Quadratic Programming

Saurav Samantaray [1]

[1]Department of Mathematics
IIT Madras

April 26, 2024

Quadratic Programming

- An optimisation problem with a quadratic objective function and linear constraints is called a quadratic program.
- Also arise as sub-problems in methods for general constrained optimisation.
- The general quadratic program (QP) can be stated as:

$$\min_x \; q(x) = \frac{1}{2}x^T G x + x^T c \tag{1}$$

$$\text{subject to} \quad a_i^T x = b_i, \quad i \in \mathcal{E}, \tag{2}$$

$$a_i^T x \geq b_i, \quad i \in \mathcal{I}. \tag{3}$$

- G is a symmetric $n \times n$ matrix, $\mathcal{E}$ and $\mathcal{I}$ are finite sets of indices.
- c, x and $\{a_i\}$, $i \in \mathcal{E} \cup \mathcal{I}$, are vectors in $\mathbb{R}^n$.

Quadratic Programming

- If the Hessian matrix G is positive semi-definite, then (1) is a convex QP.
- For convex QPs the problem is often similar in difficulty to a linear program.
- Strictly convex QPs are those in which G is positive definite.
- Non-convex QPs, in which G is an indefinite matrix, are more challenging because they can have several stationary points and local minima.
- We focus primarily on convex quadratic programs.

Equality Constrained Quadratic Programs

- Consider the case in which only equality constraints are present.
- Techniques for this special case are applicable also to problems with inequality constraints, as some algorithms for general QP require the solution of an equality-constrained QP at each iteration.
- The equality constrained QP is given by:

$$\min_x \; q(x) = \frac{1}{2}x^T G x + x^T c \tag{4}$$

$$\text{subject to} \quad Ax = b, \tag{5}$$

- A is the $m \times n$ Jacobians of constraints (with $m \leq n$) whose rows are a_i^T, $i \in \mathcal{E}$.
- b is the vector in $\mathbb{R}^m$ whose components are b_i, $i \in \mathcal{E}$.
- Assume A has full row rank (rank m) so the constraints are consistent.

First-Order Necessary Conditions

- The first-order necessary conditions for x^* to be a solution of (4) state that there is a vector λ^* such that the following system of equations is satisfied:

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix} \tag{6}$$

- These conditions are a consequence of the general result for first-order optimality conditions.
- λ^* is the vector of Lagrange multipliers.

First-Order Necessary Conditions

- The system (6) can be rewritten in a form that is useful for computation by expressing x^* as $x^* = x + p$, where x is some estimate of the solution and p is the desired step.
- By introducing this notation and rearranging the equations

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = \begin{bmatrix} g \\ h \end{bmatrix} \tag{7}$$

- $$h = Ax - b, \quad g = c + Gx, \quad p = x^* - x.$$

- The matrix in (7) is called the Karush–Kuhn–Tucker (KKT) matrix.

First-Order Necessary Conditions

- Z denotes the $n \times (n-m)$ matrix whose columns are a basis for the null space of A.
- That is Z has full rank and satisfies $AZ = 0$.

Lemma

Let A have full row rank, and assume that the reduced-Hessian matrix $Z^T G Z$ is positive definite. Then the KKT matrix

$$K = \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \tag{8}$$

is nonsingular, and hence there is a unique vector pair (x^*, λ^*) satisfying (6).

First-Order Necessary Conditions

- Suppose the KKT matrix is singular, therefore there exists vectors w and v such that

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} w \\ v \end{bmatrix} = 0 \tag{9}$$

- Since $Aw = 0$, we have from the above

$$0 = \begin{bmatrix} w \\ v \end{bmatrix}^T \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} w \\ v \end{bmatrix} = w^T G w$$

- Since w lies in the null space of A, it can be written as $w = Zu$ for some vector $u \in \mathbb{R}^{n-m}$.

- Therefore, we have

$$0 = w^T G w = u^T Z^T G Z u,$$

which by positive definiteness of $Z^T G Z$ implies that $u = 0$.

- Therefore, $w = 0$, and, $A^T v = 0$.
- Full row rank of A then implies that $v = 0$.
- We conclude that equation (9) is satisfied only if $w = 0$ and $v = 0$, so the matrix is non-singular, as claimed.

Example

- Consider the quadratic programming problem

$$\begin{aligned} \min q(x) =& 3x_1^2 + 2x_1x_2 + x_1x_3 + 2.5x_2^2 + 2x_2x_3 + 2x_3^2 - 8x_1 \\ & - 3x_2 - 3x_3, \\ & \text{subject to } \; x_1 + x_3 = 3, \quad x_2 + x_3 = 0. \end{aligned} \tag{10}$$

- We rewrite the problem by defining

$$G = \begin{bmatrix} 6 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 4 \end{bmatrix}, \; c = \begin{bmatrix} -8 \\ -3 \\ -3 \end{bmatrix}, \; A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \; b = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

- The solution x^* and the optimal Lagrange multiplier vector λ^* are: $\quad x^* = (2, -1, 1)^T, \qquad \lambda^* = (3, -2)^T.$
- G is a positive definite matrix and the null-space basis matrix can be defined as

$$Z = (-1, -1, 1)^T$$

Second-Order Sufficient Condition

- When the conditions of the above Lemma are satisfied, there exists a unique vector pair (x^*, λ^*) that satisfies the first-order necessary conditions.
- Under the above stated circumstances even the second-order sufficient conditions are also satisfied at (x^*, λ^*), so x^* is a strict local minimiser.
- It can also be shown that x^* is a global solution.

Theorem

Let A have full row rank and assume that the reduced-Hessian matrix $Z^T GZ$ is positive definite. Then the vector x^* satisfying the first-order necesary condition (6) is the unique global solution of (4)

Lecture

Saurav

DIRECT SOLUTION OF THE KKT SYSTEM

INEQUALITY-CONSTRAINED PROBLEMS

ITERATIVE SOLUTION OF THE KKT SYSTEM

Proof of Theorem

- Let x be any other feasible point (satisfying $Ax = b$).
- Let $p = x^* - x$.
- Since $Ax^* = Ax = b$, we have that $Ap = 0$.
- Substituting into the objective function we get

$$\begin{aligned} q(x) &= \frac{1}{2}(x^* - p)^T G(x^* - p) + c^T(x^* - p) \\ &= \frac{1}{2}p^T Gp - p^T Gx^* - c^T p + q(x^*) \end{aligned} \tag{11}$$

- From first-order necessary conditions we have $Gx^* = -c + A^T\lambda^*$.
- From $Ap = 0$ we have

$$p^T Gx^* = p^T(-c + A^T\lambda^*) = -p^T c.$$

Proof of Theorem

- By substituting the above into (11) we get

$$q(x) = \frac{1}{2}p^T Gp + q(x^*).$$

- Since p lies in the null space of A, we can write $p = Zu$ for some vector $u \in \mathbb{R}^{n-m}$, so that

$$q(x) = \frac{1}{2}u^T Z^T GZu + q(x^*).$$

- By positive definiteness of $Z^T GZ$, we conclude that $q(x) > q(x^*)$, except when $u = 0$, that is, when $x = x^*$.
- Therefore, x^* is the unique global solution.

Classification of the solutions

Assuming the KKT system has solutions $\begin{pmatrix} x* \\ \lambda^* \end{pmatrix}$:

1. Strong local minimiser at $x^* \iff Z^T GZ$ pd.
2. Infinite solutions if $Z^T GZ$ is psd and singular.
3. Unbounded if $Z^T GZ$ indefinite.

- The KKT system can be solved with various linear algebra techniques.

Table of Contents

Inertia of Matrix

Definition

Define the **inertia** of a symmetric matrix K to be the scalar triple that indicates the numbers n_+, n_-, and n_0 of positive, negative, and zero eigenvalues, respectively, that is,

$$\text{inertia}(K) = (n_+, n_-, n_0)$$

The following result characterizes the inertia of the KKT matrix.

Definition

Let K be defined by (8), and suppose that A has rank m. Then

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

Therefore, if $Z^T GZ$ is positive definite, $\text{inertia}(K) = (n, m, 0)$.

FACTORING THE FULL KKT SYSTEM

- Perform a triangular factorisation on the full KKT matrix and then perform backward and forward substitution with the triangular factors.
- The most effective strategy in this case is to use a symmetric indefinite factorisation.
- For a general symmetric matrix K , this factorisation has the form

$$P^T K P = L B L^T$$

- P is a permutation matrix, L is unit lower triangular, and B is block-diagonal with either 1×1 or 2×2 blocks.
- The symmetric permutations P are introduced for numerical stability of the computation and, in the case of large sparse K , for maintaining sparsity.
- Computational cost is typically about half the cost of sparse Gaussian elimination.

SCHUR-COMPLEMENT METHOD

- Assume that G is positive definite.
- The first equation in (7) can be multiplied by AG^{-1} and then subtract the second equation to obtain a linear system in the vector λ^* alone:

$$(AG^{-1}A^T)\lambda^* = (AG^{-1}g - h). \tag{12}$$

- Solve this symmetric positive definite system for λ^* and then recover p from the first equation in (7) by solving:

$$Gp = A^T\lambda^* - g. \tag{13}$$

SCHUR-COMPLEMENT METHOD

This approach requires to perform operations with G^{-1}, as well as to compute the factorisation of the $m \times m$ matrix $AG^{-1}A^T$. Therefore, it is most useful when:

- G is well conditioned and easy to invert (for instance, when G is diagonal or block-diagonal); or
- G^{-1} is known explicitly through a quasi-Newton updating formula; or
- the number of equality constraints m is small, so that the number of back solves needed to form the matrix $AG^{-1}A^T$ is not too large.

SCHUR-COMPLEMENT METHOD

- An approach like the Schur-complement method can be written to derive an explicit inverse formula for the KKT matrix in (7).
- The formula is

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix}^{-1} = \begin{bmatrix} C & E \\ E^T & F \end{bmatrix} \tag{14}$$

-

$$\begin{aligned} C &= G^{-1} - G^{-1}A^T(AG^{-1}A^T)^{-1}AG^{-1}, \\ E &= G^{-1}A^T(AG^{-1}A^T)^{-1} \\ F &= -(AG^{-1}A^T)^{-1} \end{aligned}$$

- The solution can be obtained by multiplying its right-hand side by this inverse matrix.

NULL-SPACE METHOD

- The null-space method does not require non-singularity of G.
- Has wider applicability than the Schur-complement method.
- It assumes that A has full row rank and that $Z^T GZ$ is positive definite.
- It requires knowledge of the null-space basis matrix Z.
- Like the Schur-complement method, it exploits the block structure in the KKT system to decouple it into two smaller systems.

NULL-SPACE METHOD

- Suppose that we partition the vector p into two components, as follows:

$$p = Yp_Y + Zp_Z, \tag{15}$$

- Z is the $n \times (n - m)$ null-space matrix
- Y is any $n \times m$ matrix such that $[Y|Z]$ is non-singular,
- p_Y is an m-vector,
- p_Z is an $(n - m)$-vector.
- By substituting p into the second equation of (7), and recalling that $AZ = 0$, we have:

$$(AY)p_Y = -h. \tag{16}$$

NULL-SPACE METHOD

- Since A has rank m and $[Y|Z]$ is $n \times n$ non-singular, the product $A[Y|Z] = [AY|0]$ has rank m.
- Therefore, AY is a non-singular $m \times m$ matrix, and p_Y is well determined by (16).
- Meanwhile, we can substitute (15) into the first equation of (7) to obtain

$$-GYp_Y - GZp_Z + A^T\lambda^* = g$$

- and multiply by Z^T to obtain

$$(Z^TGZ)p_Z = -Z^TGYp_Y - Z^Tg. \tag{17}$$

- This system can be solved by performing a Cholesky factorization of the reduced-Hessian matrix $Z^TGZ$to determine p_Z.

NULL-SPACE METHOD

- To obtain the Lagrange multiplier, multiply the first block row in (7) by Y^T to obtain the linear system

$$(AY)^T\lambda^* = Y^T(g + Gp), \tag{18}$$

- which can be solved for λ^*.

Example

Consider the problem (10). Choose

$$Y = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \\ 1/3 & 1/3 \end{bmatrix}$$

and set $Z = (-1, -1, 1)^T$. Note that $AY = I$.

Example

- Suppose we have $x = (0, 0, 0)^T$. Then

$$h = Ax - b = -b, \qquad g = c + Gx = c = \begin{bmatrix} -8 & -3 & -3 \end{bmatrix}^T$$

- Simple calculation shows that

$$p_Y = \begin{bmatrix} 3 & 0 \end{bmatrix}^T, \qquad p_Z = \begin{bmatrix} 0 \end{bmatrix},$$

- so that

$$p = x^* - x = Yp_Y + Zp_Z = \begin{bmatrix} 2 & -1 & 1 \end{bmatrix}^T$$

- After recovering λ^* from (18) we have:

$$x^* = \begin{bmatrix} 2 & -1 & 1 \end{bmatrix}^T, \qquad \lambda^* = \begin{bmatrix} 3 & -2 \end{bmatrix}^T$$

Table of Contents

OPTIMALITY CONDITIONS FOR INEQUALITY-CONSTRAINED PROBLEMS

$$\min_x \; q(x) = \frac{1}{2}x^T G x + x^T c \tag{19}$$

$$\text{subject to} \quad a_i^T x = b_i, \quad i \in \mathcal{E}, \tag{20}$$

$$a_i^T x \geq b_i, \quad i \in \mathcal{I}. \tag{21}$$

- The Lagrangian function for the general inequality constrained QP is given by:

$$\mathcal{L}(x, \lambda) = \frac{1}{2}x^T G x + x^T c - \sum_{i \in \mathcal{I} \cup \mathcal{E}} \lambda_i (a_i^T x - b_i). \tag{22}$$

- As defined before the active set $\mathcal{A}(x^*)$ at a point x^* consists of the indices of the constraints for which equality holds at x^*:

$$\mathcal{A}(x^*) = \{i \in \mathcal{E} \cup \mathcal{I} | a_i^T x^* = b_i\}. \tag{23}$$

OPTIMALITY CONDITIONS

- The KKT conditions for this problem can be stated as: any solution x^* of (19) satisfies the following first-order conditions, for some Lagrange multipliers λ_i^*, $i \in \mathcal{A}(x^*)$:

$$Gx^* + c - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* a_i = 0, \tag{24}$$

$$a_i^T x^* = b_i, \quad \text{for all} \quad i \in \mathcal{A}(x^*), \tag{25}$$

$$a_i^T x^* \geq b_i, \quad \text{for all} \quad i \in \mathcal{I} \setminus \mathcal{A}(x^*), \tag{26}$$

$$\lambda_i^* \geq 0, \quad \text{for all} \quad i \in \mathcal{I} \cup \mathcal{A}(x^*) \tag{27}$$

- The first optimality conditions still holds if we replace LICQ by other constraint qualifications, such as linearity of the constraints (true for QPs).
- Hence, in the optimality conditions for quadratic programming, we need not assume that the active constraints are linearly independent at the solution.

Second-Order Conditions

Theorem

If x^* satisfies the conditions (24)-(27) for some λ_i^*, $i \in \mathcal{A}(x^*)$, and G is positive semi-definite, then x^* is a global solution of (19).

Proof:

- If x is any other feasible point of (19),

$$\begin{aligned} a_i^T x = b_i & \quad \text{for all } i \in \mathcal{E} \text{ and} \\ a_i^T x \geq b_i & \quad \text{for all } i \in \mathcal{A}(x^*) \cap \mathcal{I} \end{aligned}$$

- So,

$$\begin{aligned} a_i^T (x - x^*) = 0, & \quad \text{for all } \; i \in \mathcal{E} \text{ and} \\ a_i^T (x - x^*) \geq 0, & \quad \text{for all } \; i \in \mathcal{A}(x^*) \cap \mathcal{I} \end{aligned}$$

Second-Order Conditions

- Using the above expression with (24) and (27) we have,

$$(x - x^*)(Gx^* + c) = \sum_{i \in \mathcal{E}} \lambda_i^* a_i^T (x - x^*) + \sum_{i \in \mathcal{A}(x^*) \cap \mathcal{I}} \lambda_i^* a_i^T (x - x^*) \geq 0. \tag{28}$$

- By elementary manipulation (Psd of G):

$$\begin{aligned} q(x) &= q(x^*) + (x - x^*)^T (Gx^* + c) + \frac{1}{2}(x - x^*)^T G (x - x^*) \\ &\geq q(x^*) + \frac{1}{2}(x - x^*)^T G (x - x^*) \\ &\geq q(x^*) \end{aligned}$$

- Therefore $q(x) \geq q(x^*)$ for any feasible x, so x^* is a global solution.

Lecture

Saurav

Degeneracy

Degeneracy is one of the following situations, which can cause problems for the algorithms:

1. the active constraint gradients a_i, $i \in \mathcal{A}(x^*)$, are linearly dependent at the solution x^*, and/or
2. the strict complementarity condition fails to hold, that is, there is some index $i \in \mathcal{A}(x^*)$ such that all Lagrange multipliers satisfying (24)-(26) have $\lambda_i^* = 0$. (Such constraints are weakly active.)

Table of Contents

ITERATIVE SOLUTION OF THE KKT SYSTEM

- An alternative to the direct factorisation techniques discussed is to use an iterative method to solve the KKT system.
- Iterative methods are suitable for solving very large systems and often lend themselves well to parallelization.
- The conjugate gradient (CG) method is not recommended for solving the full system , because it can be unstable on systems that are not positive definite.
- Iterative methods can be derived from the null-space approach by applying the conjugate gradient method to the reduced system (17)

Active-set methods for convex QP

- Convex QP: any local solution is also global.
- Active-set Methods are the most effective methods for small- to medium-scale problems.
- They have properties susch as:
  - efficient detection of unboundedness and infeasibility;
  - accurate estimate (typically) of the optimal active set.
- A brute-force approach to solving the KKT systems for all combinations of active constraints:
  - if the optimal active set $\mathcal{A}(x^*)$ (the active set at the optimal point x^*) was known
  - the solution could be found as the solution of the equality-constrained QP problem

$$\min_x q(x) \text{ s.t. } a_i^T x = b_i, i \in \mathcal{A}(x^*).$$

Active-set method

- start from a guess of the optimal active set;
- if not optimal, drop one index from $\mathcal{A}(x)$ and add a new index (using gradient and Lag. mult. information); repeat.
- The simplex method for LP is an active-set method.
- QP active-set methods may have iterates that aren't vertices of the feasible polytope.
- Three types of active-set methods: primal, dual, and primal-dual.
- We focus on primal methods, which generate iterates that remain feasible wrt the primal problem while steadily decreasing the objective function q.

Primal active-set method

- Primal active-set methods 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 is referred to as the working set and is denoted at the *kth* iterate x_k by $\mathcal{W}_k$.
- An important requirement imposed on $\mathcal{W}_k$, the gradients *ai* of the constraints in the $\mathcal{W}_k$ are linearly independent, even when the full set of active constraints at that point has linearly dependent gradients.

Primal active-set method

- Given an iterate x_k and the working set $\mathcal{W}_k$, first check whether x_k minimises the quadratic q in the subspace defined by the working set.
- If not, compute a step p by solving an equality-constrained QP subproblem in which the constraints corresponding to the working set $\mathcal{W}_k$ are regarded as equalities and all other constraints are temporarily disregarded.
- To express this subproblem in terms of the step p, define

$$p = x - x_k, \qquad g_k = Gx_k + c.$$

- Substituting the above expressions into the objective function q in (19) we get:

$$q(x) = q(x_k + p) = \frac{1}{2}p^T Gp + g_k^T p + \rho_k$$

Primal active-set method

- $\rho_k = \frac{1}{2} x_k^T G x_k + c^T x_k$ is independent of p.
- Therefore we can drop ρ_k from the objective without affecting the solution of the problem.
- The QP subproblem to be solved at the k^{th} iteration is:

$$\min_p = \frac{1}{2} p^T G p + g_k^T p \tag{29}$$

$$\text{subject to } a_i^T p = 0, \quad i \in \mathcal{W}_k. \tag{30}$$

- Denote the solution of the above subproblem as p_k.
- Note that for each $i \in \mathcal{W}_k$, the value of $a_i^T x$ does not change as we move along p_k

$$a_i^T (x_k + \alpha p_K) = a_i^T x_k = b_i \text{ for all } \alpha$$

Primal active-set method

- Since the constraints in W_k were satisfied at x_k, they are also satisfied at $x_k + \alpha p_k$, for any value of α.
- Since G is positive definite, the solution of (29) can be computed by any of the techniques described.
- Suppose for a moment that the optimal p_k form (29) is non-zero, we need to decide how far to move along this direction.
- If $x_k + p_k$ is feasible with respect to all the constraints, we set $x_{k+1} = x_k + p_k$.
- Otherwise, set

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

- Where the step-length parameter α_k is chosen to be the largest value in the range $[0, 1]$ for which all constraints are satisfied.

Primal active-set method

- An explicit definition of α_k can be derived by considering what happens to the constraints $i \notin \mathcal{W}_k$.
- As the constraints $i \in \mathcal{W}_k$ will certainly be satisfied regardless of the choice of α_k.
- If $a_i^T p_k \geq 0$ for some $i \notin \mathcal{W}_k$, then for all $\alpha_k \geq 0$ we have

$$a_i^T(x_k + \alpha_k p_k) \geq a_i^T x_k \geq b_i.$$

- Hence, constraint i will be satisfied for all non-negative choices of the step-length parameter.
- Whenever $a_i^T p_k < 0$ for some $i \notin \mathcal{W}_k$, however we have

$$a_i^T(x_k + \alpha_k p_k) \geq b_i \text{ iff } \alpha_k \leq \frac{b_i - a_i^T x_k}{a_i^T p_k}$$

Primal active-set method

- To maximise the decrease in q, α_k has to be as large as possible in $[0, 1]$, subject to retaining feasibility.
- So we obtain the following expression:

$$\alpha_k \stackrel{\text{def}}{=} \min \left(1, \min_{i \notin \mathcal{W}_k,\ a_i^T p_k < 0} \frac{b_i - a_i^T x_k}{a_i^T p_k} \right). \tag{31}$$

- The constraints with corresponding indices i for which the minimum in (31) is achieved are called the blocking constraints.
- If $\alpha_K = 1$ and no new constraints are active at $x_k + \alpha_k p_k$, then there are no blocking constraints on this iteration.s
- Note that it is quite possible for α_k to be zero, because we could have $a_i^T p_k < 0$ for some constraint i that is active at x_k but not a member of the current working set $\mathcal{W}_k$.

Primal active-set method

- If $\alpha_k < 1$, that is, the step along p_k was blocked by some constraint not in $\mathcal{W}_k$, a new working set $\mathcal{W}_{k+1}$ is constructed by adding one of the blocking constraints to $\mathcal{W}_k$.
- Continue to iterate in this manner, adding constraints to the working set until we reach a point $\hat{x}$ that minimises the quadratic objective function over its current working set $\hat{\mathcal{W}}$.
- It is easy to recognise such a point because the subproblem (29), has solution $p = 0$.
- Since $p = 0$ satisfies the first order optimality conditions for (29), we have:

$$\sum_{i \in \hat{\mathcal{W}}} a_i \hat{\lambda}_i = g = G\hat{x} + c, \tag{32}$$

for some Lagrange multipliers $\hat{\lambda}_i$, $i \in \hat{\mathcal{W}}$.

Primal active-set method

- $\hat{x}$ and $\hat{\lambda}$ satisfy the first KKT condition (24), if we define the multipliers corresponding to the inequality constraints that are not in the working set to be zero.
- Because of the control imposed on the step length, $\hat{x}$ is also feasible with respect to all the constraints, so the second and third KKT conditions (25) and (26) are satisfied at this point.
- Now we examine the signs of the multipliers corresponding to the inequality constraints in the working set, that is, the indices $i \in \hat{\mathcal{W}} \cap \mathcal{I}$.
- If these multipliers are all non-negative, the fourth KKT condition (27) is also satisfied.
- So we conclude that $\hat{x}$ is a KKT point for the original problem.

Primal active-set method

- Since G is positive semi-definite, we have from the previous theorem $\hat{x}$ is a global solution of the main problem.
- $\hat{x}$ is a strict local minimiser and the unique global solution if G is positive definite.
- If, one or more of the multipliers $\hat{\lambda}_j, j \in \hat{\mathcal{W}} \cap \mathcal{I}$, is negative, the condition (27) is not satisfied and the objective function $q(.)$ may be decreased by dropping one of these constraints.
- Thus, we remove an index j corresponding to one of the negative multipliers from the working set and solve a new subproblem for the new step.
- This strategy produces a direction p at the next iteration that is feasible with respect to the dropped constraint.

Primal active-set method

Theorem

Suppose that the point $\hat{x}$ satisfies first-order conditions for the equality-constrained subproblem with working set $\hat{\mathcal{W}}$; that is, equation (32) 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:

$$\min_p \frac{1}{2} p^T G p + (G\hat{x} + c)^T p, \tag{33}$$

$$\text{subject to} \quad a_i^T p = 0, \text{ for all } \quad i \in \hat{\mathcal{W}} \text{ with } i \neq j. \tag{34}$$

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 (33), then we have that $a_j^T p > 0$, and that p is a descent

Primal active-set method

- While any index j for which $\hat{\lambda}_j < 0$ usually will yield a direction p along which the algorithm can make progress, the most negative multiplier is often chosen in practice (and in the algorithm specified below).
- This choice is motivated by the sensitivity analysis, which shows that the rate of decrease in the objective function when one constraint is removed is proportional to the magnitude of the Lagrange multiplier for that constraint.
- As in linear programming, however, the step along the resulting direction may be short (as when it is blocked by a new constraint), so the amount of decrease in q is not guaranteed to be greater than for other possible choices of j.

Primal active-set method

The following theorem shows that whenever p_k is obtained from (29) is nonzero and satisfies second-order sufficient optimality conditions for the current working set, it is a direction of strict descent for $q(.)$.

Theorem

Suppose that the solution p_k of (29) is nonzero and satisfies the second-order sufficient conditions for optimality for that problem. Then the function $q(.)$ is strictly decreasing along the direction p_k .

- When G is positive definite—the strictly convex case—the second-order sufficient conditions are satisfied for all feasible subproblems.
- From the result above that we obtain a strict decrease in $q(.)$ whenever $p_k \neq 0$.

Primal active-set method

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

Compute a feasible starting point x_0;
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 (32),
      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
      $j \leftarrow \arg\min_{j \in \mathcal{W}_k \cap \mathcal{I}} \hat{\lambda}_j$;
      $x_{k+1} \leftarrow x_k$; $\mathcal{W}_{k+1} \leftarrow \mathcal{W}_k \backslash \{j\}$;
  else (* $p_k \neq 0$ *)
    Compute α_k from (31);
    $x_{k+1} \leftarrow x_k + \alpha_k p_k$;
    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)

Primal active-set method

- Various techniques can be used to determine an initial feasible point.
- One such is to use the "Phase I" approach for linear programming.
- No significant modifications are needed to generalise this method from linear programming to quadratic programming.
- A variant here that allows the user to supply an initial estimate $\tilde{x}$ of the vector x: Given $\tilde{x}$, define the following feasibility linear program:

$$\begin{aligned} &\min_{(x,z)} e^T z \\ \text{subject to } & a_i^T x + \gamma_i z_i = b_i, \qquad i \in \mathcal{E}, \\ & a_i^T x + \gamma_i z_i \geq b_i, \qquad i \in \mathcal{I}, \\ & z \geq 0, \end{aligned}$$

Primal active-set method

- $e = (1, 1, ..., 1)^T$, $\gamma_i = -\text{sign}(a_i^T \tilde{x} - b_i)$ for $i \in \mathcal{E}$, and $\gamma_i = 1$ for $i \in \mathcal{I}$.
- Afeasible initial point for this problem is then

$$x = \tilde{x},\ z_i = |a_i^T \tilde{x} - b_i|\ (i \in \mathcal{E}), \quad z_i = \max(b_i - a_i^T \tilde{x}, 0)\ (i \in \mathcal{I}).$$

- It can be verified that if $\tilde{x}$ is feasible for the original problem , then $(\tilde{x}, 0)$ is optimal for the feasibility subproblem.
- In general, if the original problem has feasible points, then the optimal objective value in the subproblem is zero, and any solution of the subproblem yields a feasible point for the original problem.
- The initial working set $\mathcal{W}_0$ for the algorithm can be found by taking a linearly independent subset of the active constraints at the solution of the feasibility problem.

Example

- Consider the following simple 2-dimensional problem:

$$\min_x q(x) = (x_1 - 1)^2 + (x_2 - 2.5)^2 \tag{35}$$

$$\text{subject to} \quad x_1 - 2x_2 + 2 \geq 0, \tag{36}$$

$$-x_1 - 2x_2 + 6 \geq 0, \tag{37}$$

$$-x_1 + 2x_2 + 2 \geq 0, \tag{38}$$

$$x_1 \geq 0, \tag{39}$$

$$x_2 \geq 0. \tag{40}$$

- The constraints are referred by the indices from 1 through 5.
- It is easy to determine a initial feasible point; $x^0 = (2, 0)^T$.
- Constraints 3 and 5 are active at this point, and we set $\mathcal{W}_0 = \{3, 5\}$.

Example

- Note that the choices $\mathcal{W}_0 = \{5\}$ or $\mathcal{W}_0 = \{3\}$ or even $\mathcal{W} = \phi$ are all valid; each choice would lead the algorithm to perform somewhat differently.

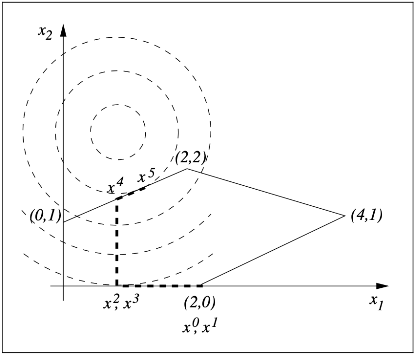

Figure: Iterates of the active-set method

Example

- Since x^0 lies on a vertex of the feasible region, it is obviously a minimiser of the objective function q with respect to the working set $\mathcal{W}_0$; that is, the solution of the subproblem with $k = 0$ is $p = 0$.
- Now (32) can be used to find the multipliers $\hat{\lambda}_3$ and $\hat{\lambda}_5$ associated with the active constraints.
- Substitution of data from the subproblem into (32) yields

$$\begin{bmatrix} -1 \\ 2 \end{bmatrix} \hat{\lambda}_3 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \hat{\lambda}_5 = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$$

- which has the solution $(\hat{\lambda}_3, \hat{\lambda}_5) = (-2, -1)$.

Example

- Now remove the constraint 3 from the working set, as it has the most negative multiplier and set $\mathcal{W}_1 = \{5\}$.
- Begin iteration 1 by finding the solution of the subproblem for $k = 1$, which is $p^1 = (-1, 0)^T$.
- The step-formula (31) yields $\alpha_1 = 1$, and the new iterate is $x^2 = (1, 0)^T$.
- There are no blocking constraints, so $\mathcal{W}_2 = \mathcal{W}_1 = \{5\}$.
- In-turn start the iteration 2 that the solution of the subproblem is $p^2 = 0$.
- From (32), deduce that the Lagrange multiplier for the lone working constraint $\tilde{\lambda}_5 = -5$.
- So drop 5 from the working set to get $\mathcal{W}_3 = \phi$.

Example

- Iteration 3 starts by solving the unconstrained problem, to obtain the solution $p^3 = (0, 2.5)^T$.
- Formula (31) yields a step length of $\alpha_3 = 0.6$ and the new iterate $x^4 = (1, 1.5)^T$.
- There is a single blocking constraint (constraint 1), so we obtain $\mathcal{W}_4 = \{1\}$.
- The solution of the subproblem for $k = 4$ is then $p^4 = (0.4, 0.2)^T$, and the new step length is 1.
- There are no blocking constraints on this step.
- So the next working set is unchanged: $\mathcal{W}_5 = \{1\}$.
- The new iterate is $x^5 = (1.4, 1.7)^T$
- Finally, we solve the subproblem for $k = 5$ to obtain a solution $p^5 = 0$.
- The formula (32) yields a multiplier $\hat{\lambda}_1 = 0.8$, so we have found the solution.
- We set $x^* = (1.4, 1.7)^T$ and terminate.

Frame Title