Published in SIAG/OPT Views-and-News, No. 1, fall 1992

EXTENDED LINEAR-QUADRATIC PROGRAMMING

Terry Rockafellar

Most work in numerical optimization starts from the convention that the problem to be solved is given in the form

$$(\mathcal{P}) \qquad \begin{aligned} &\text{minimize } f_0(x) \text{ over all } x \in X, \\ &\text{such that } f_i(x) \begin{cases} \leq 0 & \text{for } i = 1, \ldots, s, \\ = 0 & \text{for } i = s+1, \ldots, m, \end{cases} \end{aligned}$$

with $X \subset \mathbb{R}^n$. But this notion of what optimization is all about may be unnecessarily limiting, both in the kind of modeling it promotes and the computational approaches it suggests. While all optimization eventually boils down to minimizing some function over some set, the formulation $(\mathcal{P})$ says nothing about the mathematical structure of the objective and instead puts all the emphasis on the structure of feasibility, insisting on "black-and-white" constraints which don't allow for gray areas of interaction between feasibility and optimality.

For many applications several objective function candidates are in the background of any attempt at optimization. Rather than choosing one of them to minimize while the others are held within precise bounds, it would make sense to form a joint expression out of "max" terms, penalty terms and the like. That could lead to a nonsmooth objective, but with special features. In $(\mathcal{P})$ there is no built-in way of handling such features.

In fact the horizons of practical optimization modeling can be widened considerably by providing for this extra structure in a manner conducive to computation. A key seems to be the use of composite terms, as is already well understood as a means of treating nonsmoothness numerically, and by admitting infinite penalties in some situations to integrate such terms into a problem statement that builds on the conventional one. The idea will be explained briefly here with particular attention to the linear-quadratic case.

An extended problem statement appearing to offer many advantages over $(\mathcal{P})$ is

$$(\overline{\mathcal{P}}) \qquad \text{minimize } f(x) = f_0(x) + \rho\big(F(x)\big) \text{ over } x \in X, \text{ where } F(x) = \big(f_1(x), \ldots, f_m(x)\big).$$

Here, as usual in numerical treatments of $(\mathcal{P})$, the set X can be simple (polyhedral, say) and the functions $f_0, f_1, \ldots, f_m$ can be smooth (C^2, say), but the function ρ on $\mathbb{R}^m$ can be nonsmooth and even extended-real-valued, although with form amenable to elementary convex analysis. Feasibility in $(\overline{\mathcal{P}})$ means that $x \in X$ and $F(x) \in D$, where $D = \{\, u \,|\, \rho(u) < \infty \,\}$. The case where $(\overline{\mathcal{P}})$ reduces to $(\mathcal{P})$ is thus the one where the function constraints in $(\mathcal{P})$ are represented through infinite penalties:

$$\rho(u) = \rho(u_1, \ldots, u_m) = \begin{cases} 0 & \text{when } u_i \leq 0 \text{ for } i \in [1, s] \text{ and} \\ & \qquad u_i = 0 \text{ for } i \in [s+1, m], \\ \infty & \text{otherwise.} \end{cases}$$

Such infinite penalties reflect the attitudes we force on the modeler in the traditional framework of $(\mathcal{P})$. The slightest violation of any constraint is supposed to cause infinite dissatisfaction; on the other hand, there is no reward offered for keeping comfortably within a given bound. In $(\overline{\mathcal{P}})$ there is much more flexibility.

The potential is already rich when $\rho(u) = \rho_1(u_1) + \cdots + \rho_m(u_m)$, so that

$$f(x) = f_0(x) + \rho_1\big(f_1(x)\big) + \cdots + \rho_m\big(f_m(x)\big). \tag{1}$$

We can think of ρ_i in general as converting the values of a particular f_i into units facilitating a trade-off with the values of f_0 and the other f_i's, but even if we cling to the notion of a putative constraint like $f_i(x) \leq 0$, we have new ways of expressing it. For instance, we can imagine ρ_i introducing a minor penalty as $f_i(x)$ starts to exceed 0, with this becoming more serious for larger violations and perhaps infinite for violations beyond a certain amount. In the other direction, ρ_i could give a *negative* penalty when $f_i(x)$ drops below 0, at least until a level is reached where no further reward is warranted.

The extended problem model $(\overline{\mathcal{P}})$ has been studied theoretically in [1], but a linear-quadratic programming version was proposed earlier in [2] out of needs in stochastic programming. (Models with black-and-white constraints are particularly inappropriate in optimization under uncertainty.) In bridging toward the linear-quadratic context, let's concentrate now on a single class of examples of expressions ρ_i which could be invoked in (1). These expressions, first introduced in [3], typically involve two linear pieces with a smooth quadratic interpolation between, but they also cover as limiting cases expressions in which the quadratic piece or one or both of the linear pieces may be missing, or where an infinite penalty might come up. They are parameterized in general by $\beta_i \geq 0$, $\hat{y}_i \in (-\infty, \infty)$, and a closed interval

$$Y_i = \{\, y_i \in \mathbb{R} \,|\, \hat{y}_i^- \leq \hat{y}_i \leq \hat{y}_i^+ \,\},$$

where the upper bound $\hat{y}_i^+$ could be ∞ and the lower bound $\hat{y}_i^-$ could be $-\infty$. (The reason for focusing on Y_i instead of just the two values $\hat{y}_i^+$ and $\hat{y}_i^-$ will emerge through duality below.) The formula for $\rho_i(u_i)$ as dictated by these parameters is best understood by starting with $\hat{\rho}_i(u_i) = \hat{y}_i u_i + (1/2\beta_i)u_i^2$, this being the unique quadratic function with $\hat{\rho}_i(0) = 0$, $\hat{\rho}_i'(0) = \hat{y}_i$, and $\hat{\rho}_i''(0) = 1/\beta_i$. Let $\hat{u}_i^+$ be the unique value such that $\hat{\rho}_i'(\hat{u}_i^+) = \hat{y}_i^+$, and similarly let $\hat{u}_i^-$ be the unique value such that $\hat{\rho}_i'(\hat{u}_i^-) = \hat{y}_i^-$. Then

$$\rho_i(u_i) = \rho_{Y_i,\beta_i,\hat{y}_i}(u_i) = \begin{cases} \hat{\rho}_i(\hat{u}_i^+) + \hat{y}_i^+(u_i - \hat{u}_i^+) & \text{when } u_i > \hat{u}_i^+, \\ \hat{\rho}_i(u_i) & \text{when } \hat{u}_i^- \le u_i \le \hat{u}_i^+, \\ \hat{\rho}_i(\hat{u}_i^-) + \hat{y}_i^-(u_i - \hat{u}_i^-) & \text{when } u_i < \hat{u}_i^-. \end{cases} \tag{2}$$

As extreme cases, if $\hat{y}_i^+ = \infty$ this is taken to mean that the quadratic graph is followed forever to the right without switching over to a tangential linearization; the interpretation for $\hat{y}_i^- = -\infty$ is analogous. The case of $\beta_i = 0$ is taken to mean that there is no quadratic middle piece at all: the function is given by $\hat{y}_i^+ u_i$ when $u_i > 0$ and by $\hat{y}_i^- u_i$ when $u_i < 0$. Possibly infinite values for $\hat{y}_i^+$ or $\hat{y}_i^-$ then yield infinite penalties.

Already in choosing expressions ρ_i in (1) just from this class, there are many ways of incorporating the functions f_i into an optimization model. A particular f_i can be treated for instance in terms of a constraint with infinite penalties for violation,

$$\begin{cases} Y_i = [0,\infty),\ \beta_i = 0,\ \hat{y}_i = 0 & \text{(inequality mode)}, \\ Y_i = (-\infty,\infty),\ \beta_i = 0,\ \hat{y}_i = 0 & \text{(equality mode)}, \end{cases}$$

classical linear penalties $d_i > 0$,

$$\begin{cases} Y_i = [0,d_i],\ \beta_i = 0,\ \hat{y}_i = 0 & \text{(inequality mode)}, \\ Y_i = [-d_i,d_i],\ \beta_i = 0,\ \hat{y}_i = 0 & \text{(equality mode)}, \end{cases}$$

classical quadratic penalties,

$$\begin{cases} Y_i = [0,\infty),\ \beta_i > 0,\ \hat{y}_i = 0 & \text{(inequality mode)}, \\ Y_i = (-\infty,\infty),\ \beta_i > 0,\ \hat{y}_i = 0 & \text{(equality mode)}, \end{cases}$$

a constraint replaced by an augmented Lagrangian term,

$$\begin{cases} Y_i = [0,\infty),\ \beta_i > 0,\ \hat{y}_i \ge 0 & \text{(inequality mode)}, \\ Y_i = (-\infty,\infty),\ \beta_i > 0,\ \hat{y}_i \text{ arb.} & \text{(equality mode)}, \end{cases}$$

or a modified augmented Lagrangian term with "saturation" bound $d_i > 0$,

$$\begin{cases} Y_i = [0,d_i],\ \beta_i > 0,\ \hat{y}_i \ge 0 & \text{(inequality mode)}, \\ Y_i = [-d_i,d_i],\ \beta_i > 0,\ \hat{y}_i \text{ arb.} & \text{(equality mode)}. \end{cases}$$

Even expressions with more than the three pieces directly allowed for in (2) can be taken care of. For instance, if we want to model f_1 with no penalty when $f_1(x) \leq 0$, a linear penalty rate $d_1 > 0$ when $0 < f_1(x) \leq 1$ but an infinite penalty if $f_1(x) > 1$, we can choose notation so that the function f_2 is $f_1 - 1$ and put a linear penalty expression as above on f_1 but an infinite penalty expression on f_2. Clearly, the range of modeling expressions easily representable through such tricks is enormous.

A strong property of the class of functions ρ_i in (2) is a *dual representation*: one has

$$\rho_{Y_i,\beta_i,\hat{y}_i}(u_i) = \sup_{\hat{y}_i^- \leq y_i \leq \hat{y}_i^+} \Big\{ u_i y_i - \tfrac{1}{2}\beta_i (y_i - \hat{y}_i)^2 \Big\}. \tag{3}$$

This leads us to consider more generally in $(\overline{\mathcal{P}})$ the class of all functions $\rho : \mathbb{R}^m \to (-\infty, \infty]$ representable dually as

$$\rho_{Y,B,\hat{y}}(u) = \sup_{y \in Y} \Big\{ u{\cdot}y - \tfrac{1}{2}(y - \hat{y}){\cdot}B(y - \hat{y}) \Big\}, \tag{4}$$

where Y is a nonempty *polyhedral* set in $\mathbb{R}^m$, B is a symmetric, positive *semi*definite matrix in $\mathbb{R}^{m \times m}$, and $\hat{y} = (\hat{y}_1, \ldots, \hat{y}_m)$ is some vector in $\mathbb{R}^m$. The examples of $(\overline{\mathcal{P}})$ we've been discussing so far correspond to the *box-diagonal* case of such a function, where

$$Y = Y_1 \times \cdots \times Y_m, \qquad B = \operatorname{diag}[\beta_1, \ldots, \beta_m],$$

for nonnegative values β_i and closed intervals Y_i, not necessarily bounded. An example of a multidimensional ρ function *not* conforming to the box-diagonal format is

$$\begin{aligned} \rho(u) &= \rho(u_1, \ldots, u_m) = \max\{u_1, \ldots, u_m\} \\ &= \rho_{Y,B,\hat{y}}(u) \;\text{ for }\; Y = \big\{\, y \,\big|\, y_i \geq 0,\; y_1 + \cdots + y_m = 1 \,\big\},\; B = 0,\; \hat{y} = 0. \end{aligned}$$

In this case, with f_0 taken to be $\equiv 0$ for instance, $(\overline{\mathcal{P}})$ would be a nonsmooth optimization problem of the form: minimize $f(x) = \max\big\{f_1(x), \ldots, f_m(x)\big\}$ over all $x \in X$.

Note by the way that the parameter vector $\hat{y}$ really adds no generality, because $\rho_{Y,B,\hat{y}} = \rho_{Y',B,0}$ for $Y' = Y - \hat{y}$ (translation). But the inclusion of this vector is convenient because in many cases it can stand for reference values for Lagrange multipliers. These can be estimated by the modeler as rates of change of the minimum value of the objective in $(\overline{\mathcal{P}})$ relative to shifts in the f_i values, cf. [1, Section 9].

The case of problem $(\overline{\mathcal{P}})$ called *extended linear-quadratic programming*, ELQP, is the one in which ρ belongs to the class (4), the set X is polyhedral, the function f_0 is convex linear-quadratic, and the functions $f_1, \ldots, f_m$ are affine. We can state this as

$(\overline{\mathcal{P}}_{\mathrm{lq}})$ minimize $f(x) = c{\cdot}x + \tfrac{1}{2}(x - \hat{x}){\cdot}C(x - \hat{x}) + \rho_{Y,B,\hat{y}}\big(b - Ax\big)$ over $x \in X$

for vectors $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, $\hat{x} \in \mathbb{R}^n$, a symmetric, positive *semi*definite matrix $C \in \mathbb{R}^{n\times n}$, and a matrix $A \in \mathbb{R}^{m\times n}$. As with $\hat{y}$, the parameter vector $\hat{x}$ adds no real generality—it can be taken to be 0 if desired—but is convenient often as an initial estimate of an optimal solution when proximal terms are being introduced to achieve strong convexity.

The ρ function in $(\overline{\mathcal{P}}_{\rm lq})$ can take the value ∞ to the extent that exact linear constraints are modeled with infinite penalties instead of being built into the specification of X. The feasible set is generally therefore not X, but

$$\left\{ x \in X \,\middle|\, b - Ax \in D_{Y,B} \right\}, \text{ where } D_{Y,B} = \left\{ u \,\middle|\, \rho_{Y,B,\hat{y}}(u) < \infty \right\} \tag{5}$$

(this doesn't actually depend on $\hat{y}$). It was shown in [4] that $D_{Y,B}$ is always a *polyhedral cone*. (It's the sum of the barrier cone for Y and the range space for B.) The feasible set in (5) is therefore polyhedral as well; only *linear constraints* are present in $(\overline{\mathcal{P}}_{\rm lq})$ in principle. On the other hand according to [4], the objective function f is convex and *piecewise linear-quadratic* on this feasible set. Due to the different ways of setting up penalties, there may be discontinuities in the first or second derivatives of f.

From this standpoint an ELQP problem may seem quite complicated in comparison with conventional LP or QP, but simplicity resurfaces through an associated *Lagrangian representation*: in terms of

$$L(x,y) = c{\cdot}x + \tfrac{1}{2}(x-\hat{x}){\cdot}C(x-\hat{x}) + b{\cdot}y - \tfrac{1}{2}(y-\hat{y}){\cdot}B(y-\hat{y}) - y{\cdot}Ax \text{ on } X \times Y,$$

the essential objective function in $(\overline{\mathcal{P}}_{\rm lq})$ is given by $f(x) = \sup_{y\in Y} L(x,y)$ for $x \in X$, as seen from (4). Thus: *ELQP problems are precisely the problems arising from Lagrangians L that are linear-quadratic convex-concave on a product $X \times Y$ of polyhedral sets.*

The symmetry in the generalized Lagrangian leads us to dualize in terms of maximizing $g(y) = \inf_{x\in X} L(x,y)$ over all $y \in Y$. We arrive then at the *dual problem*

$$(\overline{\mathcal{D}}_{\rm lq}) \qquad \text{maximize } g(y) = b{\cdot}y - \tfrac{1}{2}(y-\hat{y}){\cdot}B(y-\hat{y}) - \rho_{X,C,\hat{x}}\big(A^T y - c\big) \text{ over } y \in Y.$$

This is an ELQP problem expressed concavely instead of convexly. Its feasible set is

$$\left\{ y \in Y \,\middle|\, A^T y - c \in D_{X,C} \right\}, \text{ where } D_{X,C} = \left\{ v \,\middle|\, \rho_{X,C,\hat{x}}(v) < \infty \right\}.$$

The ρ function examples given above provide many interesting specializations. Traditional duality in linear programming and quadratic programming are covered, but much more. The theoretical properties of this duality are every bit as strong as in the classical cases, according to the following result from [2].

Theorem. *If either* $(\overline{\mathcal{P}}_{\rm lq})$ *or* $(\overline{\mathcal{D}}_{\rm lq})$ *has finite optimal value, then both problems have optimal solutions, and*

$$\min(\overline{\mathcal{P}}_{\rm lq}) \;=\; \max(\overline{\mathcal{D}}_{\rm lq}).$$

The pairs $(\bar{x},\bar{y}) \in X \times Y$ *such that* $\bar{x}$ *is an optimal solution to* $(\overline{\mathcal{P}}_{\rm lq})$ *and* $\bar{y}$ *is an optimal solution to* $(\overline{\mathcal{D}}_{\rm lq})$ *are precisely the saddle points of the associated Lagrangian* L *on* $X \times Y$ *and are characterized by the normal cone conditions*

$$-\nabla_x L(\bar{x},\bar{y}) \in N_X(\bar{x}), \qquad \nabla_y L(\bar{x},\bar{y}) \in N_Y(\bar{y}).$$

The development of good techniques for solving ELQP problems offers many open challenges. It was shown in [2] that any ELQP problem could, in principle, be reformulated as a conventional QP problem and solved that way, but the reformulation greatly increases the dimension and introduces possibly redundant constraints, which could cause numerical troubles in some situations. It also destroys the symmetry between the primal and dual and thereby threatens disruption of the kind of problem structure that ought to be put to use, especially in large-scale applications. Generalizations of complementarity algorithms could perhaps be applied effectively to the saddle point expression of optimality. Most of the efforts so far have been directed however at exploiting new kinds of decomposability that have come to light in ELQP applications with dynamics and stochastics [2], [3], [4], [5]. In [6] a class of "envelope" methods, something like bundle methods with smoothing, has been developed. Envelope ideas have been used differently in [7] to get generalized projected algorithms which operate with a novel kind of primal-dual feedback. These algorithms have solved problems with 100,000 primal and 100,000 dual variables, derived as discretized problems in optimal control [4], in half the time as the earlier algorithms in [6]. In [8] and [9] forward-backward splitting methods have been applied to the saddle point representation to take advantage of Lagrangian separability.

Besides offering direct possibilities in optimization modeling far beyond those available in conventional linear or quadratic programming, ELQP problems $(\overline{\mathcal{P}}_{\rm lq})$ can arise from general nonlinear problems $(\overline{\mathcal{P}})$ just like QP subproblems can arise from problems $(\mathcal{P})$ in schemes of sequential quadratic programming through second-order approximations to a Lagrangian function. (Lagrangian theory for $(\overline{\mathcal{P}})$ is furnished in [1].) There is lots to do, not only with ELQP as such, but in using ELQP techniques to solve extended problems $(\overline{\mathcal{P}})$ by Newton-like approaches.

REFERENCES

1. R. T. Rockafellar, “Lagrange multipliers and optimality,” SIAM Review, to appear.
2. 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** (1986), 63–93.
3. R. T. Rockafellar and R. J-B Wets, *Linear-quadratic problems with stochastic penalties: the finite generation algorithm*, in: *Numerical Techniques for Stochastic Optimization Problems*, Y. Ermoliev and R. J-B Wets (eds.), Springer-Verlag Lecture Notes in Control and Information Sciences No. 81, 1987, 545-560.
4. R. T. Rockafellar, “Linear-quadratic programming and optimal control,” SIAM J. Control Opt. **25** (1987), 781–814.
5. R. T. Rockafellar and R. J-B Wets, “Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time,” SIAM J. Control Opt. 28 (1990), 810–822.
6. R. T. Rockafellar, “Computational schemes for solving large-scale problems in extended linear-quadratic programming,” Math. Prog., Ser. B **48** (1990), 447–474.
7. C. Zhu and R. T. Rockafellar, “Primal-dual projected gradient algorithms for extended linear-quadratic programming,” SIAM J. Optimization, accepted.
8. P. Tseng, “Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming,” Math. Programming 48 (1990), 249–263.
9. H. G. Chen and R. T. Rockafellar, “Forward-backward splitting methods in Lagrangian optimization,” SIAM J. Optimization, submitted.