, the dual of such a problem will usually not belong to the same class. Thus the adoption of this definition poses a serious conceptual and practical obstacle to the use of duality-based methods and results. Second, the possibility of penalty terms of a piecewise linear or quadratic nature is excluded.

What should really be identified as quadratic programming, at least in a generalized sense, is the widest class of problems whose optimality conditions can be written down in terms of linear complementarity. Such problems can be solved, at least in principle, by some kind of finitely terminating algorithm that pivots in a system of linear equations. It is not hard to see what the appropriate class is, after briefly considering the relationship between optimality conditions and duality.

Let us recall that saddlepoints and minimax problems provide a foundation for all discussion of optimality and duality in convex programming. Given a pair of convex sets $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^M$ and an expression $k(x, y)$ that is convex in $x \in X$ for each $y \in Y$, and also concave in $y \in Y$ for each $x \in X$, one also has an associated ***primal*** problem

$$(\mathcal{P}) \qquad \begin{aligned} &\text{minimize } f(x) \text{ over all } x \in X, \\ &\quad \text{where } f(x) = \sup_{y \in Y} k(x, y), \end{aligned}$$

and an associated *dual* problem

$$(\mathcal{Q}) \qquad \begin{aligned} &\text{maximize } g(y) \text{ over all } y \in Y, \\ &\quad \text{where } g(y) = \inf_{x \in X} k(x, y). \end{aligned}$$

In these problems f is convex and g is concave, so that $(\mathcal{P})$ and $(\mathcal{Q})$ are convex programming problems in the general sense of convex analysis (cf. [**1**]). However, f and g might be extended-real-valued, since the "sup" and "inf" defining them could be infinite.

The set of *feasible* solutions to $(\mathcal{P})$ is defined not to be X, necessarily, but rather the set of all $x \in X$ such that $f(x) < \infty$. Similarly, the set of feasible solutions to $(\mathcal{Q})$ is defined to be the set of $y \in Y$ such that $g(y) > -\infty$. The nature of these sets, and the possible expressions for f and g relative to them, must be investigated in particular cases. Whatever the details, it is always true that the optimal values in the two problems satisfy

$$\inf(\mathcal{P}) \geq \sup(\mathcal{Q}),$$

and that in the cases where actually $\inf(\mathcal{P}) = \sup(\mathcal{Q})$, a pair $(\overline{x}, \overline{y})$ is a saddle point of $k(x, y)$ relative to $x \in X$ and $y \in Y$ if and only if $\overline{x}$ is an optimal solution to $(\mathcal{P})$ and $\overline{y}$ is an optimal solution to $(\mathcal{Q})$. A full discussion of this kind of duality is provided in [**2**].

The traditional duality scheme in linear programming serves as a good illustration. Suppose $X = \mathbb{R}^n_+$, $Y = \mathbb{R}^m_+$, and

$$k(x, y) = c \cdot x + b \cdot y - y \cdot Ax. \tag{1}$$

Then

$$f(x) = \begin{cases} b \cdot x & \text{if } Ax \geq c, \\ \infty & \text{if } Ax \not\geq c, \end{cases}$$

so that $(\mathcal{P})$ amounts to minimizing $b \cdot x$ subject to $x \geq 0$ and $Ax \geq c$. At the same time

$$g(y) = \begin{cases} c \cdot y & \text{if } A^* y \leq b, \\ -\infty & \text{if } A^* y \not\leq b \end{cases}$$

(where A^* is the transpose of A), so $(\mathcal{Q})$ amounts to maximizing $c \cdot x$ subject to $y \geq 0$ and $A^* y \leq b$.

In the general case the condition that $(\overline{x}, \overline{y})$ be a saddlepoint of k relative to $X \times Y$ serves as a joint optimality condition for $(\mathcal{P})$ and $(\mathcal{Q})$, as already mentioned. If k is differentiable on a neighborhood of $X \times Y$, as will be assumed for simplicity, this condition can be written as

$$-\nabla_x k(\overline{x}, \overline{y}) \in N_X(\overline{x}) \text{ and } \nabla_y k(\overline{x}, \overline{y}) \in N_Y(\overline{y}),$$

where $N_X(\overline{x})$ and $N_Y(\overline{y})$ are *normal* cones to X and Y as defined in convex analysis. When can (2) be viewed as a "complementarity problem" in $\overline{x}$ and $\overline{y}$? In the strict sense this is true only when $X = \mathbb{R}^n_+$ and $Y = \mathbb{R}^m_+$, in which case the normal cone relations describe "complementary slackness". As a matter of fact, though, whenever X and Y are *polyhedral* it is fair to say that one has a complementarity problem, because the introduction of explicit constraint representations for X and Y (involving finitely many linear inequalities) and corresponding vectors of Lagrange multipliers leads immediately in that case to a re-expression of (2) along the stricter lines. Of course, the complementarity problem is *linear* if and only if the vectors $\nabla_x k(\overline{x}, \overline{y})$ and $\nabla_y k(\overline{x}, \overline{y})$ depend in an affine way on $\overline{x}$ and $\overline{y}$, and this is obviously equivalent to k being a function that is (no worse than) quadratic.

We are led by this route to a very natural class of problems.

Definition 1. Problems $(\mathcal{P})$ and $(\mathcal{Q})$ are problems of *generalized linear-quadratic programming* if X and Y are polyhedral convex sets, and $k(x, y)$ is quadratic in (x, y), convex in x and concave in y:

$$k(y, x) = c \cdot x + \tfrac{1}{2} x \cdot Cx + b \cdot y - \tfrac{1}{2} y \cdot By - y \cdot Ax, \tag{3}$$

where C and B are symmetric and positive semidefinite.

What exactly is the nature of $(\mathcal{P})$ and $(\mathcal{Q})$ in the case described in this definition, and does it justify the name? In answering this question, some notation is helpful. Let us associate with the polyhedral sets Y and matrix B the expression

$$\rho_{Y,B}(s) = \sup_{y \in Y} \{ s \cdot y - \tfrac{1}{2} y \cdot By \}$$

and likewise with X and C the expression

$$\rho_{X,C}(r) = \sup_{x \in X} \{ r \cdot x - \tfrac{1}{2} x \cdot Cx \}.$$

The two problems can then be written as follows,

$$(\mathcal{P}) \qquad \text{minimize } c \cdot x + \tfrac{1}{2} x \cdot Cx + \rho_{Y,B}(b - Ax) \text{ over } x \in X,$$

$$(\mathcal{Q}) \qquad \text{maximize } b \cdot y + \tfrac{1}{2} y \cdot By - \rho_{X,C}(A^* y - c) \text{ over } y \in Y.$$

Here $\rho_{Y,B}$ and $\rho_{X,C}$ are convex functions that could have $+\infty$ as a value. We have proved in general in [**3**] that each of these functions in *piecewise linear-quadratic* in the sense that its effective domain can be expressed as a union of finitely many polyhedral convex sets, relative to each of which the function value is given by a quadratic (or linear) formula. Thus $(\mathcal{P})$ consists of minimizing a piecewise linear-quadratic convex function over a certain convex polyhedron, whereas $(\mathcal{Q})$ consists of maximizing a piecewise linear-quadratic concave function over a convex polyhedron.

Linear programming is the case where $k(x, y)$ has the form (1) and $X = \mathbb{R}^n_+$, $Y = \mathbb{R}^m_+$, as already observed. General piecewise linear programming is obtained by keeping this form for $k(x, y)$ but allowing X and Y to be arbitrary convex polyhedra. Many other cases, corresponding for instance to having X and Y be boxes (preducts of intervals) and C and B be diagonal, are discussed in detail in [**3**]. In particular, an interesting correspondence between bounded variables and penalty terms is revealed. In all cases, however, the following duality theorem–as strong as the familiar one for linear programming–applies.

Theorem 1. *If either $(\mathcal{P})$ or $(\mathcal{Q})$ has finite optimal value, or if both $(\mathcal{P})$ and $(\mathcal{Q})$ have optimal solutions, then both have optimal solutions, and* $\min(\mathcal{P}) = \max(\mathcal{Q})$.

A proof of this theorem has been given in [**4**].

2. Solution Methods and Applications.

Any generalized linear-quadratic programming problem can be reformulated as a quadratic programming problem in the traditional sense and solved that way, if it is not too large. The reformulation tends to introduce numerical instabilities, however, as well as destroy the symmetry between primal and dual.

For problem $(\mathcal{P})$ as in Definition 1, the dual constraint representation

$$Y = \{y \in \mathbb{R}^m \mid Dy \leq d\}$$

leads, for instance, to $(\mathcal{P})$ being reformulated as the problem of minimizing

$$c \cdot x + \tfrac{1}{2} x \cdot Cx + u \cdot d + \tfrac{1}{2} v \cdot Bv$$

over all (x, u, v) that satisfy

$$x \in X,\ u \geq 0, \text{ and } Ax + Bv + D^* u = b.$$

(The optimality conditions for this encompass the ones for $(\mathcal{P})$.)
Here v is a sort of vector of "dummy variables" that causes some trouble. If $(\overline{x}, \overline{u}, \overline{v})$ solves the reformulated problem, then $\overline{x}$ solves $(\mathcal{P})$; a Lagrange multiplier vector for the constraint $Ax + Bv + D^* u = b$ gives at the same time an optimal solution to $(\mathcal{Q})$. The difficulty is that in this case any triple $(\overline{x}, \overline{u}, \overline{v}')$ with $B\overline{v}' = B\overline{v}$ also solves the reformulated problem, so there is an inherent degeneracy or indeterminancy (as long as B is not positive definite). In fact $(\overline{x}, \overline{u}, \overline{y})$ solves the reformulated problem in particular, so the constraints $Dv \leq d$ could be added without changing the optimal value; but this only serves to narrow the choice of $\overline{v}$ and is no real remedy. The indeterminancy in $\overline{v}$ can cause failure in computating solutions even when one is using an otherwise reliable and highly refined code for quadratic programming, as has been learned from hard experience.

A direct approach to solving $(\mathcal{P})$ and $(\mathcal{Q})$ without reformulation ought to be possible. For the case where the sets X and Y are *boxes* and the matrices B and C are *diagonal* (which is less of a restriction in practice than it may seem), the problems fall in the category of "monotropic programming", as developed in [5]. This case has been investigated by J. Sun in his recent dissertation [6]. Sun has found a method that combines the "active set" approach to quadratic programming with subgradient techniques of convex analysis so as to solve $(\mathcal{P})$ and $(\mathcal{Q})$ in finitely many steps. This method has many appealing features and certainly illustrates that the subject of quadratic programming is by no means finished!

Obviously, one cannot expect to solve any generalized linear-quadratic programming problem by a finitely terminating algorithm if the number of variables is very large. Yet it is just in connection with such large-scale problems that this class of problems takes on a special appeal.

A prime motivating example occurs in two-stage stochastic programming. There it is valuable to set up models that can be viewed as generalized linear-quadratic programming where the dual vector y belongs to a space of very high dimension, but the primal vector x is still low dimensional. There is no space here for a discussion of such stochastic programming problems; see [**4**] and [**7**]. Another example occurs in optimal control [**3**] and its discretization; there both x and y are high dimensional.

The beauty of large-scale linear-quadratic programming problems along such lines is that they can be reduced to solving a sequence of small-scale problems of the same sort. One such technique, introduced in [**4**] and now under sharp investigation, is called the ***finite generation method.*** It focuses on solving "approximate" subproblems $(\mathcal{P}^\nu)$ and $(\mathcal{Q}^\nu)$ that correspond to restricting $k(x,y)$ from $X \times Y$ to $(\text{ co } X^\nu) \times (\text{ co } Y^\nu)$, where X^ν and Y^ν are finite subsets of X and Y, and "co" denotes convex hull. (The superscript $\nu = 1, 2, \ldots$ counts the iterations of the procedure.)

The dimensionality of $(\mathcal{P}^\nu)$ and $(\mathcal{Q}^\nu)$ depends only on the number of points in X^ν and Y^ν, rather than on the spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ in which these points are given. To see this, consider the representation of a point $x \in \text{ co } X^\nu$ as

(4) $$x = \sum_{k=1}^{n^\nu} \xi_k x_k^\nu \text{ with } \xi_k \geq 0, \sum_{k=1}^{n^\nu} \xi_k = 1,$$

and similarly the representation of a point $y \in \text{ co } Y^\nu$ as

(5) $$y = \sum_{\ell=1}^{m^\nu} \eta_\ell y_\ell^\nu \text{ with } \eta_\ell \geq 0, \sum_{\ell=1}^{m^\nu} \eta_\ell = 1.$$

(Here x_k^ν for $k = 1, \ldots, n^\nu$ are the points that make up X^ν, while y_ℓ^ν for $\ell = 1, \ldots, m^\nu$ are the points that make up Y^ν.) The crucial observation is that when (4) and (5) are substituted into the expression (3) for $k(x,y)$, one obtains another quadratic form

$$k^\nu(\xi,\eta) = c^\nu \cdot \xi + \tfrac{1}{2}\xi \cdot C^\nu \xi + b^\nu \cdot \eta - \tfrac{1}{2}\eta \cdot B^\nu \eta - \eta^\nu \cdot A^\nu \xi^\nu$$

with *explicit coefficients.* Problems $(\mathcal{P}^\nu)$ and $(\mathcal{Q}^\nu)$ correspond to this form relative to $\xi \in S_{n^\nu}$ and $\eta \in S_{m^\nu}$, the unit *simplexes* of dimensions n^ν and m^ν. Thus they are generalized linear-quadratic programming problems of dimensions n^ν and m^ν, respectively.

Techniques such as have already been described in this section could be used therefore to solve $(\mathcal{P}^\nu)$ and $(\mathcal{Q}^\nu)$ as long as n^ν and m^ν are kept relatively small. The question then is how to pass from X^ν and Y^ν to sets $X^{\nu+1}$ and $Y^{\nu+1}$, within this limitation, and be assured that from the sequence of saddlepoints $(\overline{x}^\nu, \overline{y}^\nu)$ generated by the method one will be able to construct "approximate" solutions to the original problems $(\mathcal{P})$ and $(\mathcal{Q})$. The answer, as it has emerged so far, seems to involve taking advantage of separate decomposability properties of $k(x,y)$.

Suppose it is possible in a practical sense to solve for fixed $\overline{x}^\nu$ the quadratic programming problem

(6) $$\text{maximize } k(\overline{x}^\nu, y) \text{ over all } x \in X.$$

and likewise to solve for fixed $\overline{y}^\nu$ the quadratic program problem

$$\text{minimize } k(x, \overline{y}^\nu) \text{ over all } x \in X. \tag{7}$$

This can be true despite the potentially large dimensions of X and Y if X and Y are boxes and $k(x,y)$ is separable in x and separable in y (but not jointly). The optimal value in (6) is the primal objective value $f(\overline{x}^\nu)$ in $(\mathcal{P})$, and in calculating it one obtains a certain new point $y^\nu \in Y$. Likewise, the optimal value in (7) is the objective value $g(\overline{y}^\nu)$ in $(\mathcal{Q}^\nu)$, and in calculating it one obtains a new point $x^\nu \in X$.

The difference $\varepsilon^\nu = f(\overline{x}^\nu) - g(\overline{y}^\nu)$ is a measure of how close to optimality $\overline{x}^\nu$ and $\overline{y}^\nu$ are. (They come within ε^ν of giving the optimal values in $(\mathcal{P}^\nu)$ and $(\mathcal{Q}^\nu)$, because $f(\overline{x}^\nu) \geq \min(\mathcal{P}) = \max(\mathcal{Q}) \geq g(\overline{y}^\nu)$.) The new points x^ν and y^ν are obvious candidates for use along with $\overline{x}^\nu$ and $\overline{y}^\nu$ in forming $X^{\nu+1}$ and $Y^{\nu+1}$.

There are too many details to enter into here. A considerable discussion in the case of two-stage stochastic programming is given in [4]. Let it suffice to say that this is now an active new area of research in which we may soon see progress towards the solution of large-scale problems that until now have been beyond our capability.

References

1. R.T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, 1970.
2. R.T. Rockafellar, *Conjugate Duality and Optimization*, Regional Conference Series in Appl. Math. 16, SIAM Publications, 1974.
3. R.T. Rockafellar, "Linear-quadratic programming and optimal control", *SIAM J. Control Opt.*, forthcoming.
4. R.T. Rockafellar and R.J.B. Wets, "A Lagrangian finite generation technique for solving linear-quadratic problems in stochastic programming", *Math. Prog. Studies* **28**(1985), 63–93.
5. R.T. Rockafellar, *Network Flows and Monotropic Optimization*, Wiley and Sons, 1984.
6. J. Sun, *On Monotropic Piecewise Quadratic Programming*, dissertation, University of Washington, Seattle; 1986.
7. R.T. Rockafellar and R.J.-B. Wets, "Linear-quadratic programming problems with stochastic penalties", in *Numerical Techniques for Stochastic Optimization Problems* (Y. Ermoliev and R.J.-B. Wets, eds.), Springer-Verlag, Berlin, forthcoming in 1986.