Quadratic Programming

Note Title 5/15/2022

As mentioned, QPs are important in their own right, and also arise as subproblems in methods for general constrained optimization.

A general QP:

$$\min_x q(x) = \frac{1}{2}x^TGx + c^Tx \quad \text{s.t.} \quad \begin{array}{ll} a_i^Tx = b_i & i \in \mathcal{E} \\ a_i^Tx \geq b_i & i \in \mathcal{I}. \end{array} \qquad \text{(QP)}$$

WLOG, we can always assume G is symmetric.

Recall: $S = \{x \in \mathbb{R}^n : a_i^Tx = b_i \;\; i \in \mathcal{E}, \;\; a_i^Tx \geq b_i \;\; i \in \mathcal{I}\}$ is always a convex polyhedron.

$q(x) = \frac{1}{2}x^TGx + c^Tx$ is a convex function on $\mathbb{R}^n \iff G \succeq 0$.

So (QP) is a convex optimization problem $\iff G \succeq 0$.

When G is positive semidefinite, we call the QP a convex QP. This is the case we shall focus primarily on.

(Strictly convex QPs are those in which $G \succ 0$.)

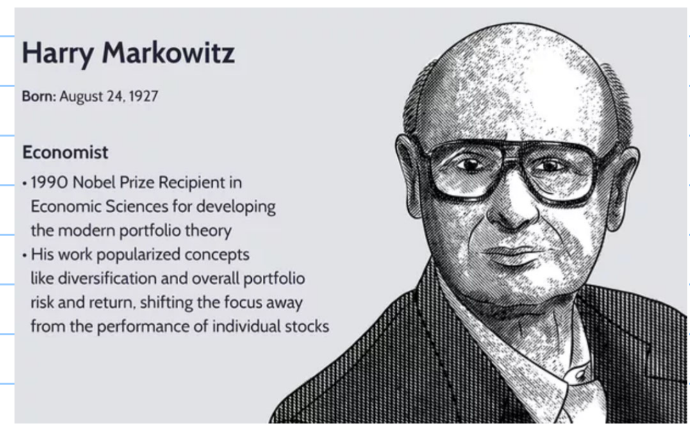
Recall some lovely facts about convex problems :

In CO_3 : x^* solves (QP) $\overset{\text{any } G}{\Longrightarrow}$ $\underset{G \succcurlyeq 0}{\Longleftarrow}$ x^* satisfies KKT conditions

In CO_4 : no duality gap when $G \succcurlyeq 0$.

(Just like duality theory gives rise to the successful primal-dual interior point methods for solving LPs, the same happens to QPs.)

Earlier we showed that the max. margin linear classification problem can be reformulated as a convex QP. Another well-known application is portfolio optimization in finance :



There are n available investment products with (annualized %age) returns $R_1, \cdots, R_n$.

How to optimally diversify one's investment so as to maximize return and minimize risk?

Let's make the following bold assumption:

(i) It is reasonable to think of $(R_1, \cdots, R_n)$ as what mathematicians call a jointly distributed random vector.

So $(R_1, \cdots, R_n)$ has a joint distribution (albeit unknown.)

Furthermore, assume $\underbrace{E[R_i]}_{=\mu_i}$ and $\underbrace{COV(R_i, R_j)}_{=E[(R_i-\mu_i)(R_j-\mu_j)] =: \sigma_{ij}}$ $i, j = 1, \cdots, n$ are finite.

(ii) Assume $E[R_i]$ and $COV(R_i, R_j)$ $i, j = 1, \cdots, n$ can be estimated from historical data.

Recall: $CORR(R_i, R_j) = COV(R_i, R_j) / \sqrt{VAR(R_i)}\sqrt{VAR(R_j)} \overset{\text{essentially Cauchy-Schwartz}}{\in} [-1, 1]$.

Interpretation: If $CORR(R_i, R_j) > 0$ $(i \neq j)$, then the two investments' returns tend to rise and fall together. (E.g. two specific stocks in the same sector.)

If $CORR(R_i, R_j) < 0$, the two investments' returns tend to move in opposite directions (E.g. stocks and bonds)

Assume you have perfect knowledge of $E[R_i]$ and $COV(R_i, R_j)$.

Assume also you have some capital to invest (1M USD, say).

An investor constructs a portfolio by putting a fraction x_i of the capital into investment product i, $i=1,\cdots,n$.

Assume that all capitals are invested, and no short-selling is allowed, then

$$\sum_{i=1}^{n} x_i = 1 \quad , \quad x_i \geq 0.$$

The return of the portfolio is $R = \sum x_i R_i$, with mean $E[R] = \sum_{i=1}^{n} x_i \mu_i = x^T \mu$.
while the variance is given by

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

$$\begin{aligned} VAR[R] = E[(R-E[R])^2] &= E[(\sum x_i (R_i - \mu_i))^2] \\ &= E[\sum_i \sum_j x_i x_j (R_i - \mu_i)(R_j - \mu_j)] \\ &= \sum_i \sum_j x_i x_j \underbrace{E[(R_i - \mu_i)(R_j - \mu_j)]}_{COV(x_i, x_j) = G_{ij}} = x^T G x. \end{aligned}$$

One would like $E[R]$ to be large and $VAR[R]$ to be small.
There is usually a tradeoff between the two.

Depending on an investor's 'risk appetite', he may aim for a higher or lower target expected return r_0 (say $r_0 = 10\%$ or 15%), while always aiming to minimize risk.
quantified by $VAR[R]$

So he aims to solve:

$$\min_x \; x^T G x \quad \text{s.t.} \quad \begin{aligned} \mu^T x &= r_0 \\ e^T x &= 1 \\ x &\geq 0 . \end{aligned} \qquad e = [1, \cdots, 1]^T$$

It is a convex QP.

Recall that if there are no inequality constraints, a QP amounts to solving a linear system — see below.

If shorting is allowed (and an investor does not mind that), the '$x \geq 0$' constraint above can be dispensed with. In this case, the math is simpler and the investor may make more money with a 'long-short portfolio'.

Properties of Equality-Constrained QP

Assume $\mathcal{I} = \phi$, the QP written in matrix form is:

$$\min_x q(x) := \frac{1}{2} x^T G x + c^T x \quad \text{s.t.} \quad \underset{\substack{\uparrow \\ m\times n,\ m\le n}}{A} x = b$$

rows of A denoted by a_i^T, $i \in \mathcal{E}$

Assume the rows are linearly independent. ($\Leftrightarrow$ rank $A = m$)

The (unconstrained) critical point of q is at $x^* = -G^{-1}c$.

$$q|_S(t) = \frac{1}{2}(x^0 + Zt)^T G (x^0 + Zt) + c^T x$$

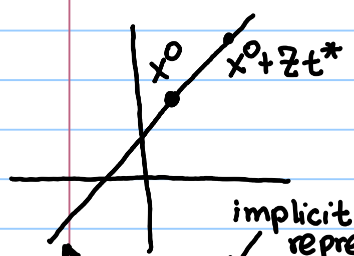
(restricting a polynomial to an affine subspace is still a poly. of the same degree.)

$$= \frac{1}{2} t^T Z^T G Z t + \tilde{c}^T t + (\text{constant})$$

$$= \frac{1}{2}(t - t^*)^T (Z^T G Z)(t - t^*) + (\text{constant})' \leftarrow$$

completing the square

$t^* = (Z^T G Z)^{-1} \tilde{c}$



So there is a unique minimizer if $\underbrace{Z^T G Z}_{\substack{\uparrow \\ \text{depends only on } G \text{ and } A \\ \text{(independent of } c \text{ and } b)}} \succ 0$.

implicit representation

$$S := \{x : Ax = b\} = \{x^0 + Zt : t \in \mathbb{R}^{n-m}\}$$

$AZ = 0$

parametric/explicit representation

KKT conditions

(*) $$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}.$$

Even when rank $A < m$, it is necessary for optimality, and is also sufficient for optimality when $A \succeq 0$.

In an iterative setting, express x^* as $x^* = x + p$ ← desired step,

↑ some estimate of the solution

the linear system above becomes

$$\underbrace{\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix}}_{} \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = \begin{bmatrix} c + Gx \\ Ax - b \end{bmatrix}$$

called the KKT matrix (it is symmetric)

The argument given in the previous page shows:

Thm: Let A have full rank (m), $AZ = 0$, rank $Z = n-m$. If $Z^T G Z \succ 0$, then

(i) the KKT matrix is non-singular

(ii) the (unique) solution of (*) is the unique global solution of the QP.

If $Z^T G Z$ is positive semidefinite with zero eigenvalues, a vector x^* satisfying

(*) is a local minimizer but not a strict local minimizer.

If Z^TGZ has negative eigenvalues, then x^* is only a stationary point, not a local minimizer.

E.g. $G = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, $A = [1, 0]$, $c = 0$, $b = 0$

$\min \frac{1}{2}(x_1^2 - x_2^2)$ st $x_1 = 0$ $\longleftrightarrow$ $\min -\frac{1}{2}x_2^2$ no minimizer

(0,0) is a stationary pt, in fact a maximizer

$Z = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $Z^TGZ = [-1]$

same G, c, b, but $A = [0, 1]$

$\min \frac{1}{2}(x_1^2 - x_2^2)$ st. $x_2 = 0$ $\longleftrightarrow$ $\min \frac{1}{2}x_1^2$ minimizer at (0,0)

$Z = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $Z^TAZ = [1]$.

In this case, the KKT matrix is $\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ ← eigenvalues: 1, 0.618, −1.618

What if $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$? $\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ ← eigenvalues: 1, 1.618, −0.618

An unpleasant surprise:

Even when the reduced Hessian is positive definite, the KKT matrix is always indefinite.

For any symmetric matrix M, write $\text{inertia}(M) = \underbrace{(n_+, n_-, n_0)}_{\substack{\#\text{ of positive, negative, and zero} \\ \text{eigenvalues of } M\text{, respectively}}}$

Thm: If $A \in \mathbb{R}^{m\times n}$ $(m \le n)$, $AZ = 0$ (rank $Z = n-m$), $G \in \mathbb{R}^{n\times n}$ $(G^T = G)$, $K = \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix}$ then

$$\text{inertia}(K) = \text{inertia}(Z^T G Z) + (m, m, 0).$$

In particular, if $Z^T G Z \succ 0$, then $\text{inertia}(K) = (n, m, 0)$.

i.e. # of negative eigenvalues = # of constraints

Because of this situation, standard methods for solving symmetric positive definite linear systems, such as Cholesky factorization or conjugate gradient method, cannot be directly applied to solve the KKT system.

See N&W Sec 16.2 and 16.3 for the interesting numerical linear algebra for solving the indefinite KKT system.

From now on, we assume/pretend we know how to solve QPs with only equality constraints.

QP with inequality constraints

We shall discuss active-set and interior point methods for solving convex QPs.

Basic theory for $\min \frac{1}{2}x^TQx + c^Tx$ st. $a_i^Tx - b_i = 0 \quad i \in \mathcal{E}$, $a_i^Tx - b_i \geq 0 \quad i \in \mathcal{I}$

$$\mathcal{L}(x,\lambda) = \frac{1}{2}x^TGx + c^Tx - \sum_{i \in \mathcal{E}\cup\mathcal{I}} \lambda_i(a_i^Tx - b_i)$$

KKT conditions : $\exists x^* \in \mathbb{R}^n$ st. for some λ_i^*, $i \in \mathcal{A}(x^*) := \{i \in \mathcal{E}\cup\mathcal{I} \mid a_i^Tx^* = b_i\}$

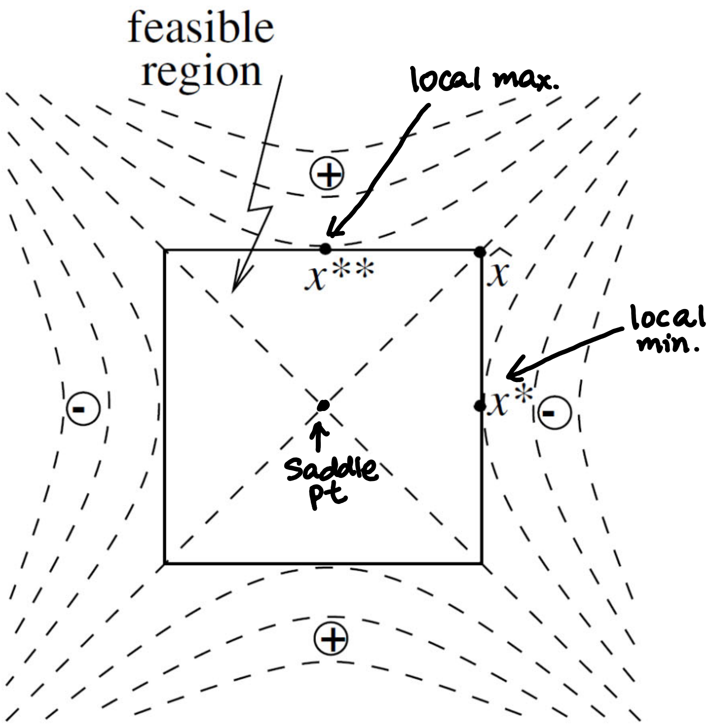
$$Gx^* + c - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* a_i = 0$$

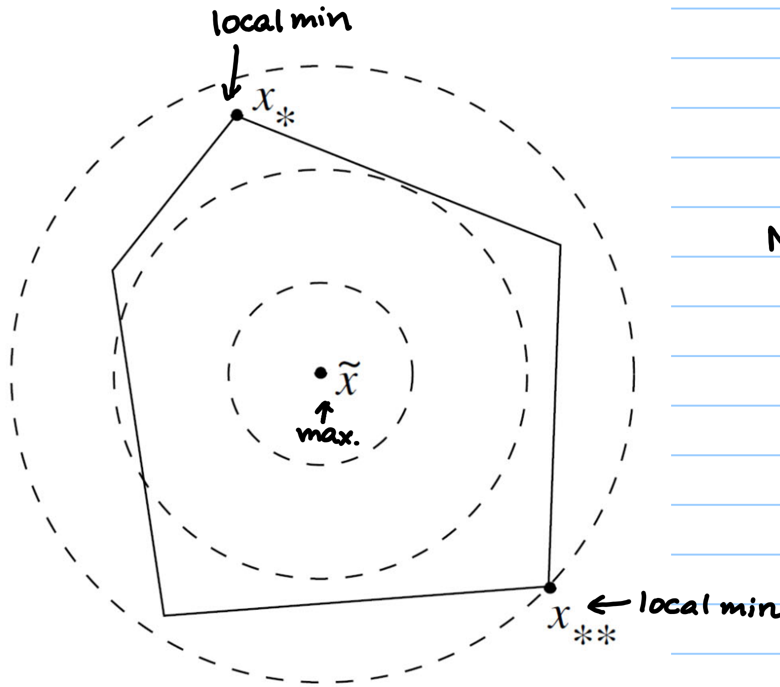
$$\begin{aligned} a_i^Tx^* &= b_i & i &\in \mathcal{A}(x^*) \\ a_i^Tx^* &> b_i & i &\in \mathcal{I}\setminus\mathcal{A}(x^*) \\ \lambda_i^* &\geq 0 & i &\in \mathcal{I}\cap\mathcal{A}(x^*) \end{aligned}$$

Since all constraints are linear, these conditions are still necessary for optimality even if $a_i : i \in \mathcal{A}(x^*)$ are linearly dependent.

Recall: If $G \succeq 0$, then KKT conditions $\Leftrightarrow$ x^* is a global minimizer of the QP.

Some examples:

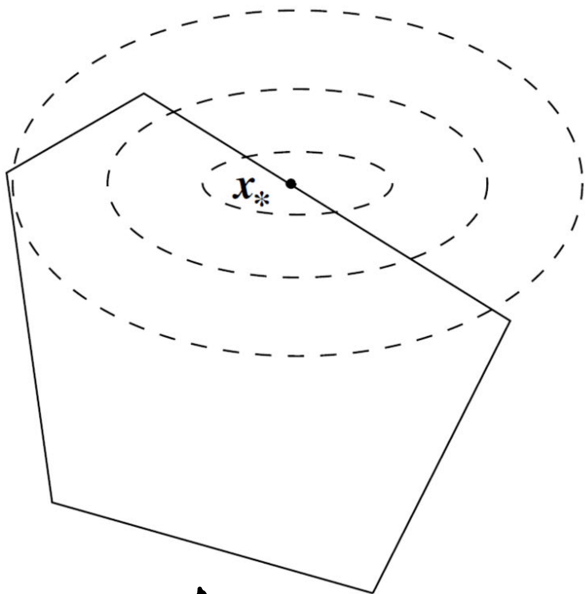




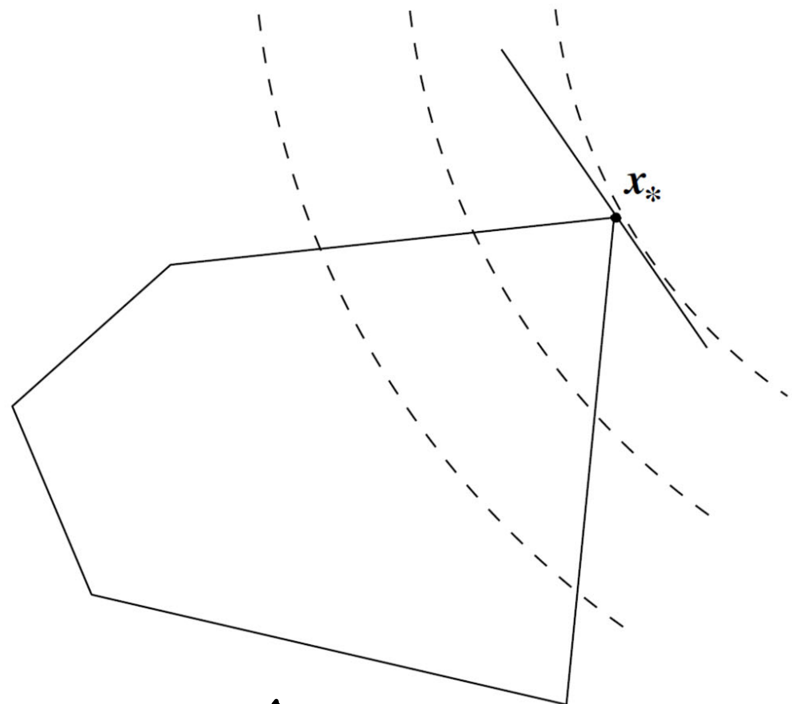
Nonconvex QPs

$G = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ $\quad x_1^2 - x_2^2 = (x_1+x_2)(x_1-x_2)$ $\qquad G = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$

Two kinds of degeneracy

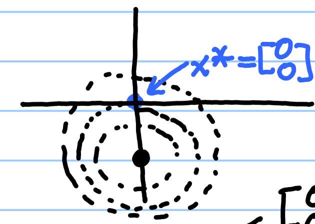


$\uparrow$
non-strict complimentarity
i.e. a constraint is active,
but the corresponding
Lagrange multiplier is 0



$\uparrow$
the active constraint gradients,
i.e. a_i, $i \in \mathcal{A}(x^*)$, are linearly dependent.

Another example:
$\min\ x_1^2 + (x_2+1)^2/2$
s.t. $x_1, x_2 \geqslant 0$



$= \begin{bmatrix}0\\0\end{bmatrix}$

$\nabla \mathcal{L}(x^*, \lambda^*) =$
$\begin{bmatrix}x_1^*\\x_2^*+1\end{bmatrix} - \lambda_1^* \begin{bmatrix}1\\0\end{bmatrix} - \lambda_2^* \begin{bmatrix}0\\1\end{bmatrix}$
$\begin{bmatrix}0\\1\end{bmatrix}$ $\Rightarrow \lambda_1^* = 0$, while
both constraints
are active at x^*.

$\uparrow$
non-strict
complimentarity

Such degeneracies can trip up algorithms. We shall see how to deal with them...

(Primal) Active-set Methods

In a nutshell,

- Generates iterates that remain feasible (for the QP) while steadily decreasing the objective $q(x)$.

- Find a step from one iterate to the next by solving a quadratic subproblem in which some of the inequality constraints, and all the equality constraints, are imposed as equalities. This subset at the k-th iterate x_k is denoted by

$$\mathcal{W}_k \quad (\mathcal{E} \subset \mathcal{W}_k \subset \mathcal{E} \cup \mathcal{I}). \leftarrow \text{the "working set"}$$

  (Both features are similar to the simplex method for LPs, except that the iterates x_k are not necessarily vertices of the feasible polytope.)

- $\mathcal{W}_k$ is maintained so that $\{a_i : i \in \mathcal{W}_k\}$ is a linearly independent set, even when the full set of active constraint gradients at x_k, i.e. $\{a_i : a_i^T x_k = b_i\}$, is linearly dependent.

  (i.e. $\mathcal{W}_k$ may be a proper subset of $\mathcal{A}(x_k)$)

Idea of a single step :

- Given an iterate x_k (feasible) and the working set W_k ($a_i^T x_k = b_i$, $i \in W_k$) first compute a step p by solving

$$\min \; q(x_k + p) \quad \text{s.t.} \quad a_i^T(x_k+p) = b_i \quad i \in W_k$$

$$q(x_k+p) = \frac{1}{2} p^T G p + (Gx_k)^T p + \frac{1}{2} x_k^T G x_k + c^T(x_k+p)$$

(ignore the other inequality constraints at this point)

It is equivalent to solving :

$$\min \; \frac{1}{2} p^T G p + (Gx_k + c)^T p \quad \text{s.t.} \quad a_i^T p = 0 \;, \; i \in W_k .$$

Call the solution p_k. (computed by the techniques in sec 16.2 or 16.3.)

Note : since $a_i^T x_k = b_i$ and $a_i^T p_k = 0 \;\; \forall i \in W_k$, $a_i^T(x_k + \alpha p_k) = b_i \;\; \forall \alpha \in \mathbb{R}$.

So all constraints in W_k are satisfied (and active) at $x_k + \alpha p_k \;\; \forall \alpha$.

- Suppose for now $p_k \neq 0$, we next need to decide how far to move along p_k in order to get a better estimate for the original QP.

If $x_k + p_k$ satisfies all the remaining constraints (ie. those in $\mathcal{I} \setminus \mathcal{W}_k$), set

$$x_{k+1} = x_k + p_k .$$

Otherwise, set $x_{k+1} = x_k + \alpha_k p_k$ where α_k is chosen to be the largest value in $[0,1]$ for which all the constraints are satisfied.

Explicit expression for α_k :

For each $i \in \mathcal{I} \setminus \mathcal{W}_k$,

if $a_i^T p_k \geq 0$, then $a_i^T(x_k + \alpha_k p_k) \geq a_i^T x_k \geq b_i \quad \forall \alpha_k \geq 0$

if $a_i^T p_k < 0$, $\quad a_i^T(x_k + \alpha_k p_k) \geq b_i \iff \alpha_k \leq (b_i - a_i^T x_k)/a_i^T p_k$

So the $\alpha_k \in [0,1]$ maximizing the decrease in q and retaining feasibility is :

$$\alpha_k = \min\left(1, \min_{\substack{i \in \mathcal{I} \setminus \mathcal{W}_k \\ a_i^T p_k < 0}} \frac{b_i - a_i^T x_k}{a_i^T p_k}\right). \qquad (16.41)$$

(similar to the 'ratio test' in the simplex method)

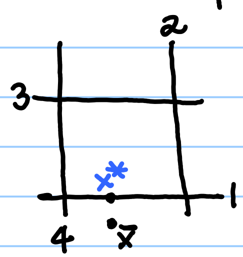
We call the constraints i for which the minimum in (16.41) is achieved the blocking constraints. (If $\alpha_k = 1$ and no new constraints are active at $x_k + p_k$, then there are no blocking constraints on this iteration.)

Note: α_k above can be zero, as we could have $a_i^T p_k < 0$ for some constraint i that is active at x_k but not a member of $\mathcal{W}_k$.

If $\alpha_k < 1$, we add one of the blocking constraint to $\mathcal{W}_k$ to form $\mathcal{W}_{k+1}$.

If $\alpha_k = 1$, set $\mathcal{W}_{k+1} = \mathcal{W}_k$, then in the next iteration, we must have $p_{k+1} = 0$.

Example: $\min\ (x-\bar{x})^T(x-\bar{x})$ s.t. $0 \le x_1 \le 1$, $0 \le x_2 \le 1$.



$\bar{x} = \begin{bmatrix} 0.3 \\ -0.2 \end{bmatrix}$ $x^* \begin{bmatrix} 0.3 \\ 0 \end{bmatrix}$

Consider two initial $x^0, \mathcal{W}_0$:

I. $\mathcal{W}_0 = \emptyset$

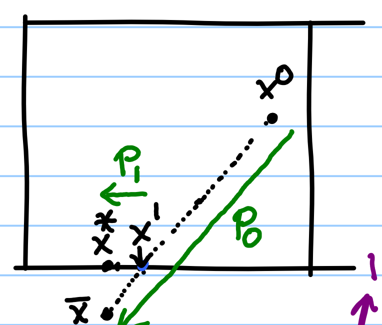
$x^0 + p_0$ solves $\min\ (x-\bar{x})^T(x-\bar{x})$

$x^1 \leftarrow (*, 0)^T$, $\mathcal{W}_1 \leftarrow \{1\}$

x^2 solves $\min\ (x-\bar{x})^T(x-\bar{x})$ s.t. $x_2 = 0$

$\Rightarrow x^2 = \begin{bmatrix} 0.3 \\ 0 \end{bmatrix} = x^*$

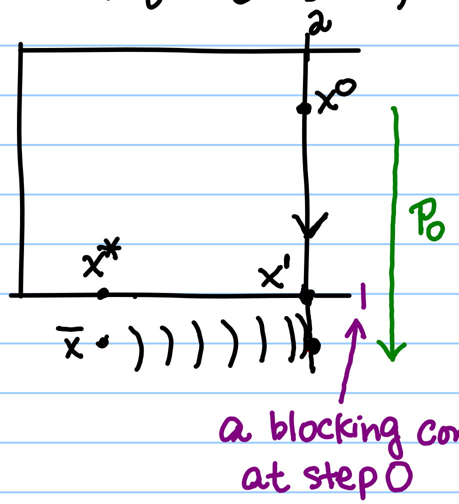
Done!



a blocking constraint at step 0

Apology: I have to write x^k instead of x_k

II. $\mathcal{W}_0 = \{2\}$, $x^0 = [1, 0.8]^T$



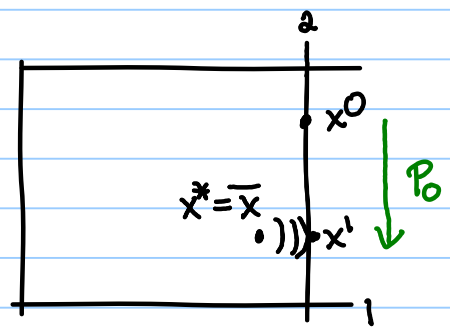
$x^0 + p_0$ solves $\min (x-\bar{x})^T(x-\bar{x})$ s.t. $x_1 = 1$

$x^1 \leftarrow [1, 0]^T$, $\mathcal{W}_1 = \{2, 1\}$

Clearly, x^1 already solves $\min (x-\bar{x})^T(x-\bar{x})$ s.t. $x_1 = 1$, $x_2 = 0$

$\Rightarrow p_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. What to do next?

Another example: $\min (x-\bar{x})^T(x-\bar{x})$ s.t. $0 \le x_1 \le 1$, $0 \le x_2 \le 1$, $\bar{x} = [0.8, 0.3]^T$



$\mathcal{W}_0 = \{2\}$, $x^0 = [1, 0.8]^T$

$x^1 = x^0 + p_0$ no block constraint, so $\mathcal{W}_1 \leftarrow \{2\}$

x^1 solves $\min (x-\bar{x})^T(x-\bar{x})$ s.t. $x_1 = 1$ $\Rightarrow p_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

What to do next?

- When we arrive at an iterate $\hat{x}$ minimizes the quadrative objective function over the current working set $\hat{W}$, the corresponding step size is $p=0$.
  In this case, the solution of

$$\min_p \frac{1}{2} p^T G p + (G\hat{x}+c)^T p \quad \text{st} \quad a_i^T p = 0 \;, i \in \hat{W},$$

  or that of

$$\begin{bmatrix} G & \hat{A}^T \\ \hat{A} & 0 \end{bmatrix} \begin{bmatrix} -p \\ \lambda \end{bmatrix} = \begin{bmatrix} G\hat{x}+c \\ 0 \end{bmatrix} \qquad \hat{A}^T := [\cdots a_i \cdots]_{i \in \hat{W}},$$

  satisfies $p=0$. Denote by $\hat{\lambda}$ the optimal Lagrange multipliers of the above subproblem, which satisfies

$$\sum_{i \in \hat{W}} a_i \hat{\lambda}_i = G\hat{x}+c \qquad (16.42)$$

If $\hat{\lambda}_j \geq 0$, $\forall j \in \hat{W} \cap \mathcal{I}$, $\hat{x}$ must be a solution of the original QP. (Ex: explain why.)

Otherwise, $\hat{\lambda}$ must have a negative component. (Ex: check this for the examples above by solving the three KKT linear systems.)

What to do next? Ans: remove from $\hat{W}$ an index corresponding to such a negative Lagrange multiplier.

To the least, by dropping a constraint, there is a chance to further decrease $q(\cdot)$.

more is true:

Theorem 16.5.

Suppose that the point $\hat{x}$ satisfies first-order conditions for the equality-constrained subproblem with working set $\hat{\mathcal{W}}$; that is, equation (16.42) is satisfied along with $a_i^T \hat{x} = b_i$ for all $i \in \hat{\mathcal{W}}$. Suppose, too, that the constraint gradients a_i, $i \in \hat{\mathcal{W}}$, are linearly independent and that there is an index $j \in \hat{\mathcal{W}}$ such that $\hat{\lambda}_j < 0$. Let p be the solution obtained by dropping the constraint j and solving the following subproblem:

$$\min_p \ \tfrac{1}{2} p^T G p + (G\hat{x} + c)^T p, \tag{16.43a}$$

$$\text{subject to } a_i^T p = 0, \text{ for all } i \in \hat{\mathcal{W}} \text{ with } i \neq j. \tag{16.43b}$$

Then p is a feasible direction for constraint j, that is, $a_j^T p \geq 0$. Moreover, if p satisfies second-order sufficient conditions for (16.43), then we have that $a_j^T p > 0$, and that p is a descent direction for $q(\cdot)$.

The 2nd order sufficient conditions for (16.43) is:

$Z^T G Z \succ 0$ where the columns of Z form a basis of the null space of $\begin{bmatrix} \vdots \\ - a_i^T - \\ \vdots \end{bmatrix}_{i \in \hat{W}}$.

If $G \succ 0$, it is always satisfied. (So only when $G \succeq 0$ and G has at least a zero eigenvalue do we have to be concerned with it.)

The proof of the theorem above is nice and clear / not hard; see N&W.

The steps described above give the following algorithm.

Ex: If $G = [0]$ (so the QP becomes a LP), x_0 chosen to be a vertex, W_0 set accordingly, does the following algorithm boils down to the simplex method?

It seems like every two steps of the following algorithm is one step of the simplex method.

Algorithm 16.3 (Active-Set Method for Convex QP).

Compute a feasible starting point x_0; $\leftarrow$ use Phase I or a related method ; see sec 16.5 N&W

Set $\mathcal{W}_0$ to be a subset of the active constraints at x_0;

for $k = 0, 1, 2, \ldots$

  Solve (16.39) to find p_k;

  if $p_k = 0$

    Compute Lagrange multipliers $\hat{\lambda}_i$ that satisfy (16.42), with $\hat{\mathcal{W}} = \mathcal{W}_k$;

    if $\hat{\lambda}_i \geq 0$ for all $i \in \mathcal{W}_k \cap \mathcal{I}$

      stop with solution $x^* = x_k$;

    else

      $j \leftarrow \arg\min_{j \in \mathcal{W}_k \cap \mathcal{I}} \hat{\lambda}_j$;

      $x_{k+1} \leftarrow x_k$; $\mathcal{W}_{k+1} \leftarrow \mathcal{W}_k \setminus \{j\}$;

  else (* $p_k \neq 0$ *)

    Compute α_k from (16.41);

    $x_{k+1} \leftarrow x_k + \alpha_k p_k$;

    if there are blocking constraints

      Obtain $\mathcal{W}_{k+1}$ by adding one of the blocking constraints to $\mathcal{W}_k$;

    else

      $\mathcal{W}_{k+1} \leftarrow \mathcal{W}_k$;

end (for)

(16.39): $\min_p \frac{1}{2} p^T G p + (G x_k + c)^T p$ s.t. $a_i^T p = 0$, $i \in \mathcal{W}_k$

(16.42): $\sum_{i \in \hat{\mathcal{W}}} a_i \hat{\lambda}_i = G x_k + c$

(16.41): $\alpha_k = \min\left(1, \min_{\substack{i \in \mathcal{I} \setminus \mathcal{W}_k \\ a_i^T p_k < 0}} \frac{b_i - a_i^T x_k}{a_i^T p_k}\right)$

Example

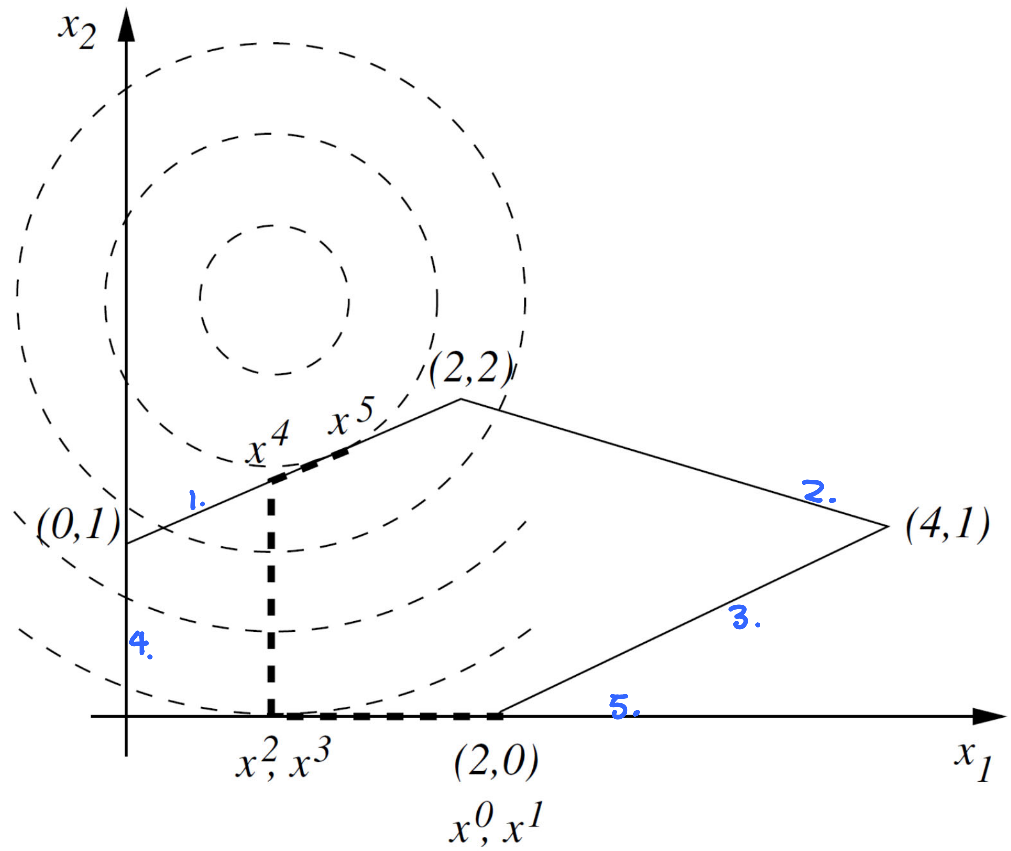
$\min\ (x-\bar{x})^T(x-\bar{x})$ s.t. $x_1 - 2x_2 + 2 \geq 0$ (1.)

$- x_1 - 2x_2 + 6 \geq 0$ (2.)

$\bar{x} = \begin{bmatrix} 1 \\ 2.5 \end{bmatrix}$ $-x_1 + 2x_2 + 2 \geq 0$ (3.)

$x_1 \geq 0$ (4.)

$x_2 \geq 0$ (5.)



Remaining issues:

- The algorithm guarantees that the constraint gradients in $\mathcal{W}_k$ ($\subset \mathcal{A}(x_k)$) are linearly independent.
- finite termination / cycling
- KKT matrix factorization updates

see the second half of sec 16.5 for details.

Interior Point Methods (a very brief discussion)

For simplicity, assume $\mathcal{E}=\emptyset$, i.e. the QP is of the form

$$\min q(x) = \tfrac{1}{2}x^T Q x + c^T x \quad \text{s.t.} \ \overset{m\times n}{A} x \geq b. \qquad (G \succeq 0)$$

The case of $\mathcal{E} \neq \emptyset$ can be easily accomdated by the method below. (To the very least, albeit inefficient, we can always write an equality constraint as two inequality constraints.)

KKT conditions:

$$\begin{aligned} Gx - A^T\lambda + c &= 0 \\ Ax - b &\geq 0 \\ (Ax-b)_i \lambda_i &= 0, \quad i=1,\cdots,m \\ \lambda &\geq 0 \end{aligned} \quad \xrightarrow[y\geq 0]{\text{slack vector}} \quad \begin{aligned} Gx - A^T\lambda + c &= 0 \\ Ax - y - b &= 0 \\ y_i \lambda_i &= 0 \quad i=1,\cdots,m \\ y, \lambda &\geq 0 \end{aligned}$$

Again, when $G \succeq 0$, these conditions are necessary and sufficient for optimality.

We would like to solve the (mildly?) nonlinear square system

$$\begin{bmatrix} Gx - A^T\lambda + c \\ Ax - y - b \\ Y\Lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 & & \\ & \ddots & \\ & & y_m \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{bmatrix}.$$

And we want to stay away from the 'spurious solutions', i.e. those that do not satisfy $y, \lambda \geq 0$.

As in the interior point method for LP, the idea is to approximately solve

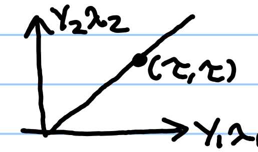
$$\underbrace{\begin{bmatrix} Gx - A^T\lambda + c \\ Ax - y - b \\ Y\Lambda - \begin{bmatrix}\tau \\ \vdots \\ \tau\end{bmatrix} \end{bmatrix}}_{F(x,y,\lambda;\tau)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

with a choice of $\tau > 0$, and gradually shrink τ towards 0.

$\{(x_\tau, y_\tau, \lambda_\tau) : F(x_\tau, y_\tau, \lambda_\tau; \tau) = 0 \text{ for some } \tau > 0\}$ is called the central path of the QP.

Being close to the central curve keeps the approximate solution away from the boundary of the feasible region and staying away from the 'spurious solutions'.

Define a complimentarity measure μ by $\mu := y^T\lambda/m = \frac{1}{m}\sum_{i}^{m} y_i\lambda_i$



Given $(\bar{x}, \bar{y}, \bar{\lambda})$, and applying a damped Newton's method to the system above with τ set to $\sigma\bar{y}^T\bar{\lambda}/m = \sigma\bar{\mu}$ for some $\sigma \in [0,1]$ and $(\bar{x}, \bar{y}, \bar{\lambda})$ as the current iterate, the next iterate is given by $\underbrace{(\bar{x}, \bar{y}, \bar{\lambda}) + \alpha(\Delta x, \Delta y, \Delta\lambda)}_{(x^+, y^+, \lambda^+)}$ where

$$\begin{bmatrix} G & 0 & -A^T \\ A & -I & 0 \\ 0 & \bar{\Lambda} & \bar{Y} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta\lambda \end{bmatrix} = \begin{bmatrix} -G\bar{x} + A^T\bar{\lambda} - c \\ -A\bar{x} + \bar{y} + b \\ -\bar{\Lambda}\bar{Y}e + \sigma\bar{\mu}e \end{bmatrix} \quad e = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \qquad (*)$$

The damping factor α is chosen to retain the inequality $y^+, \lambda^+ > 0$.

The σ is called the 'centrality parameter':

if $\sigma = 1$, $(*)$ is the Newton step for solving $F(x, y, \lambda; \bar{\mu}) = 0$, meaning that it tries to drives the iterate closer to the central curve. It does not drive μ closer to zero.
But, by moving closer to the central curve, it sets the scene for a substantial reduction in μ on the next iteration.

if $\sigma = 0$, $(*)$ is the Newton step for solving $F(x, y, \lambda; 0) = 0$, but likely pushing the iterate to the boundary of the feasible region, causing difficulties in subsequent steps.

The idea is to chose an intermediate value $\sigma \in (0,1)$ to balance the twin goals of reducing μ and improving centrality.

quadprog

Quadratic programming

Syntax

```
x = quadprog(H,f)
x = quadprog(H,f,A,b)
x = quadprog(H,f,A,b,Aeq,beq)
x = quadprog(H,f,A,b,Aeq,beq,lb,ub)
x = quadprog(H,f,A,b,Aeq,beq,lb,ub,x0)
x = quadprog(H,f,A,b,Aeq,beq,lb,ub,x0,options)
x = quadprog(problem)
[x,fval] = quadprog( __ )
[x,fval,exitflag,output] = quadprog( __ )
[x,fval,exitflag,output,lambda] = quadprog( __ )
```

```
[wsout,fval,exitflag,output,lambda] = quadprog(H,f,A,b,Aeq,beq,lb,ub,ws)
```

Description

Solver for quadratic objective functions with linear constraints.

quadprog finds a minimum for a problem specified by

$$\min_x \frac{1}{2} x^T H x + f^T x \text{ such that } \begin{cases} A \cdot x \leq b, \\ Aeq \cdot x = beq, \\ lb \leq x \leq ub. \end{cases}$$

H, A, and Aeq are matrices, and f, b, beq, lb, ub, and x are vectors.

You can pass f, lb, and ub as vectors or matrices; see Matrix Arguments.

16.7 THE GRADIENT PROJECTION METHOD

In the active-set method described in Section 16.5, the active set and working set change slowly, usually by a single index at each iteration. This method may thus require many iterations to converge on large-scale problems. For instance, if the starting point x^0 has no active constraints, while 200 constraints are active at the (nondegenerate) solution, then at least 200 iterations of the active-set method will be required to reach the solution.

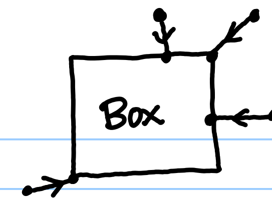
The gradient projection method allows the active set to change rapidly from iteration to iteration. It is most efficient when the constraints are simple in form—in particular, when there are only bounds on the variables. Accordingly, we restrict our attention to the following bound-constrained problem:

$$\left.\begin{aligned} &\min_x \quad q(x) = \tfrac{1}{2}x^T G x + x^T c && (16.68a) \\ &\text{subject to} \quad l \le x \le u, \quad \text{a 'box'} && (16.68b) \end{aligned}\right\} \text{ a bound-constrained QP}$$

In fact, the gradient projection method applies to a general convex problem:

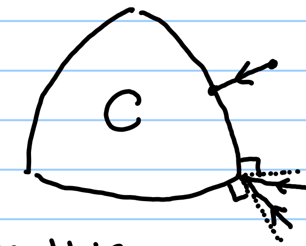
$\min f(x)$ s.t. $x \in C$. $C \subseteq \mathbb{R}^n$ a convex set, $f: C \to \mathbb{R}$ a convex function

The method is simply based on gradient descent followed by projection to C, i.e.



$$x^{k+1} = P_C(x_k - t_k \nabla f(x_k)),$$

where $P_C : \mathbb{R}^n \to C$ is the orthogonal projector onto C,



$$P_C(x) = \operatorname{argmin}\{\|y-x\|_2^2 : y \in C\}.$$ (see [Beck] Ch. 8.)

For a rigorous convergence analysis of the gradient projection method in this general context, see Ch 9 of [Beck].

The problem is that the computation of P_C — a convex optimization problem by itself — can be as costly as solving the original convex problem. (This is the case even when C is a convex polytope.)

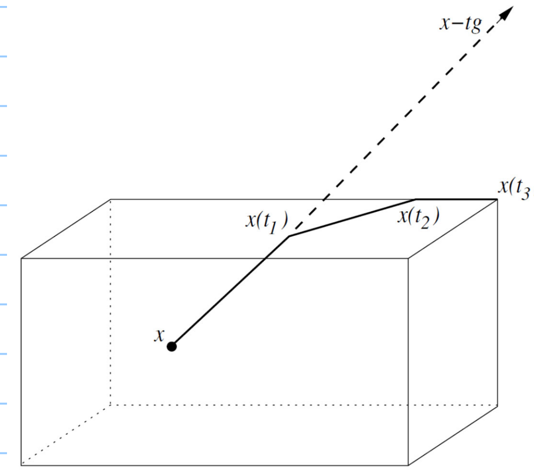
However, when C is a box $P_C(x)$ is very easy to compute. Moreover, we can design a more specialized and efficient variant of the general gradient projection method for bound constrained QPs.

Each iteration of the gradient projection algorithm consists of two stages:

- Find the first local minimizer of q along the following piecewise-linear path

$$P(x - t\,g\,;\,l, u)$$

projection onto the box; current iterate; gradient of q $= Gx + c$



$$P(x; l, u) = \begin{cases} l_i & \text{if } x_i < l_i \\ x_i & \text{if } x_i \in [l_i, u_i] \\ u_i & \text{if } x_i > u_i \end{cases}$$

The solution of this subproblem is referred to as the Cauchy point, and is denoted by x^c.

- Further improve x^c by solving $\min_x q(x)$ s.t. $\left.\begin{array}{ll} x_i = x_i^c & i \in \mathcal{A}(x^c) \\ l_i \le x_i \le u_i & i \notin \mathcal{A}(x^c) \end{array}\right\}$ meaning x is constrained to a (lower-dimensional) 'face' of the box.

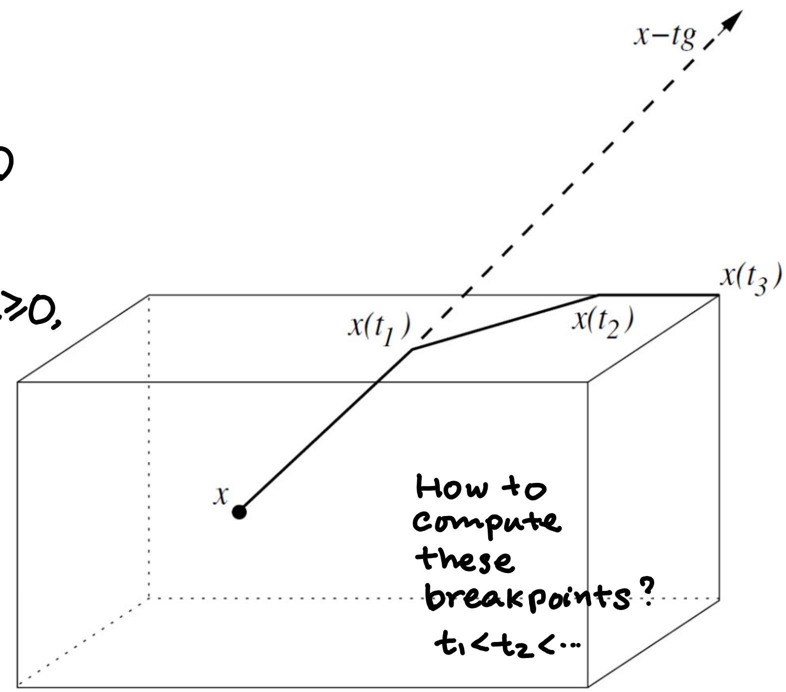
This subproblem (a QP itself) is only solved approximately, as it can be as difficult to solve as the original QP.

Cauchy Point computation

Write $x(t) = P(x - tg\,;\, l, u)$ (piecewise linear)

$q(x(t))$ is a piecewise quadratic (univariate) function in t

To find x^c = the first local minimizer of $q(x(t))$, $t \geq 0$, we need to find the values of t at which the kinks in $x(t)$ (or breakpoints) occur.



Explicit formulae:

First, $\bar{t}_i$ = the value of t s.t. $(x - tg)_i$ reaches either l_i or u_i

$$= \begin{cases} (x_i - u_i)/g_i & \text{if } g_i < 0 \text{ and } u_i < +\infty \\ (x_i - l_i)/g_i & \text{if } g_i > 0 \text{ and } l_i > -\infty \\ \infty & \text{otherwise} \end{cases}.$$

(Note: some $\bar{t}_i$ may be 0.)

The components of $x(t) = P(x - tg\,;\, l, u)$ are therefore $x_i(t) = \begin{cases} x_i - t g_i & \text{if } 0 \leq t \leq \bar{t}_i \\ x_i - \bar{t}_i g_i & \text{if } t \geq \bar{t}_i \end{cases}$.

To search for the first local minimizer of $q(x)$ along $P(x - tg\,;\, l, u)$, we eliminate the duplicate values and zero values from the set $\{\bar{t}_1, \cdots, \bar{t}_n\}$ to obtain a sorted, reduced set of breakpoints $\{t_1, \cdots, t_\ell\}$ with $t_1 < \cdots < t_\ell$. (see Figure.)

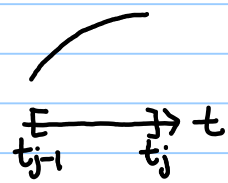
We then have the following simple formula for $x(t)$:

$$\text{For } t \in [t_{j-1}, t_j], \quad x(t) = x(t_{j-1}) + \underbrace{(t - t_{j-1})}_{=:\,\Delta t} p^{j-1}, \quad p_i^{j-1} = \begin{cases} -g_i & \text{if } t_{j-1} < \bar{t}_i \\ 0 & \text{otherwise} \end{cases}$$

So, for $t \in [t_{j-1}, t_j]$, $q(x(t)) = c^T x(t) + \frac{1}{2} x(t)^T G\, x(t) = f_{j-1} + f_{j-1}' \Delta t + \frac{1}{2} f_{j-1}'' (\Delta t)^2$,

$$f_{j-1} := c^T x(t_{j-1}) + \tfrac{1}{2} x(t_{j-1})^T G\, x(t_{j-1}), \quad f_{j-1}' := c^T p^{j-1} + x(t_{j-1})^T G\, p^{j-1}, \quad f_{j-1}'' := (p^{j-1})^T G\, p^{j-1}.$$

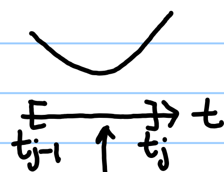
A local min found on $[t_{j-1}, t_j)$ if:



a local min

if $f_{j-1}' < 0$

OR



a local min at

$t_{j-1} \underbrace{- f_{j-1}'/f_{j-1}''}_{\Delta t^*}$

$\Delta t^* \in [0, t_j - t_{j-1})$.

If not found, move on to the next interval $[t_j, t_{j+1}]$ and continue the search.

Subspace minimization

After x^c has been computed, set $\mathcal{A}(x^c) = \{ i : x_i^c = l_i \text{ or } u_i \}$.

In the second stage of the gradient projection iteration, we approximately solve the QP by fixing components x_i for $i \in \mathcal{A}(x^c)$ at the values x_i^c. The remaining components are determined from the subproblem

$$\min_x q(x) = \tfrac{1}{2} x^T G x + c^T x \quad \text{s.t.} \quad \begin{array}{ll} x_i = x_i^c, & i \in \mathcal{A}(x^c), \\ l_i \le x_i \le u_i & i \notin \mathcal{A}(x^c). \end{array}$$

Idea for solving the above problem approximately when $G \succeq 0$:

Apply conjugate gradient (CG) to iteratively solve $\min q(x)$ st $x_i = x_i^c$, $i \in \mathcal{A}(x^c)$. Stop as soon as a CG iterate has a component $i \notin \mathcal{A}(x^c)$ is out-of-bound, followed by projecting the iterate back to the box.

Call the approximate solution x^+. Analysis shows that global convergence of the gradient projection procedure requires only that $q(x^+) \le q(x^c)$.

Algorithm 16.5 (Gradient Projection Method for QP).

Compute a feasible starting point x^0;

for $k = 0, 1, 2, \ldots$

min $q(x)$ s.t. $l \leq x \leq u$

if x^k satisfies the KKT conditions for (16.68)

stop with solution $x^* = x^k$;

Set $x = x^k$ and find the Cauchy point x^c;

Find an approximate solution x^+ of (16.74) such that $q(x^+) \leq q(x^c)$ and x^+ is feasible;

$x^{k+1} \leftarrow x^+$;

end (for)

The gradient projection method can be applied in principle to QPs with general linear constraints, with the "projection onto a box" replaced by a "projection onto a convex polytope". The latter may be as costly to compute as solving the original QP.

The gradient projection method can potentially be applied to the dual of a general QP. Recall: the dual of $\min_x \frac{1}{2}x^T G x + c^T x$ st $Ax \geq b$ is:

$$\max_\lambda -\frac{1}{2}(A^T\lambda - c)^T G^{-1}(A^T\lambda - c) + b^T\lambda, \quad \lambda \geq 0.$$

a bound constrained QP!

if G^{-1} is efficient to compute with, applying the gradient projection method to the dual (as a way to solve the primal) can be a good idea.