Quadratical Programming and Active Set Method

LI-PING ZHANG

Department of Mathematical Sciences

Tsinghua University, Beijing 100084

Office: New Science Building #A302, Tel: 62798531

E-mail: lipingzhang@tsinghua.edu.cn

Outline

- Quadratical Programming
- Dual of QP
- Equality-Constrained QP
- Active Set Method

投资回报问题

假设有一笔可用的资金（不妨设为一个单位），全部用来进行项目投资，为了方便分析，我们把第i种项目投资简记为$i(i=1,2,\cdots n)$，并且不考虑投资收益的再投资问题。假设第i种项目投资的回报率为r_i，由于投资收益具有不确定性，所以通常把r_i看成服从正态分布的随机变量，其均值为$\mu_i = E[r_i]$，方差为$\sigma_i^2 = E[(r_i-\mu_i)^2]$，其中$\sigma_i$表示回报率关于均值的偏离程度，实际上反映了投资风险的大小，如何确定一个最优的投资方案，使得在一定的条件下投资回报最大？

设投资者将x_i比例的资金用于第i种投资

$\Rightarrow x = (x_1, x_2, \cdots, x_n)^T$为一种投资方案

对应的回报为 $R = x^T r$

期望回报为 $E[R] = E[x^T r] = x^T \mu$

记i种和j种投资的相关系数为

$$\rho_{ij} = \frac{E[(r_i - \mu_i)(r_j - \mu_j)]}{\sigma_i \sigma_j}, i, j = 1, \cdots, n$$

则投资组合x的回报R的方差为

$$E[(R - E[R])^2] = E\left[\left(\sum_i x_i (r_i - \mu_i)\right)^2\right]$$

$$= \sum_i \sum_j x_i x_j \rho_{ij} \sigma_i \sigma_j \overset{def}{=} x^T Q x$$

$$\max\ x^T\mu - \kappa x^T Qx$$

$$s.t.\quad \sum_{i=1}^{n} x_i = 1$$

$$x_i \geq 0, i = 1, \cdots, n$$

其中$\kappa \geq 0$为风险容忍参数。κ越接近0，说明投资者越具有冒险精神。

Quadratic programming problem

考虑约束的择一系统

$$
(\mathbf{QP}) \quad
\begin{array}{lll}
\textbf{minimize} & f(x) := \frac{1}{2}x^TQx + c^Tx & \\
\textbf{subject to} & a_i^Tx = b_i & i \in \mathcal{E} \\
 & a_i^Tx \geq b_i & i \in \mathcal{I}.
\end{array}
\qquad (1)
$$

It is usually assumed that the matrix Q is symmetric. This is not a restrictive assumption as long as the value of the quadratic form x^TQx is what matters. When $Q = 0$, the objective function reduces to a linear function. There is a significant difference between linear and quadratic functions vis-á-vis convexity. As we know, a function is both convex and concave if and only if it is affine. The objective function could be convex, concave, or neither. In quadratic programming *minimization* problems with convex objective functions, the necessary first-order optimality conditions are also sufficient for optimality. For *maximization* problems, we would want a *concave* objective function.

By Farkas' lemma, we have

Theorem 1 *The feasible region $\mathcal{F}$ of (QP) is nonempty if and only if for any nonzero vector μ with*

利用Farkas引理考虑约束的择一系统

$$\text{若} \quad \sum_{i\in\mathcal{E}\cup\mathcal{I}} \mu_i a_i = 0, \quad \mu_i \geq 0, i \in \mathcal{I},$$

the following inequality holds:

$$\text{则} \quad \sum_{i\in\mathcal{E}\cup\mathcal{I}} \mu_i b_i \leq 0.$$

Dual of QP

Let (QP) be a strictly convex quadratic programming problem. Denote $A := (a_i^T)_{\mathcal{E}\cup\mathcal{I}}$ and $b := (b_i)_{\mathcal{E}\cup\mathcal{I}}$. Its Lagrangian function

$$L(x,\lambda) = \frac{1}{2}x^T Qx + c^T x - \lambda^T(Ax - b).$$

Its Lagrangian dual is: 对凸函数的Lagrangian对偶即为Wolfe对偶

$$\begin{array}{ll} \max & \varphi(\lambda) := -\frac{1}{2}(c - A^T\lambda)^T Q^{-1}(c - A^T\lambda) + b^T\lambda \\ s.t. & \lambda_i \geq 0, i \in \mathcal{I}. \end{array}$$

It can be rewritten as

$$\begin{array}{ll} \max & -\frac{1}{2}\lambda^T(AQ^{-1}A^T)\lambda + (b + AQ^{-1}c)^T\lambda \\ s.t. & \lambda_{\mathcal{I}} \geq 0, \quad \lambda_{\mathcal{E}} \textbf{ free}. \end{array} \tag{2}$$

凸规划时KKT条件即为最优解条件

Theorem 2 *Suppose that the matrix Q in (1) is symmetric and positive definite. If (λ^*, μ^*) is a KKT-pair of (2), then $x^* = Q^{-1}(A^T\lambda^* - c)$ is the (unique) optimal point of (1).*

Proof: Since (λ^*, μ^*) is a KKT-pair of (2),

$$\begin{aligned} & AQ^{-1}A^T\lambda^* - (b + AQ^{-1}c) = \mu^* \\ & \lambda_i^* \geq 0, \mu_i^* \geq 0, \lambda_i^*\mu_i^* = 0, i \in \mathcal{I} \\ & \mu_i^* = 0, i \in \mathcal{E}. \end{aligned} \tag{3}$$

By the first equality in (3),

$$AQ^{-1}(A^T\lambda^* - c) - b = \mu^*, \quad \textbf{i.e.,} \quad Ax^* - b = \mu^*.$$

By the second and the last equations in (3),

$$
\begin{aligned}
&a_i^T x^* - b_i \geq 0, \lambda_i^* \geq 0, \lambda_i^*(a_i^T x^* - b_i) = 0, i \in \mathcal{I} \\
&a_i^T x^* - b_i = 0, i \in \mathcal{E},
\end{aligned}
$$

which, together with $Qx^* + c = A^T\lambda^*$, implies that x^* is a KKT point of (1). $\square$

Duality Gap

Let x and λ be feasible solution to the primal and dual problems (1) and (2), respectively. Denote $y = c - A^T\lambda$, the duality gap is

$$\begin{aligned} f(x) - \varphi(\lambda) &= \frac{1}{2}x^TQx + c^Tx + \frac{1}{2}(c - A^T\lambda)^TQ^{-1}(c - A^T\lambda) - b^T\lambda \\ &= \frac{1}{2}(x^TQx + y^TQ^{-1}y) + c^Tx - b^T\lambda \\ &= \frac{1}{2}(x^TQx + y^TQ^{-1}y) + y^Tx + (Ax - b)^T\lambda \\ &= \frac{1}{2}(Qx + y)^T(x + Q^{-1}y) + (Ax - b)^T\lambda. \end{aligned}$$

By Theorem 2, the duality gap between the primal and dual problems is zero.

Equality-Constrained QP

We begin this algorithmic discussion with equality-constrained quadratic programming problem such as

$$\begin{aligned} \min \quad & \frac{1}{2}x^T Qx + c^T x \\ s.t. \quad & Ax = b. \end{aligned} \tag{4}$$

Let us make assumption that $A \in \mathcal{R}^{m\times n}$ and that A has row full rank.

We may assume, without loss of generality, that

$$A = (A_B \ \ A_N), \quad x = \begin{pmatrix} x_B \\ x_N \end{pmatrix},$$

where A_B is a basis of A. Thus, x_B can be reexpressed as

$$x_B = A_B^{-1}(b - A_N x_N).$$

We also divide Q and c into

$$c = \begin{pmatrix} c_B \\ c_N \end{pmatrix}, \quad Q = \begin{pmatrix} Q_{BB} & Q_{BN} \\ Q_{NB} & Q_{NN} \end{pmatrix},$$

with respect to the basis A_B. Then the problem (4) can be rewritten as the following unconstrained quadratic optimization problem

$$\min_{x_N \in \mathcal{R}^{n-m}} \psi(x_N),$$

where

$$\begin{aligned} \psi(x_N) \quad &= \tfrac{1}{2} x_N^T (Q_{NN} - Q_{NB} A_B^{-1} A_N - (A_B^{-1} A_N)^T Q_{BN} \\ &\quad + (A_B^{-1} A_N)^T Q_{BB} A_B^{-1} A_N) x_N \\ &\quad + x_N^T (Q_{NB} - (A_B^{-1} A_N)^T Q_{BB}) A_B^{-1} b + \tfrac{1}{2} (A_B^{-1} b)^T Q_{BB} A_B^{-1} b \\ &\quad + x_N^T (c_N - (A_B^{-1} A_N)^T c_B) + c_B^T A_B^{-1} b. \end{aligned}$$

The equality-constrained quadratic optimization problem (4) is a matter of minimizing the unconstrained quadratic function $\psi(x_N)$. The first-order optimality condition $\nabla\psi(x_N) = 0$ will have to hold at a local minimizer x_N of ψ. If the Hessian matrix of ψ is positive semidefinite, these necessary conditions will also be sufficient for the optimality of such stationary point of ψ.

This method is elimination method.

Lagrange Method

The KKT condition of QP (4) is written as

$$Qx + c = A^T\lambda, \quad Ax = b,$$

which is just a system of linear equations:

$$\begin{pmatrix} Q & -A^T \\ -A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -c \\ -b \end{pmatrix}. \tag{5}$$

Theorem 3 *Suppose that $A \in \mathcal{R}^{m\times n}$ has row full rank, and $x^TQx > 0$ holds for any nonzero vector x with $Ax = 0$. Then the system (5) has one solution, denoted by (x^*, λ^*). Moreover, x^* is a global optimal solution of (4).*

Proof

The matrix

$$\begin{pmatrix} Q & -A^T \\ -A & 0 \end{pmatrix}$$

is nonsingular. Indeed, suppose $(x, \lambda) \in \mathcal{R}^{n \times m}$ such that

$$Qx - A^T\lambda = 0, \quad Ax = 0.$$ 没有非 解故可逆

Multiplication of the first equation by x^T yields

$$0 = x^T Qx - x^T A^T y = x^T Qx$$

since $Ax = 0$. Hence the hypothesis on Q implies that $x = 0$ and then the full rank condition on A implies that $\lambda = 0$.

Thus the system (5) has one solution. Let x be any feasible point of the QP problem (4) and define $d := x - x^*$. Then $Ad = 0$ and the hypothesis on Q

then implies that

$$d^T Q d > 0.$$

Consider

$$\begin{aligned} f(x) &= \tfrac{1}{2}(x^* + d)^T Q (x^* + d) + c^T (x^* + d) \\ &= f(x^*) + \tfrac{1}{2} d^T Q d + d^T Q x^* + c^T d \\ &> f(x^*) + d^T (A^T \lambda^*) \\ &= f(x^*). \end{aligned}$$

This implies that x^* is a global optimal solution. □

Active set method

The idea underlying active set methods is to partition inequality constraints into two groups: those that are to be treated as active and those that are to be treated as inactive.

The idea of active set method is to define at each step of an algorithm a set of constraints, termed the working set, that is to be treated as the active set. The working set is chosen to be a subset of the constraints that are actually active at the current point, and hence the current point is feasible for the working set. The algorithm then proceeds to move on the surface defined by the working set of constraints to an improved point.

不等式里面的积极约束以及原等式约束

An active set method consists of the following components: (i) determination of a current working set that is a subset of the current active constraints, and (ii) movement on the surface defined by the working set to an improved point.

Theorem 4 *Let x^* be a feasible point for (1) and also be a KKT point of the following problem*

这个问题的kkt
对里面的λ

$$\boxed{\begin{array}{ll} \textbf{min} & \frac{1}{2}x^TQx + c^Tx \\ s.t. & a_i^Tx = b_i, \quad i \in \mathcal{A}(x^*). \end{array}}$$

If there exist $\lambda_i \geq 0,\ i \in \mathcal{I}(x^*)$, then x^* is also a KKT point of (1).

Here

$$\mathcal{I}(x^*) = \{i \in \mathcal{I} : a_i^Tx^* = b_i\}$$

and

$$\mathcal{A}(x^*) = \mathcal{I}(x^*) \cup \mathcal{E}.$$

工作集

For the QP problem (1) with the case where Q is positive definite, at iteration k a point x^k is given that is feasible for all constraints and satisfies all the equality constraints of the current working set W_k. The working set always includes the equality constraints $\mathcal{E}$ and possibly some of the inequality constraints $\mathcal{I}$.

Consider the quadratic program corresponding to the working set is then defined, by translating to the point x^k, in the form

$$\begin{array}{ll} \textbf{min} & \frac{1}{2}d^TQd + g_k^Td \\ s.t. & a_i^Td = 0, \quad i \in W_k, \end{array} \qquad (6)$$

$$\begin{cases} Qd^k + g_k - A_w^T\lambda_w = 0 \\ A_w d^k = 0 \end{cases}$$ KKT条件

where $g_k = c + Qx^k$. Let d^k solve the quadratic program. If $d^k = 0$, the current point x^k is optimal w.r.t the working set W_k. If $d^k \neq 0$, it is a descent direction of $f(x)$.

Theorem 5 *Let d^k solves (6). If $d^k \neq 0$ then $f(x^k + \alpha d^k) < f(x^k)$ holds for any $\alpha \in (0, 1]$.*

proof: Obviously, d^k is an optimal solution for

$$\begin{array}{ll} \textbf{min} & \frac{1}{2}(x^k + d)^T Q(x^k + d) + c^T(x^k + d) \\ s.t. & a_i^T d = 0, \quad i \in W_k. \end{array}$$

Hence

$$f(x^k) > f(x^k + d^k) = \frac{1}{2}(x^k + d^k)^T Q(x^k + d^k) + c^T(x^k + d^k),$$

which implies that

$$\frac{1}{2}(d^k)^T Q d^k + g_k^T d^k < 0.$$

For any $\alpha \in (0, 1]$,

$$
\begin{aligned}
f(x^k + \alpha d^k) &= f(x^k) + \alpha g_k^T d^k + \tfrac{1}{2}\alpha^2 (d^k)^T Q d^k \\
&= f(x^k) + \alpha(g_k^T d^k + \tfrac{1}{2}(d^k)^T Q d^k) + \tfrac{1}{2}(\alpha^2 - \alpha)(d^k)^T Q d^k \\
&< f(x^k).
\end{aligned}
$$

If $d^k \neq 0$, then d^k is a descent direction of $f(x)$ at x^k. If $x^k + d^k$ is feasible, the new point $x^{k+1} = x^k + d^k$is obtained and $W_{k+1} = W_k$. Otherwise, we choose the step size α_k such that $x^k + \alpha_k d^k$ is feasible. Hence α_k satisfies

$$a_i^T(x^k + \alpha_k d^k) \geq b_i, \quad i \in \mathcal{I} \backslash W_k.$$

Define

$$\hat{\alpha}_k := \min_{a_i^T d^k < 0,\; i \in \mathcal{I} \backslash W_k} \left\{ \frac{b_i - a_i^T x^k}{a_i^T d^k} \right\}. \tag{7}$$

Thus, the general move is $x^{k+1} = x^k + \alpha_k d^k$ where $\alpha_k = \min\{1, \hat{\alpha}_k\}$. At this point a new inequality constraint is satisfied by equality, and this constraint is adjoined to the working set W_{k+1}. That is, when $\alpha_k = \hat{\alpha}_k$, there exists $i_0 \in \mathcal{I} \backslash W_k$ such that $a_{i_0}^T(x^k + \alpha_k d^k) = b_{i_0}$, then set $W_{k+1} = W_k \cup \{i_0\}$.

If $d^k = 0$, then (x^k, λ^k) is a KKT-pair of (1) w.r.t. the working set W_k. In this case, if $\lambda_i^k \geq 0, \ \forall i \in W_k \cap \mathcal{I}(x^k)$ then x^k is optimal for (1). Otherwise, there exists $i_0 \in W_k \cap \mathcal{I}(x^k)$ such that $\lambda_{i_0}^k < 0$. Take

$$i_k = \textbf{arg} \min_{i \in W_k \cap \mathcal{I}(x^k)} \{\lambda_i^k \,|\, \lambda_i^k < 0\}, \quad W_{k+1} = W_k \backslash \{i_k\}, \quad x^{k+1} = x^k.$$

Active Set Method for (1)

0. Start with a feasible point x^0 and a working set $W_0 = \mathcal{A}(x^0)$. Set $k = 0$.

1. Solve the equality constrained quadratic program (6). If $d^k = 0$, go to Step 3.

2. Set $x^{k+1} = x^k + \alpha_k d^k$ where $\alpha_k = \min\{1, \hat{\alpha}_k\}$ and $\hat{\alpha}_k$ is defined by (7). If $\alpha_k < 1$, adjoin the minimizing index in (7) to W_k to form W_{k+1}. Set $k = k + 1$ and return to Step 1.

3. Compute the Lagrange multipliers of (6), and then let $\lambda_q = \min_{i \in \mathcal{I} \cap W_k} \lambda_i$. If $\lambda_q \geq 0$, stop; x^k is optimal. Otherwise, drop q from W_k to define $W_{k+1} = W_k \backslash \{q\}$. Set $k = k + 1$ and return to Step 1.

有限步终止

Example

The following is a strictly convex quadratic program with a compact feasible region containing the vector $\mathbf{x}^0 = (0, 0)$.

$$\begin{array}{ll} \min & \frac{1}{2}(2x_1^2 - 4x_1x_2 + 4x_2^2) - 2x_1 - 6x_2 \\ s.t. & -x_1 - x_2 \geq -2 \\ & x_1 - 2x_2 \geq -2 \\ & x_1 \geq 0, \quad x_2 \geq 0. \end{array}$$

Obviously, here

$$Q = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}, \quad c = \begin{pmatrix} -2 \\ -6 \end{pmatrix},$$

$$A = \begin{pmatrix} -1 & -1 \\ 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} -2 \\ -2 \\ 0 \\ 0 \end{pmatrix}, \quad \mathcal{I} = \{1, 2, 3, 4\}.$$

By terms of the KKT conditions of (6) we have

$$d^k = [Q^{-1}A_W^T(A_W Q^{-1} A_W^T)^{-1} A_W - I](x^k + Q^{-1}c)$$

$$\lambda^k = (A_W Q^{-1} A_W^T)^{-1} A_W (x^k + Q^{-1}c).$$

We also have

$$Q^{-1} = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

At the starting point $\mathbf{x}^0 = (0, 0)$,

$$W = \{3, 4\}, \quad A_W = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then we have

$$\mathbf{d} = \mathbf{0}, \quad \lambda = (-2, -6).$$

Since λ is not a nonnegative vector, we drop $q = 4$ from W. Accordingly, we revise some definitions as follows:

$$W = \{3\}, \quad A_W = (1 \quad 0).$$

Thus we have

$$\mathbf{d} = (0, 3/2).$$

The idea now is to move from $\mathbf{x}^0$ to $\mathbf{x}^0 + \alpha\mathbf{d}$ for some $\alpha > 0$.

When $\alpha = 1$ it is easy to check that this point is infeasible. We find a smaller α by enforcing the feasibility condition $A(\mathbf{0} + \alpha \mathbf{d}) \geq b$, which turns out to imply that $\alpha_k = 2/3$. We define the next iterate to be $\mathbf{x}^1 = \mathbf{x}^0 + (2/3)\mathbf{d} = (0, 1)$. For this iteration, we have

$$W = \{2, 3\}, \quad A_W = \begin{pmatrix} 1 & -2 \\ 1 & 0 \end{pmatrix}.$$

Hence,

$$\mathbf{d} = \mathbf{0}, \quad \lambda = (1, -5).$$

Updating, we get $W = \{2\}$ and $A_W = (1 \ -2)$. Thus we get

$$\mathbf{d} = (5, 5/2).$$

TO compute the step length, we look at $A(\mathbf{x}^1 + \alpha \mathbf{d}) \geq b$, which implies that

$$\alpha_k = \min\{1, 2/15\} = 2/15.$$

Using this, we define a new iterate $\mathbf{x}^2 = \mathbf{x}^1 + (2/15)\mathbf{d} = (2/3, 4/3)$. We then have

$$W = \{1, 2\},\ \ A_W = \begin{pmatrix} -1 & -1 \\ 1 & -2 \end{pmatrix},\quad \mathbf{d} = \mathbf{0},\ \ \lambda = (26/9, -4/9).$$

We drop $q = 2$ from W and compute $\mathbf{d} = (2/15, -2/15)$. It is easy to check that $\mathbf{x}^2 + \mathbf{d} = (4/5, 6/5)$ is feasible. Hence we set

$$\mathbf{x}^3 = \mathbf{x}^2 + \mathbf{d} = (4/5, 6/5)$$

and then we have

$$W = \{1\},\ \ A_W = (-1 - 1),\ \ \textbf{and then } \mathbf{d} = \mathbf{0},\ \ \lambda = 14/5.$$

This tells us that $\mathbf{x}^3$ is a local minimizer and then a unique global minimizer since this problem is a strictly convex quadratic program.

Remarks

The sequence $\{x^k\}$ is feasible and $f(x^{k+1}) < f(x^k)$ when $x^{k+1} \neq x^k$. If the active set method terminates in a finite number of iterations then x^k is a KKT point and also optimal solution of the original convex quadratic program. If the active set method generate a sequence, then there exists k_0 such that $f(x^k) = f(x^{k_0})$ when $k \geq k_0$. We consider the following two cases:

(i) There exists a subsequence $\mathcal{N}_0 \subset \mathcal{N}$ such that $d^k = 0, \ \forall k \in \mathcal{N}_0$.

(ii) There exists $k_0 > 0$ such that $f(x^k) = f(x^{k_0}), \ \forall k \geq k_0$.

To prove (i), we will prove that for any $k > 0$ there exists $t_k \geq 0$ such that $d^{k+t_k} = 0$. Indeed, if $d^k = 0$ take $t_k = 0$; otherwise, according to the process of the algorithm, if $\alpha_k = 1 < \hat{\alpha}_k$ then $d^{k+1} = 0$ and so take $t_k = 1$. If $\alpha_k = \hat{\alpha}_k$ then $|W_{k+1}| > |W_k|$ and hence $\{W_k\}$ is monotonically increasing. Since $|W_k| \leq |\mathcal{I}| + |\mathcal{E}|$ the step size is taken 1 after finitely many iterations. In this case we can get t_k such that $d^{k+t_k} = 0$.

Prove (ii). It is clear that the sequence $\{f(x^k)\}$ is nonincreasing. Due to (i) we have x^k is a KKT point of the program (1) w.r.t. the working set W_k for any $k \in \mathcal{N}_0$. Since the number of W_k is finite, $\{f(x^k)\}_{k \in \mathcal{N}_0}$ has finite different values. Hence, there exists $k_0 > 0$ such that $f(x^k) = f(x^{k_0}),\ \forall k \geq k_0$.

Convergence of active set method

We know that there exists $k_0 > 0$ such that $x^k = x^{k_0}, \; \forall k \geq k_0$. We discuss the case where $k \geq k_0$.

Lemma 1 *If $d^{k_1} = d^{k_2} = 0$ and $d^k \neq 0, \; \forall k_1 < k < k_2$, then $W_{k_1} \neq W_{k_2}$.*

Proof: If $k_2 = k_1 + 1$ then $W_{k_2} = W_{k_1} \backslash \{i_{k_1}\} \neq W_{k_1}$. Now we assume that $k_2 > k_1 + 1$. We only need to prove that when $k_2 = k_1 + 2$, outgoing index i_{k_1} in Step $k_1 (d^{k_1} = 0)$ and the ingoing index i_{k_1+1} in Step $k_1 + 1 (\alpha_{k_1+1} = 0)$ is different.

By $d^{k_1} = 0$, x^{k_0} is a KKT point of (1) w.r.t W_{k_1} and then there exists λ^{k_1} such that

$$Qx^{k_0} + c = \sum_{i \in W_{k_1}} \lambda_i^{k_1} a_i, \tag{8}$$

and there exists i_{k_1} such that $\lambda_{i_{k_1}}^{k_1} < 0$, and $W_{k_1+1} = W_{k_1} \backslash \{i_{k_1}\}$.

In Step k_1+1, $\alpha_{k_1+1}=0$, i.e.,

$$\alpha_{k_1+1}=\hat{\alpha}_{k_1+1}=\min_{\substack{a_i^T d^{k_1+1}<0\\ i\in\mathcal{I}\backslash W_{k_1+1}}}\left\{\frac{b_i-a_i^T x^{k_0}}{a_i^T d^{k_1+1}}\right\}=0.$$

Thus there is $i_{k_1+1}\in\mathcal{I}\backslash W_{k_1+1}$ such that $W_{k_1+2}=W_{k_1+1}\cup\{i_{k_1+1}\}$, and

$$a_{i_{k_1+1}}^T d^{k_1+1}<0. \tag{9}$$

Since d^{k_1+1} is optimal for (6) w.r.t W_{k_1+1}, d^{k_1+1} is descent direction of $f(x)$ at point x^{k_0}, i.e.,

$$(Qx^{k_0}+c)^T d^{k_1+1}\le 0,$$

which together with (8) implies that

$$\sum_{i\in W_{k_1}}\lambda_i^{k_1}a_i^T d^{k_1+1}\le 0.$$

Hence

$$\lambda_{i_{k_1}}^{k_1} a_{i_{k_1}}^T d^{k_1+1} \leq 0$$

due to $a_i^T d^{k_1+1} = 0,\ \forall i \in W_{k_1+1}$. Since $\lambda_{i_{k_1}}^{k_1} < 0$,

$$a_{i_{k_1}}^T d^{k_1+1} \geq 0,$$

which together with (9) implies that $i_{k_1} \neq i_{k_1+1}$. Hence, $W_{k_1} \neq W_{k_2}$ since $i_{k_1+1} \in W_{k_1+2} = W_{k_2}$ and $i_{k_1+1} \notin W_{k_1}$.

Theorem 6 *Let* $\{x^k\}$ *be generated by the active set method. If the vectors* $[a_i, \ \ i \in \mathcal{A}(x^k)]$ *are linearly independent, then the algorithm terminates at a KKT point of (1) after a finite number of iterations, or the objective function of (1) is unbounded below.*

Proof: Suppose that the objective function f is bounded below and the infinite sequence $\{x^k\}$ is generated by the active set method. There exists $k_0 > 0$ such that $x^k = x^{k_0}, \ \forall k \geq k_0$. That is, for any $k \geq k_0$, $d^k = 0$ or $\alpha_k = 0$.

Let $\{d^{k_t}\}$ denote the subsequence with $d^k = 0$.

Then for any $t,\ k_{t+1} > k_t + 1$.

Indeed, suppose, reasoning by contradiction, that there is i such that $k_{i+1} = k_i + 1$, $d^{k_i} = d^{k_i+1} = 0$, then $W_{k_i}\backslash\{i_{k_i}\} = W_{k_i+1}$ and there exist $\lambda,\ \gamma$ such that

$$g_{k_0} = \sum_{j \in W_{k_i}} \lambda_j a_j = \sum_{j \in W_{k_i+1}} \gamma_j a_j,$$

and $\lambda_{i_{k_i}} < 0$. This contradicts with the fact that $[a_i,\ \ i \in W_{k_i}]$ are linearly independent. Hence

$$|W_{k_1}| \leq |W_{k_2}| \leq \cdots \leq |W_{k_t}| \leq \cdots.$$

Since $|W_{k_t}| \leq |\mathcal{I} \cup \mathcal{E}|$, when t is sufficiently large $|W_{k_t}|$ does not change, and then $k_{t+1} = k_t + 2$.

Since $d^{k_t} = d^{k_{t+1}} = 0$, x^{k_0} are KKT point of (1) w.r.t. W_{k_t} and $W_{k_{t+1}}$ respectively. There exist $\beta, \ \rho$ such that

$$g_{k_0} = \sum_{j \in W_{k_t}} \beta_j a_j = \sum_{j \in W_{k_{t+1}}} \rho_j a_j. \tag{10}$$

By the above lemma, $W_{k_t} \neq W_{k_{t+1}}$, and hence there exists i_{k_t} such that $i_{k_t} \in W_{k_t}, \ \ i_{k_t} \notin W_{k_{t+1}}$ and $\beta_{i_{k_t}} < 0$. This indicates that (10) contradicts with the hypothesis on $[a_i, \ \ i \in \mathcal{A}(x^k)]$.

Initial feasible solution

Given $\tilde{x}$, we define the following feasibility linear program:

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

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

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