Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Piecewise Linear Solution Paths for Parametric Piecewise Quadratic Programs

Jacob Roll *

* *Division of Automatic Control, Linköping University, SE-581 83 Linköping, Sweden (e-mail: roll@isy.liu.se).*

Abstract: Recently, pathfollowing algorithms for parametric optimization problems with piecewise linear solution paths have been developed within the field of regularized regression. This paper presents a generalization of these algorithms to a wider class of problems, namely a class of parametric piecewise quadratic programs and related problems. By using pathfollowing algorithms that exploit the piecewise linearity, the entire solution paths can be very efficiently computed. Possible applications include design parameter selection for identification methods such as Direct Weight Optimization.

Keywords: Parametric programming; convex optimization; multi-objective optimization; piecewise quadratic; pathfollowing algorithm.

1. INTRODUCTION

In many applications, one encounters optimization problems involving a trade-off between two terms to optimize, i.e., problems of the type

$$\min_{x \in P} L(x) + \lambda J(x) \tag{1}$$

where λ is a design parameter controlling the trade-off, and P is the feasible region. The problem (1) is a parametric optimization problem [Guddat et al., 1990], or can also be viewed as a special case of multi-objective optimization [Boyd and Vandenberghe, 2004].

Examples of (single-)parametric optimization problems in the form (1) can be found, e.g., in the field of regularized regression. In this case, x should be interpreted as the parameters to estimate, $L(x)$ may be the ordinary least-squares cost function, and $J(x)$ represents an extra penalty on the parameters (e.g., the 1-norm as in LASSO [Tibshirani, 1996]). In Efron et al. [2004], the authors presented a new estimation method, least angle regression (LARS), and showed that the solutions to both LARS and LASSO can be efficiently computed for all values of λ simultaneously. As pointed out in [Rosset and Zhu, 2004, 2007], the key to these algorithms is that the *solution paths* (i.e., the optimal solutions x to the parametric optimization problem as a function of λ) are *piecewise linear* as λ varies from 0 to ∞. Similar results have recently also been shown for the related nn-garrote method and grouped versions of all these methods [Yuan and Lin, 2006]. In all these cases, having a single-parametric optimization problem allows for developing pathfollowing algorithms that exploit the piecewise linearity to efficiently find and represent the solution path.

Other related applications include support vector machines [Hastie et al., 2004], where the solution for different regularization penalties can be computed using parametric programming; and nonlinear system identification by Direct Weight Optimization [Roll et al., 2005], where parametric programming can be used for selection of a design parameter controlling the bias-variance trade-off [Roll, 2007a].

This paper presents a generalization of the framework of pathfollowing algorithms for piecewise linear solution paths in [Efron et al., 2004, Rosset and Zhu, 2007, Yuan and Lin, 2006], and extends the problem class to a broad class of (single-)parametric piecewise quadratic programs and related problems. It is shown that the solution paths are piecewise linear, and a pathfollowing algorithm is given. For the case of quadratic plus piecewise affine cost functions, an algorithm with explicit expressions for computation of the solution path is given.

Related work can also be found in the area of model predictive control, where in recent years results in explicit model predictive control has led to a growing interest in multiparametric linear and quadratic programming. It has been shown that the solutions to different classes of problems are piecewise affine functions of the parameters [see, e.g., Tøndel et al., 2003, Bemporad et al., 2002, Borrelli, 2003]. However, it seems that piecewise quadratic problems has only very recently begun to receive attention [Mayne et al., 2007].

The paper is organized as follows: In the following two sections, we will consider some specific classes of optimization problems of the type (1), which will be shown to have piecewise linear solution paths. Section 2 considers piecewise quadratic problems and describes a pathfollowing algorithm. In Section 3, a class of quadratic plus piecewise affine problems is considered, for which the directions of the linear parts of the solution path can be computed explicitly. Section 4 points out how the problem classes can be extended, while Section 5 gives an illustrating example.

2. PIECEWISE QUADRATIC PLUS PIECEWISE AFFINE COST FUNCTION

First, we will consider a class of optimization problems in the form (1) where $J(x)$ is piecewise affine and $L(x)$ is a piecewise quadratic function. A general piecewise affine convex function can be written as a maximum of a number of affine functions [Boyd and Vandenberghe, 2004]:

$$J(x) = \max_k \{c_k^T x + d_k\} \tag{2}$$

$L(x)$ is supposed to be strictly convex and in the form

$$L(x) = \frac{1}{2} x^T Q_i x + f_i^T x + r_i \qquad \text{if } x \in \mathcal{X}_i \tag{3}$$

where the matrices $Q_i = Q_i^T$ are positive definite, and the regions $\mathcal{X}_i$ are polyhedral and defined by $\mathcal{X}_i = \{x \,|\, \tilde{H}_i x \preccurlyeq \tilde{q}_i\}$, $i \in \mathcal{I}$ (here $\preccurlyeq$ denotes componentwise inequalities). Together, $\mathcal{X}_i$ form a partition of the x space [1]. Furthermore, we assume that for each $\lambda \geq 0$, problem (1) has a unique, finite optimal solution.

We can now show the following lemma.

Lemma 1. The problem

$$\begin{aligned} \min_x \quad & \lambda \max_k \{c_k^T x + d_k\} + L(x) \\ \text{subj. to} \quad & Ax = b \\ & \bar{A}x \preccurlyeq \bar{b} \end{aligned} \tag{4}$$

with $L(x)$ given by (3) has a piecewise linear solution path, i.e., the optimal $x \in \mathbb{R}^n$ is a piecewise affine function of $\lambda \in [0, \infty]$.

Proof: It is easy to see that the optimum of (4), which is unique and finite for given λ according to the assumptions, changes continuously with λ.

Now, we can partition the feasible set into a number of relatively open polyhedra together with a number of points (the corners of the polyhedra), denoted P_j (i.e., either $P_j = \text{relint}(P_j)$ or P_j is a single point; for the definition of relative interior, see Boyd and Vandenberghe [2004]), such that on P_j, the cost function of (1) equals

$$\lambda(c_{k_j}^T x + d_{k_j}) + \frac{1}{2} x^T Q_{i_j} x + f_{i_j}^T x + r_{i_j}$$

Let the affine hull of P_j [Boyd and Vandenberghe, 2004] be described by

$$\text{aff}(P_j) = \{x \,|\, \tilde{A}_j x = \tilde{b}_j\}$$

where $\tilde{A}_j$ is chosen such that it has full row rank.

Assume that the solution to (4) for a given λ lies in P_j. Then, since this solution is either in the relative interior of P_j or the only point of P_j, it is also the solution to

$$\begin{aligned} \min_x \quad & \lambda(c_{k_j}^T x + d_{k_j}) + \frac{1}{2} x^T Q_{i_j} x + f_{i_j}^T x + r_{i_j} \\ \text{subj. to} \quad & \tilde{A}_j x = \tilde{b}_j \end{aligned} \tag{5}$$

But the solution to this problem can be computed as

$$\begin{aligned} x = Q_{i_j}^{-1} \Big(& \big(\tilde{A}_j^T (\tilde{A}_j Q_{i_j}^{-1} \tilde{A}_j^T)^{-1} \tilde{A}_j Q_{i_j}^{-1} - I\big)(f_{i_j} + c_{k_j}\lambda) \\ & + \tilde{A}_j^T (\tilde{A}_j Q_{i_j}^{-1} \tilde{A}_j^T)^{-1} \tilde{b}_j \Big) \end{aligned} \tag{6}$$

(see Roll [2007b]). Here, x is linear in λ. This means that the solution to (4) must consist of a number of such linear pieces, one piece for every P_j that the solution path passes through. Hence, the solution path is piecewise linear. □

[1] However, for simplicity we let $\mathcal{X}_i$ be closed sets, which means that they will intersect at the boundaries.

Remark 2. The strict convexity condition for $L(x)$ can be relaxed. It is sufficient that $L(x)$ is strictly convex in a neighborhood of each point on the solution path, and convex elsewhere.

Having shown Lemma 1, let us try to derive an algorithm that computes the entire solution path. For simplicity, we assume that A has full row rank. We can rewrite the problem by introducing a slack variable s according to

$$\begin{aligned} \min_{x,s} \quad & \lambda s + L(x) \\ \text{subj. to} \quad & s \geq c_k^T x + d_k \\ & Ax = b \\ & \bar{A}x \preccurlyeq \bar{b} \end{aligned} \tag{7}$$

The Lagrangian function of (7) becomes

$$\begin{aligned} \mathcal{L}(s, x; \mu, \mu^{\bar{A}}, \mu^A) &= \lambda s + L(x) \\ &- \sum_{k=1}^m \mu_k (s - c_k^T x - d_k) - \mu^{\bar{A}^T}(\bar{b} - \bar{A}x) - \mu^{A^T}(b - Ax) \end{aligned} \tag{8}$$

where μ, $\mu^{\bar{A}}$ and μ^A are Lagrangian multipliers for the different constraints of (7).

Using a version of the *Karush-Kuhn-Tucker (KKT) conditions* [Rockafellar, 1970, Cor. 28.3.1] we can see that (x, s) is the optimal solution to (7) if and only if the following conditions are satisfied for some subgradient ν of $L(x)$:

$$\nu + \sum_{k=1}^m \mu_k c_k + \bar{A}^T \mu^{\bar{A}} + A^T \mu^A = 0 \tag{9a}$$

$$\lambda - \sum_{k=1}^m \mu_k = 0 \tag{9b}$$

$$s \geq c_k^T x + d_k \tag{9c}$$

$$\bar{A}x \preccurlyeq \bar{b} \tag{9d}$$

$$Ax = b \tag{9e}$$

$$\mu_k (s - c_k^T x - d_k) = 0 \tag{9f}$$

$$\mu_j^{\bar{A}} (\bar{b}_j - \bar{A}_j x) = 0 \tag{9g}$$

$$\mu_k \geq 0, \quad \mu^{\bar{A}} \succcurlyeq 0 \tag{9h}$$

The subgradient ν can be written as a convex combination of linear expressions:

$$\nu = \sum_{i \in \mathcal{I}^a} \alpha_i (Q_i x + f_i), \quad \text{where } \mathcal{I}^a = \{i \,|\, x \in \mathcal{X}_i\}, \tag{10}$$

and the coefficients α_i satisfy $\sum_{i \in \mathcal{I}^a} \alpha_i = 1, \quad \alpha_i \geq 0.$

The KKT conditions have a solution that is unique in (x, s), but not necessarily in $(\mu, \mu^{\bar{A}}, \mu^A)$. Hence, we should keep the number of active constraints at a minimum, to ensure that we get a unique solution.

Denote the different sets of active constraints by $\mathcal{K}^a$ (for μ_k) and $\mathcal{J}^a$ (for $\mu_j^{\bar{A}}$), and let us also introduce the following notation:

$$\mathcal{K}^a = \{k_1, k_2, \ldots, k_{n^{\mathcal{K}}}\} \qquad \mathcal{J}^a = \{j_1, j_2, \ldots, j_{n^{\mathcal{J}}}\}$$

$$\mu_{\mathcal{K}^a} = \big(\mu_{k_1} \; \ldots \; \mu_{k_{n^{\mathcal{K}}}}\big)^T \qquad \mu_{\mathcal{J}^a}^{\bar{A}} = \Big(\mu_{j_1}^{\bar{A}} \; \ldots \; \mu_{j_{n^{\mathcal{J}}}}^{\bar{A}}\Big)^T$$

$$C_{\mathcal{K}^a} = \big(c_{k_1} \; \ldots \; c_{k_{n^{\mathcal{K}}}}\big)^T \qquad \bar{A}_{\mathcal{J}^a} = \Big(\bar{A}_{j_1}^T \; \ldots \; \bar{A}_{j_{n^{\mathcal{J}}}}^T\Big)^T$$

$$d_{\mathcal{K}^a} = \begin{pmatrix} d_{k_1} & \dots & d_{k_{n^{\mathcal{K}}}} \end{pmatrix}^T \qquad \bar{b}_{\mathcal{J}^a} = \begin{pmatrix} \bar{b}_{j_1} & \dots & \bar{b}_{j_{n^{\mathcal{J}}}} \end{pmatrix}^T$$

Here, $\bar{A}_j$ denotes the jth row of $\bar{A}$, while $\mu_j^{\bar{A}}$ and $\bar{b}_j$ denote the jth element of $\mu^{\bar{A}}$ and $\bar{b}$, respectively. If we combine (9a), (9b), (9f), (9g) and (9e), to obtain the solution we then need to solve

$$\begin{pmatrix} \sum_{i \in \mathcal{I}^a} \alpha_i Q_i & 0 & C_{\mathcal{K}^a}^T & \bar{A}_{\mathcal{J}^a}^T & A^T \\ 0 & 0 & \mathbf{1}_{n^{\mathcal{K}}}^T & 0 & 0 \\ C_{\mathcal{K}^a} & \mathbf{1}_{n^{\mathcal{K}}} & 0 & 0 & 0 \\ \bar{A}_{\mathcal{J}^a} & 0 & 0 & 0 & 0 \\ A & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ -s \\ \mu_{\mathcal{K}^a} \\ \mu_{\mathcal{J}^a}^{\bar{A}} \\ \mu^A \end{pmatrix} = \begin{pmatrix} -\sum_{i \in \mathcal{I}^a} \alpha_i f_i \\ \lambda \\ -d_{\mathcal{K}^a} \\ \bar{b}_{\mathcal{J}^a} \\ b \end{pmatrix} \tag{11}$$

If the solution for the current λ is in the interior of a $\mathcal{X}_i$ region, then the α values are all zero except for $\alpha_i = 1$, and the system of equations in the remaining unknowns is linear. Furthermore, for given α it follows from [Roll, 2007b, Lemma 2] that if

$$\begin{pmatrix} C_{\mathcal{K}^a} & \mathbf{1}_{n^{\mathcal{K}}} \\ \bar{A}_{\mathcal{J}^a} & 0 \\ A & 0 \end{pmatrix} \tag{12}$$

has full row rank, then the solution to (11) is unique.

If x instead belongs to an intersection between a number of regions $\bigcap_{i \in \mathcal{I}^a} \mathcal{X}_i$, the α values are in general unknown. Then (11) is not a system of linear equations, but as we will see, it can still be handled using mostly linear techniques. Let

$$H_{\mathcal{I}^a} x = q_{\mathcal{I}^a} \tag{13}$$

be a minimal number of constraints that restrict x to aff $\left(\bigcap_{i \in \mathcal{I}^a} \mathcal{X}_i\right)$ (taking into account also the last three block rows of (11)). What we need to find is a solution to the combined problem (11) and (13).

Extend $H_{\mathcal{I}^a}$ to a square, non-singular matrix according to

$$\begin{pmatrix} H_{\mathcal{I}^a} \\ H_{\mathcal{I}^a}^{\perp} \end{pmatrix} \in \mathbb{R}^{n \times n}, \qquad H_{\mathcal{I}^a} {H_{\mathcal{I}^a}^{\perp}}^T = 0$$

Given a particular solution x^* to (13), the general solution can be written as

$$x = x^* + {H_{\mathcal{I}^a}^{\perp}}^T \beta$$

where β is arbitrary. Inserting this into the first block row of (11) and multiplying from left by $H_{\mathcal{I}^a}^{\perp}$ gives

$$\sum_{i \in \mathcal{I}^a} \alpha_i H_{\mathcal{I}^a}^{\perp} \left(Q_i (x^* + {H_{\mathcal{I}^a}^{\perp}}^T \beta) + f_i \right) \tag{14}$$
$$+ H_{\mathcal{I}^a}^{\perp} C_{\mathcal{K}^a}^T \mu_{\mathcal{K}^a} + H_{\mathcal{I}^a}^{\perp} \bar{A}_{\mathcal{J}^a}^T \mu_{\mathcal{J}^a}^{\bar{A}} + H_{\mathcal{I}^a}^{\perp} A^T \mu^A = 0$$

Now, since $L(x)$ is continuous, the gradients of all the quadratic functions with indices in $\mathcal{I}^a$ have the same component along the common boundary of the regions $\mathcal{X}_i$, $i \in \mathcal{I}^a$. Hence, the first sum of (14) is independent of α, and we can choose any index $l \in \mathcal{I}^a$ and replace the sum according to

$$\sum_{i \in \mathcal{I}^a} \alpha_i H_{\mathcal{I}^a}^{\perp} \left(Q_i (x^* + {H_{\mathcal{I}^a}^{\perp}}^T \beta) + f_i \right) = H_{\mathcal{I}^a}^{\perp} \left(Q_l (x^* + {H_{\mathcal{I}^a}^{\perp}}^T \beta) + f_l \right)$$

Hence, we can solve

$$\underbrace{\begin{pmatrix} H_{\mathcal{I}^a}^{\perp} Q_l {H_{\mathcal{I}^a}^{\perp}}^T & 0 & H_{\mathcal{I}^a}^{\perp} C_{\mathcal{K}^a}^T & H_{\mathcal{I}^a}^{\perp} \bar{A}_{\mathcal{J}^a}^T & H_{\mathcal{I}^a}^{\perp} A^T \\ 0 & 0 & \mathbf{1}_{n^{\mathcal{K}}}^T & 0 & 0 \\ C_{\mathcal{K}^a} {H_{\mathcal{I}^a}^{\perp}}^T & \mathbf{1}_{n^{\mathcal{K}}} & 0 & 0 & 0 \\ \bar{A}_{\mathcal{J}^a} {H_{\mathcal{I}^a}^{\perp}}^T & 0 & 0 & 0 & 0 \\ A {H_{\mathcal{I}^a}^{\perp}}^T & 0 & 0 & 0 & 0 \end{pmatrix}}_{\Gamma(\mathcal{I}^a, \mathcal{J}^a, \mathcal{K}^a)} \underbrace{\begin{pmatrix} \beta \\ -s \\ \mu_{\mathcal{K}^a} \\ \mu_{\mathcal{J}^a}^{\bar{A}} \\ \mu^A \end{pmatrix}}_{w(\mathcal{I}^a, \mathcal{J}^a, \mathcal{K}^a)}$$
$$= \begin{pmatrix} -H_{\mathcal{I}^a}^{\perp} (Q_l x^* + f_l) \\ \lambda \\ -d_{\mathcal{K}^a} \\ \bar{b}_{\mathcal{J}^a} \\ b \end{pmatrix} \tag{15}$$

which, due to the minimality of (13), is still uniquely solvable. The solution can then be inserted into the first block row of (11) to solve for α.

To compute what happens for a small change in λ, we solve

$$\Gamma(\mathcal{I}^a, \mathcal{J}^a, \mathcal{K}^a) \frac{\partial w(\mathcal{I}^a, \mathcal{J}^a, \mathcal{K}^a)}{\partial \lambda} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{16}$$

To start the algorithm, we first have to compute the optimal solution for $\lambda = 0$ (let us denote this solution by x_0, and the solution for any given λ by x_λ). This is an ordinary convex optimization problem. After this, we select a maximal subset of the constraints that are active at x_0, such that (12) has full row rank. (However, note that in $\mathcal{K}^a$, it is sufficient to include only one index k to start with.)

The algorithm can now be described as follows:

Algorithm 1. Given: A parametric optimization problem of the type (4).

(1) Set $\lambda = 0$.
(2) Compute the solution x_0 to (4).
(3) Let $S = \{(\lambda, x_\lambda)\} = \{(0, x_0)\}$.
(4) Let $\mathcal{K}^a = \{k\}$ for some k in $\arg\max_k \{c_k^T x_\lambda + d_k\}$, and let $\mathcal{J}^a$, $\mathcal{I}^a$ be maximal subsets of indices for which corresponding constraints are active at x_0, such that (12) has full row rank. Compute $H_{\mathcal{I}^a}$ and $H_{\mathcal{I}^a}^{\perp}$. (If $\mathcal{I}^a$ has only one member, $H_{\mathcal{I}^a}$ will be empty, and we can let $H_{\mathcal{I}^a}^{\perp} = I$.)
(5) Compute (15) to get s, $\mu_{\mathcal{K}^a}$, $\mu_{\mathcal{J}^a}^{\bar{A}}$, and μ^A.
(6) Compute the directions given by (16).
(7) Find the minimal $\delta\lambda \geq 0$ such that one of the following conditions are satisfied:
    (a) $s + \frac{\partial s}{\partial \lambda} \delta\lambda = c_k^T (x_\lambda + \frac{\partial x}{\partial \lambda} \delta\lambda) + d_k$ and $\frac{\partial s}{\partial \lambda} < c_k^T \frac{\partial x}{\partial \lambda}$ for some $k \notin \mathcal{K}^a$. Then move the corresponding k to $\mathcal{K}^a$.
    (b) $\mu_k + \frac{\partial \mu_k}{\partial \lambda} \delta\lambda = 0$ and $\frac{\partial \mu_k}{\partial \lambda} < 0$ for some $k \in \mathcal{K}^a$. Then remove the corresponding k from $\mathcal{K}^a$.
    (c) $\bar{A}_j (x_\lambda + \frac{\partial x}{\partial \lambda} \delta\lambda) = \bar{b}_j$ and $\bar{A}_j \frac{\partial x}{\partial \lambda} > 0$ for some $j \notin \mathcal{J}^a$. Then move the corresponding j to $\mathcal{J}^a$.
    (d) $\mu_j^{\bar{A}} + \frac{\partial \mu_j^{\bar{A}}}{\partial \lambda} \delta\lambda = 0$ and $\frac{\partial \mu_j^{\bar{A}}}{\partial \lambda} < 0$ for some $j \in \mathcal{J}^a$. Then remove the corresponding j from $\mathcal{J}^a$.
    (e) $x_\lambda + \frac{\partial x}{\partial \lambda} \delta\lambda \in \mathcal{X}_l$ for some $l \notin \mathcal{I}^a$, and $x_\lambda + \frac{\partial x}{\partial \lambda} (\delta\lambda + \varepsilon) \notin \mathcal{X}_i$ for some $\varepsilon > 0$ and $i \in \mathcal{I}^a$. Then include l in $\mathcal{I}^a$.
    (f) There would not be a subgradient ν satisfying (10) (with x replaced by $x_\lambda + \frac{\partial x}{\partial \lambda} \delta\lambda$) if increasing

$\delta\lambda$ further. Then remove one i with corresponding $\alpha_i = 0$ from $\mathcal{I}^a$.

Add $\delta\lambda$ to λ, update x_λ, s, $\mu_{\mathcal{K}^a}$, $\mu^{\bar{A}}_{\mathcal{J}^a}$, and μ^A, and recompute $H_{\mathcal{I}^a}$ and $H^\perp_{\mathcal{I}^a}$. If there is no $\delta\lambda \geq 0$ for which the conditions are satisfied, set $\lambda = \infty$.

(8) Add the new pair (λ, x_λ) to S; $S := \{S, (\lambda, x_\lambda)\}$.

(9) If $\lambda = \infty$, stop. Otherwise, go to step 6.

When the algorithm has finished, S will contain the knots of the piecewise linear solution path, and the solution to problem (4) for any arbitrary λ can be obtained by linear interpolation between two neighboring knots.

Note that only linear techniques are needed to find the solution path, except for step 7f, where α_i may be a nonlinear function of λ. This step can be handled by simple line search. It can also be shown that all the changes of step 7 will lead to full row rank of (12).

To handle cases where $x_\lambda \to \infty$ when $\lambda \to \infty$, we only need a small modification of the algorithm. When λ is set to ∞, we should store $\frac{\partial x}{\partial \lambda}$ instead of x_∞. This allows for finding the solutions x_λ by extrapolation if λ is large.

3. QUADRATIC PLUS PIECEWISE AFFINE COST FUNCTION

Now let us consider a class of optimization problems where simple explicit expressions can be given for the directions of the linear parts of the solution paths. The problems considered will have a positive definite quadratic $L(x)$ and a $J(x)$ which is a sum of absolute values of affine functions. We will also allow linear equality constraints. In other words, the problems we will consider can be written as

$$\begin{aligned} \min_z \quad & \lambda \sum_{k=1}^m |h_k^T z + g_k| + \frac{1}{2} z^T Q z + p^T z \\ \text{subj. to} \quad & Az = b \end{aligned} \tag{17}$$

where $Q = Q^T$ is positive definite. Provided that the problem is feasible, we can use the constraints to eliminate variables, to get an equivalent problem in the form

$$\min_{\tilde{z}} \quad \lambda \sum_{k=1}^m |\tilde{h}_k^T \tilde{z} + \tilde{g}_k| + \frac{1}{2} \tilde{z}^T \tilde{Q} \tilde{z} + \tilde{p}^T \tilde{z} \tag{18}$$

where the elements of $\tilde{z}$ are a subset of the elements in z. We can now make the variable substitution $x = \tilde{Q}^{1/2}\tilde{z} + \tilde{Q}^{-1/2}\tilde{p}$. Hence, it turns out that it is sufficient to consider problems of the type

$$\min_x \lambda \sum_{k=1}^m |c_k^T x + d_k| + \frac{1}{2} x^T x \tag{19}$$

in more detail. This is a special case of (4), and hence (19) has a piecewise linear solution path.

We will now derive explicit expressions for the directions of the linear parts of the solution path. For simplicity, we will assume that the problem is non-degenerate in the sense that for all x, the vectors of the set $\{c_k | c_k^T x + d_k = 0\}$ are linearly independent.

Introducing slack variables s_k, we can rewrite (19) as

$$\begin{aligned} \min_{x,s} \quad & \lambda \sum_{k=1}^m s_k + \frac{1}{2} x^T x \\ \text{subj. to} \quad & s_k \geq c_k^T x + d_k \\ & s_k \geq -c_k^T x - d_k \end{aligned} \tag{20}$$

The Lagrangian function of (20) becomes

$$\begin{aligned} \mathcal{L}(s, x; \mu^+, \mu^-) = {} & \lambda \sum_{k=1}^m s_k + \frac{1}{2} x^T x \\ & - \sum_{k=1}^m \mu_k^+ (s_k - c_k^T x - d_k) - \sum_{k=1}^m \mu_k^- (s_k + c_k^T x + d_k) \end{aligned}$$

with μ^+ and μ^- being Lagrangian multipliers for the different constraints of (20). The KKT conditions for the problem are as follows:

$$x + \sum_{k=1}^m c_k \mu_k^+ - \sum_{k=1}^m c_k \mu_k^- = 0 \tag{21a}$$

$$\lambda - \mu_k^+ - \mu_k^- = 0 \tag{21b}$$

$$\mu_k^+ (s_k - c_k^T x - d_k) = 0 \tag{21c}$$

$$\mu_k^- (s_k + c_k^T x + d_k) = 0 \tag{21d}$$

$$\mu_k^\pm \geq 0 \tag{21e}$$

Define the sets

$$\begin{aligned} K^+ &= \{k : c_k^T x + d_k > 0\} \\ K^- &= \{k : c_k^T x + d_k < 0\} \\ K^0 &= \{k : c_k^T x + d_k = 0\} \end{aligned} \tag{22}$$

For $k \in K^+$, we get $\mu_k^+ = \lambda$ and $\mu_k^- = 0$ from (21b) and (21d). Similarly, for $k \in K^-$ we obtain $\mu_k^+ = 0$ and $\mu_k^- = \lambda$. Using (21a) this implies that

$$x + \lambda \sum_{k \in K^+} c_k - \lambda \sum_{k \in K^-} c_k + \sum_{k \in K^0} c_k (2\mu_k^+ - \lambda) = 0 \tag{23}$$

Since we would like to consider the linear parts of the solution path, we can assume $0 < \mu_k^+ < \lambda$ in the last sum. To compute the effect of a small change in λ, we can use (23) to get

$$\frac{\partial x}{\partial \lambda} + \sum_{k \in K^+} c_k - \sum_{k \in K^-} c_k + \sum_{k \in K^0} c_k \left(2\frac{\partial \mu_k^+}{\partial \lambda} - 1\right) = 0 \tag{24}$$

If we introduce the notation $K^0 = \{k_1, \ldots, k_{n^0}\}$ and

$$C^0 = \begin{pmatrix} c_{k_1} & \ldots & c_{k_{n^0}} \end{pmatrix}^T, \quad M^+ = \begin{pmatrix} \mu_{k_1}^+ & \ldots & \mu_{k_{n^0}}^+ \end{pmatrix}^T \tag{25}$$

and similarly for C^+ and C^-, we can write (24) as

$$\begin{aligned} & \frac{\partial x}{\partial \lambda} + \left({C^+}^T \mathbf{1}_{n^+} - {C^-}^T \mathbf{1}_{n^-} \right) \\ & \quad + {C^0}^T \left(2\frac{\partial M^+}{\partial \lambda} - \mathbf{1}_{n^0} \right) = 0 \end{aligned} \tag{26}$$

At the same time, for $j \in K^0$ it must hold that

$$c_j^T \frac{\partial x}{\partial \lambda} = 0$$

Hence, multiplying (26) by C^0 yields

$$\begin{aligned} & C^0 \left({C^+}^T \mathbf{1}_{n^+} - {C^-}^T \mathbf{1}_{n^-} \right) \\ & \quad + C^0 {C^0}^T \left(2\frac{\partial M^+}{\partial \lambda} - \mathbf{1}_{n^0} \right) = 0 \end{aligned} \tag{27}$$

Due to the non-degeneracy assumption, $C^0 {C^0}^T$ is invertible, and we get

$$\frac{\partial M^+}{\partial \lambda} = \tag{28}$$
$$\frac{1}{2}\left(C^0 {C^0}^T\right)^{-1} C^0 \left(-{C^+}^T \mathbf{1}_{n^+} + {C^-}^T \mathbf{1}_{n^-} + {C^0}^T \mathbf{1}_{n^0}\right)$$

Inserting this into (26) results in

$$\frac{\partial x}{\partial \lambda} = \left(I - {C^0}^T \left(C^0 {C^0}^T\right)^{-1} C^0\right) \cdot \left(-{C^+}^T \mathbf{1}_{n^+} + {C^-}^T \mathbf{1}_{n^-} + {C^0}^T \mathbf{1}_{n^0}\right) \tag{29}$$

Note that since this expression is locally constant, the solution x will locally change linearly as λ changes. Just as for the problem class considered in Section 2, this means that when computing the solution path, we only need to store the solutions and values of λ for the knots of the solution path, as the values in between can be obtained afterwards by simple linear interpolation.

We can now give an algorithm for finding the solution path to a problem in the form (19).

Algorithm 2. Given: A problem of the type (19).

(1) Set $\lambda = 0$, $x_\lambda = 0$, $\mu^+ = 0$ and $S = \{(\lambda, x_\lambda)\}$.
(2) Compute the sets K^+, K^- and K^0, as defined in (22).
(3) Compute the directions given by (28) and (29).
(4) Find the minimal $\delta\lambda \geq 0$ such that one of the following conditions are satisfied:
  (a) $c_k^T(x_\lambda + \frac{\partial x}{\partial \lambda}\delta\lambda) + d_k = 0$ and $c^T \frac{\partial x}{\partial \lambda} < 0$ for some $k \in K^+$. Then move k from K^+ to K^0.
  (b) $c_k^T(x_\lambda + \frac{\partial x}{\partial \lambda}\delta\lambda) + d_k = 0$ and $c^T \frac{\partial x}{\partial \lambda} > 0$ for some $k \in K^-$. Then move k from K^- to K^0.
  (c) $\mu_k^+ + \frac{\partial \mu_k^+}{\partial \lambda}\delta\lambda = \lambda$ and $\frac{\partial \mu_k^+}{\partial \lambda} > 0$ for some $k \in K^0$. Then move k from K^0 to K^+.
  (d) $\mu_k^+ + \frac{\partial \mu_k^+}{\partial \lambda}\delta\lambda = 0$ and $\frac{\partial \mu_k^+}{\partial \lambda} < 0$ for some $k \in K^0$. Then move k from K^0 to K^-.

  Add $\delta\lambda$ to λ. If there is no $\delta\lambda \geq 0$ for which the conditions are satisfied, set $\lambda = \infty$.
(5) Add the new pair (λ, x_λ) to S; $S := \{S, (\lambda, x_\lambda)\}$.
(6) If $\lambda = \infty$, stop. Otherwise, go to step 3.

4. RELATED SOLUTION PATHS FOR DIFFERENT PROBLEMS

Apart for the classes described in the previous sections, there are several other problem classes that have piecewise linear solution paths. In fact, starting from one problem, we can derive a family of problems having the same solution path. This can be seen from the following observation (cf. Boyd and Vandenberghe [2004, Exercise 4.51]).

Observation 1. Suppose that $L : \mathcal{D}(L) \subseteq \mathbb{R}^n \to \mathcal{R}(L) \subseteq \mathbb{R}$ and $J : \mathcal{D}(L) \to \mathcal{R}(J) \subseteq \mathbb{R}$ are convex functions defined on the same convex domain $\mathcal{D}(L)$, that

$$\min_x L(x) + \lambda J(x) \tag{30}$$

has a well-defined solution path [2] for $\lambda \in [0, \infty]$, and that $f_1 : \mathcal{R}(L) \to \mathbb{R}$ and $f_2 : \mathcal{R}(J) \to \mathbb{R}$ are strictly increasing functions. Then the solution path of

$$\min_x f_1(L(x)) + \eta f_2(J(x)) \tag{31}$$

for $\eta \in [0, \infty]$ is a subset of the solution path of (30).

Proof: See [Roll, 2007b]. □

Remark 3. If $f_1(L(x))$ and $f_2(J(x))$ are convex, the solution paths are identical. This can be seen by applying Observation 1 to $f_1(L(x))$, $f_2(J(x))$, f_1^{-1}, and f_2^{-1}.

Remark 4. If we do not assume convexity of $L(x)$ and $J(x)$, Observation 1 does not necessarily hold (for a counterexample, see [Roll, 2007b]).

Where f_1, f_2, J and L are differentiable, the relationship between η and λ can be established as follows (for simplicity, we only consider the case of Remark 3): For an optimal point x_η on the solution path, for some λ it holds that

$$\begin{aligned}\nabla L(x_\eta) + \lambda \nabla J(x_\eta) &= 0 \\ &= f_1'(L(x_\eta))\nabla L(x_\eta) + \eta f_2'(J(x_\eta))\nabla J(x_\eta)\end{aligned}$$

If $\nabla J(x_\eta) \neq 0$, this yields

$$\eta = \lambda \frac{f_1'(L(x_\eta))}{f_2'(J(x_\eta))} \tag{32}$$

If $\nabla J(x_\eta) = 0$, then $\nabla L(x_\eta) = 0$, and x_η will be a minimum point for all λ and η.

5. EXAMPLE

To illustrate Algorithm 1, let us study a simple example. Consider the parametric optimization problem

$$\min_x L(x) + \lambda J(x) \tag{33a}$$

where

$$L(x) = \begin{cases} \frac{1}{2}x^T x + \begin{pmatrix}0 & 2\end{pmatrix} x & \text{if } \begin{pmatrix}0 & 1\\ 1 & 2\end{pmatrix} x \leq \begin{pmatrix}0\\0\end{pmatrix} \\ \frac{1}{2}x^T x + \begin{pmatrix}2 & 3\end{pmatrix} x & \text{if } \begin{pmatrix}-1 & 1\\ -1 & -2\end{pmatrix} x \leq \begin{pmatrix}0\\0\end{pmatrix} \\ \frac{1}{4}x^T \begin{pmatrix}2 & -1\\ -1 & 4\end{pmatrix} x + \begin{pmatrix}0 & 5\end{pmatrix} x & \text{if } \begin{pmatrix}0 & -1\\ 1 & -1\end{pmatrix} x \leq \begin{pmatrix}0\\0\end{pmatrix} \end{cases} \tag{33b}$$

and

$$J(x) = \max\{(-1\ \ -1)\, x,\ (2\ \ -1)\, x - 3,\ \frac{1}{2}(1\ \ 1)\, x - \frac{9}{2}\} \tag{33c}$$

Applying Algorithm 1 to this problem will give a solution path with knots given by Table 1. The solution path is also plotted in Figure 1, where it is apparent that it moves from the minimum of $L(x)$ ($x = \begin{pmatrix}0\\-2\end{pmatrix}$, for $\lambda = 0$) to a minimum of $J(x)$ ($x = \begin{pmatrix}1\\2\end{pmatrix}$, for $\lambda \geq \frac{17}{2}$). From Table 1, we can see that we reach the boundary between the regions of two different quadratic functions when $\lambda = \frac{2}{3}$, and when $\lambda = 4$. To know for how long the solution should move along this boundary, we must compute the α values of (10) using the first block row of (11). The resulting expressions are shown in Table 2. Note that, in this example, the expressions become linear, which means that we can find the entire solution path using only linear techniques.

6. CONCLUSIONS

This paper has extended the use of piecewise linear solution paths suggested for LARS and LASSO in [Efron et al.,

[2] By well-defined solution path we here mean that for all values of λ, there is at least one minimum to the problem, and that all minimum points are finite.

λ	x_λ^T	s	μ^T	$\mathcal{K}^a$	$\mathcal{I}^a$
0	$\begin{pmatrix}0 & -2\end{pmatrix}$	2	$\begin{pmatrix}0 & 0 & 0\end{pmatrix}$	$\{1\}$	$\{1\}$
$\frac{2}{3}$	$\begin{pmatrix}\frac{2}{3} & -\frac{4}{3}\end{pmatrix}$	$\frac{2}{3}$	$\begin{pmatrix}\frac{2}{3} & 0 & 0\end{pmatrix}$	$\{1\}$	$\{1,2\}$
$\frac{7}{3}$	$\begin{pmatrix}\frac{1}{3} & -\frac{2}{3}\end{pmatrix}$	$\frac{1}{3}$	$\begin{pmatrix}\frac{7}{3} & 0 & 0\end{pmatrix}$	$\{1\}$	$\{2\}$
3	$\begin{pmatrix}1 & 0\end{pmatrix}$	-1	$\begin{pmatrix}3 & 0 & 0\end{pmatrix}$	$\{1,2\}$	$\{2\}$
4	$\begin{pmatrix}1 & 1\end{pmatrix}$	-2	$\begin{pmatrix}\frac{11}{3} & \frac{1}{3} & 0\end{pmatrix}$	$\{1,2\}$	$\{2,3\}$
$\frac{13}{2}$	$\begin{pmatrix}1 & 1\end{pmatrix}$	-2	$\begin{pmatrix}\frac{9}{2} & 2 & 0\end{pmatrix}$	$\{1,2\}$	$\{3\}$
$\frac{17}{2}$	$\begin{pmatrix}1 & 2\end{pmatrix}$	-3	$\begin{pmatrix}\frac{17}{3} & \frac{17}{6} & 0\end{pmatrix}$	$\{1,2,3\}$	$\{3\}$
∞	$\begin{pmatrix}1 & 2\end{pmatrix}$	-3	$\begin{pmatrix}\infty & \frac{17}{6} & \infty\end{pmatrix}$	$\{1,2,3\}$	$\{3\}$

Table 1. Intermediate results of Algorithm 1 for problem (33).

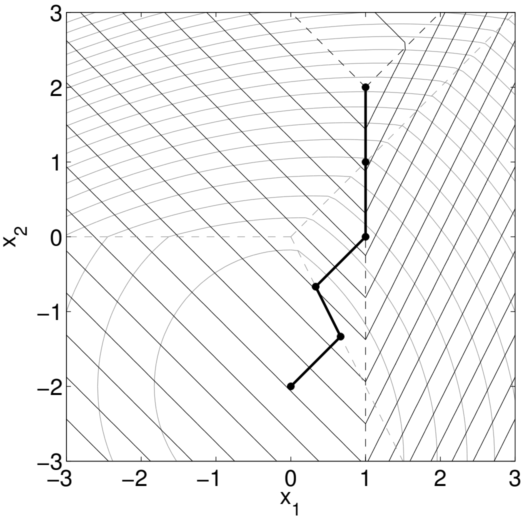

Fig. 1. Solution path (thick line) for problem (33) of Section 5, together with contour plots of $L(x)$ (gray) and $J(x)$ (black).

λ	α^T
$\frac{2}{3}+\delta\lambda$	$\begin{pmatrix}1-\frac{3}{5}\delta\lambda, & \frac{3}{5}\delta\lambda, & 0\end{pmatrix}$
$4+\delta\lambda$	$\begin{pmatrix}0, & 1-\frac{2}{5}\delta\lambda, & \frac{2}{5}\delta\lambda\end{pmatrix}$

Table 2. Checking the α values.

2004] to a more general setting of piecewise quadratic functions. The benefit of exploiting the piecewise linear solution paths is that we can efficiently compute all solutions to a parametric optimization problem, which is important in many applications, including LARS and LASSO [Efron et al., 2004] and variants using Huber norm [Rosset and Zhu, 2007], nn-garrote [Yuan and Lin, 2006], regularization in support vector machines [Hastie et al., 2004], and design parameter selection in Direct Weight Optimization [Roll, 2007a].

A topic for further studies is to enhance the computational complexity of Algorithms 1 and 2 by taking advantage of the specific problem structures. For instance, to further increase the efficiency of Algorithm 2, one could consider to compute the expressions (28) and (29) recursively, similarly to what is done in, for instance, the RLS algorithm. Also numerical issues should be studied in more detail. Furthermore, it would be interesting to compute the solutions in the other direction, i.e., for λ starting at ∞ and decreasing to 0. These could all be topics for further research.

REFERENCES

Alberto Bemporad, Manfred Morari, Vivek Dua, and Efstratios N. Pistikopoulos. The explicit linear quadratic regulator for constrained systems. *Automatica*, 38(1): 3–20, January 2002.

Francesco Borrelli. *Constrained Optimal Control of Linear and Hybrid Systems*, volume 290 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, 2003.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):407–499, 2004.

J. Guddat, F. Guerra Vasquez, and H. Th. Jongen. *Parametric Optimization: Singularities, Pathfollowing and Jumps*. John Wiley & Sons, 1990.

Trevor Hastie, Saharon Rosset, Robert Tibshirani, and Ji Zhu. The entire regularization path for the support vector machine. *Journal of Machine Learning Research*, 5:1391–1415, October 2004.

D. Q. Mayne, S. V. Raković, and E. C. Kerrigan. Optimal control and piecewise parametric programming. In *European Control Conference*, pages 2762–2767, Kos, Greece, July 2007.

R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.

Jacob Roll. Piecewise linear solution paths with application to direct weight optimization. Provisionally accepted for Automatica, 2007a.

Jacob Roll. Piecewise linear solution paths for parametric piecewise quadratic programs with application to direct weight optimization. Technical Report LiTH-ISY-R-2816, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, August 2007b.

Jacob Roll, Alexander Nazin, and Lennart Ljung. Nonlinear system identification via direct weight optimization. *Automatica*, 41(3):475–490, March 2005.

Saharon Rosset and Ji Zhu. Least angle regression - discussion. *The Annals of Statistics*, 32(2):469–475, April 2004.

Saharon Rosset and Ji Zhu. Piecewise linear regularized solution paths. *The Annals of Statistics*, 35(3), June 2007.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

Petter Tøndel, Tor Arne Johansen, and Alberto Bemporad. An algorithm for multi-parametric quadratic programming and explicit MPC solutions. *Automatica*, 39(3):489–497, March 2003.

Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 68 (1):49–67, 2006.