Numerical Optimization

Lecture 17: Quadratic Programming

王浩

信息科学与技术学院

Email: wanghao1@shanghaitech.edu.cn

1、Active-set Method for QPs

Unconstrained quadratic optimization

Consider the quadratic optimization problem

$$\min_{x} \ f(x) = g^T x + \tfrac{1}{2} x^T H x,$$

where H is symmetric. Necessary optimality conditions are

$$\nabla f(x) = g + Hx = 0.$$

If $H \succeq 0$, then f is convex, in which case the above conditions are also sufficient.

Thus, we have the following cases based on g and H:

- If $H \succeq 0$, then any solution to $g + Hx = 0$ is optimal.
- If $H \succeq 0$ and $g + Hx = 0$ has no solution, or if

$$d^T H d < 0 \ \text{ for some } \ d \neq 0,$$

then the problem is unbounded.

投资组合优化(Portfolio Optimization)

投资组合$\{1,\ldots,n\}$, 可能受益为r_i（**random variable**）

对受益与风险的这种进行建模

$$\mu_i = E[r_i], \; \sigma_i^2 = E[(r_i - \mu_i)^2], \; \rho_{ij} = \frac{E[(r_i - \mu_i)(r_j - \mu_j)]}{\sigma_i \sigma_j}$$

假定所有资金均为投资，不允许卖空

投资组合(portfolio): 设对第i项投资的资金投放比例为x_i

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

投资组合的收益(return): $R = \sum_{i=1}^{n} x_i r_i$

投资组合的风险(risk): $E[(R - E[R])^2] = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \sigma_i \sigma_j \rho_{ij} = x^T G x$, G协方差矩阵

投资组合优化(Portfolio Optimization)

投资组合的风险和收益的折衷

$$\min_{x} \quad -\mu^T x + \alpha x^T G x \qquad \textbf{s.t.} \quad \sum_{i=1}^{n} x_i = 1, \quad x \geq 0$$

$$\min_{x} \quad x^T G x \qquad \textbf{s.t.} \quad \sum_{i=1}^{n} x_i = 1, \quad x \geq 0, \quad \mu^T x \geq \bar{r}$$

$$\min_{x} \quad -\mu^T x \qquad \textbf{s.t.} \quad \sum_{i=1}^{n} x_i = 1, \quad x \geq 0, \quad x^T G x \leq \bar{\sigma}$$

Many other variants!

Equality constrained quadratic optimization

Consider the equality constrained quadratic optimization problem

$$\min_x f(x) = g^T x + \tfrac{1}{2} x^T H x$$

$$\text{s.t. } c(x) = Ax + b = 0.$$

The Lagrangian is

$$L(x, \lambda) = g^T x + \tfrac{1}{2} x^T H x + \lambda^T (Ax - b),$$

so necessary optimality conditions are

$$\begin{bmatrix} \nabla_x L(x, \lambda) \\ c(x) \end{bmatrix} = \begin{bmatrix} g + Hx + A^T \lambda \\ Ax + b \end{bmatrix} = 0 \implies \begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = - \begin{bmatrix} g \\ b \end{bmatrix}.$$

▶ The critical cone at (x, λ) is $\{d : \nabla c(x)^T d = 0\} = \{d : Ad = 0\}$.

▶ Thus, the necessary conditions above are sufficient as long as

$$d^T \nabla^2_{xx} L(x, \lambda) d = d^T H d \geq 0 \text{ for all } d \text{ such that } Ad = 0.$$

Equality constrained quadratic optimization

Consider the equality constrained quadratic optimization problem

$$\min_x f(x) = g^T x + \tfrac{1}{2} x^T H x$$

$$\text{s.t. } c(x) = Ax + b = 0.$$

The Lagrangian is

$$L(x, \lambda) = g^T x + \tfrac{1}{2} x^T H x + \lambda^T (Ax - b),$$

so necessary optimality conditions are

$$\begin{bmatrix} \nabla_x L(x, \lambda) \\ c(x) \end{bmatrix} = \begin{bmatrix} g + Hx + A^T \lambda \\ Ax + b \end{bmatrix} = 0 \implies \begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = - \begin{bmatrix} g \\ b \end{bmatrix}.$$

▶ The critical cone at (x, λ) is $\{d : \nabla c(x)^T d = 0\} = \{d : Ad = 0\}$.

▶ Thus, the necessary conditions above are sufficient as long as

$$d^T \nabla^2_{xx} L(x, \lambda) d = d^T H d \geq 0 \text{ for all } d \text{ such that } Ad = 0.$$

Inequality constrained EPs

Now we also consider inequalities in this problem

$$\min_x f(x) = g^T x + \tfrac{1}{2} x^T H x$$

$$\text{s.t.} \quad A^i x + b^i = 0, \quad i \in \mathscr{E}$$

$$A^i x + b^i \leq 0, \quad i \in \mathscr{I}$$

Optimal active-set

If an optimal active-set $\mathcal{A}_*$ (i.e., a set of inequalities satisfied as equalities at a solution) is known in advance, then a solution x_* can be found as a solution to

$$\min_x g^T x + \tfrac{1}{2} x^T H x$$

$$\text{s.t. } A^i x + b^i = 0, \ i \in \mathcal{E} \cup \mathcal{A}_*,$$

i.e., a solution to

$$g + Hx + \sum_{i \in \mathcal{E} \cup \mathcal{A}_*} A^{i^T} \lambda^i = 0$$

$$A^i x + b^i = 0, \ i \in \mathcal{E} \cup \mathcal{A}_*.$$

Active-set iteration

Suppose we have an iterate x_k and a guess $\mathcal{A}_k$ of an optimal active set.

Compute d_k as the solution to the subproblem

$$\min_{d} \; g^T(x_k + d) + \tfrac{1}{2}(x_k + d)^T H(x_k + d)$$

$$\text{s.t. } A^i(x_k + d) + b^i = 0, \; i \in \mathcal{E} \cup \mathcal{A}_k.$$

▶ If $x_k + d_k$ is feasible, then set $x_{k+1} \leftarrow x_k + d_k$ and let $\mathcal{A}_{k+1} \leftarrow \mathcal{A}_k$.

▶ Else, set $x_{k+1} \leftarrow x_k + \alpha_k d_k$, where α_k is the largest value such that x_{k+1} satisfies all constraints. Let $\mathcal{A}_{k+1}$ be the set of constraints active at x_{k+1}.

Continue this process until $d_k = 0$ for some k...

证明：d_k=0时，$\lambda_k^j \geq 0, \, j \in \mathscr{A}_k$，$x_k$达到原问题最优

Optimality check

Eventually, we obtain a solution (x_k, λ_k) of the KKT conditions

$$g + Hx + \sum_{i \in \mathcal{E} \cup \mathcal{A}_k} A^{i^T} \lambda^i = 0$$

$$A^i x + b^i = 0, \ i \in \mathcal{E} \cup \mathcal{A}_k$$

(with $\lambda^i = 0$ for $i \in \mathcal{I} \backslash \mathcal{A}_k$). The KKT conditions for the quadratic problem

$$g + Hx + A^{\mathcal{E}^T} \lambda^{\mathcal{E}} + A^{\mathcal{I}^T} \lambda^{\mathcal{I}} = 0$$

$$A^{\mathcal{E}} x + b^{\mathcal{E}} = 0$$

$$A^{\mathcal{I}} x + b^{\mathcal{I}} \leq 0$$

$$\lambda^{\mathcal{I}} \geq 0$$

$$\lambda^{\mathcal{I}} \cdot (A^{\mathcal{I}} x + b^{\mathcal{I}}) = 0$$

will be satisfied as long as x_k is feasible and $\lambda_k^i \geq 0, \ i \in \mathcal{A}_k$.

数值最优化 非线性规划 **ShanghaiTech-SIST-CS**

Finding an improving direction d_k=0时，如果没有达到最优

Suppose a solution x_k to

$$\min_x \ g^T x + \tfrac{1}{2} x^T H x$$

$$\text{s.t. } A^i x + b^i = 0, \ i \in \mathcal{E} \cup \mathcal{A}_k$$

is feasible, but not optimal for the linear optimization problem.

▶ Consider $j \in \mathcal{A}_k$ such that $\lambda_k^j < 0$.

▶ An improving direction is obtained by considering the problem

$$\min_d \ g^T (x_k + d) + \tfrac{1}{2} (x_k + d)^T H (x_k + d)$$

$$\text{s.t. } A^i (x_k + d) + b^i = 0, \ i \in \mathcal{E} \cup (\mathcal{A}_k \backslash j),$$

or, equivalently,

$$\min_d \ (g + H x_k)^T d + \tfrac{1}{2} d^T H d$$

$$\text{s.t. } A^i d = 0, \ i \in \mathcal{E} \cup (\mathcal{A}_k \backslash j).$$

▶ If this problem is unbounded or has a solution $d \neq 0$, then such a d with

$$g^T d < 0 \ \text{ and } \ A^i d = 0, \ i \in \mathcal{E} \cup (\mathcal{A}_k \backslash j),$$

is an improving direction from x_k.

Improving directions and feasibility

We don't worry about the constraints,

$$A^i x + b^i < 0, \ i \in \mathcal{I} \backslash \mathcal{A}_k$$

since we maintain feasibility for these constraints for any small displacement

What about the constraint *j*? Could it be violated?

The answer is that the new direction is feasible for this constraint.

Summary of active set method

Let $\mathcal{A}_k$ be a guess of the optimal active set corresponding to a feasible x_k.

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

1. Solve the active-set (equality constrained) QOP to obtain (d_k, λ_k)
2. If $d_k = 0$ and $\lambda_k^{\mathcal{I}} \geq 0$, then stop; x_k is optimal
3. If $d_k \neq 0$ and $x_k + d_k$ is feasible, then set

$$x_{k+1} \leftarrow x_k + d_k \quad \text{and} \quad \mathcal{A}_{k+1} \leftarrow \mathcal{A}_k$$

   and go to step 1

4. If $d_k \neq 0$ and $x_k + d_k$ is infeasible (for any constraint), then set α_k as the largest value such that $x_{k+1} \leftarrow x_k + \alpha_k d_k$ is feasible, set $\mathcal{A}_k$ as the active set at x_{k+1}, and go to step 1

5. Choose any j such that $\lambda_k^j < 0$ and set $\mathcal{A}_{k+1} \leftarrow \mathcal{A}_k \backslash j$
6. Return to step 1

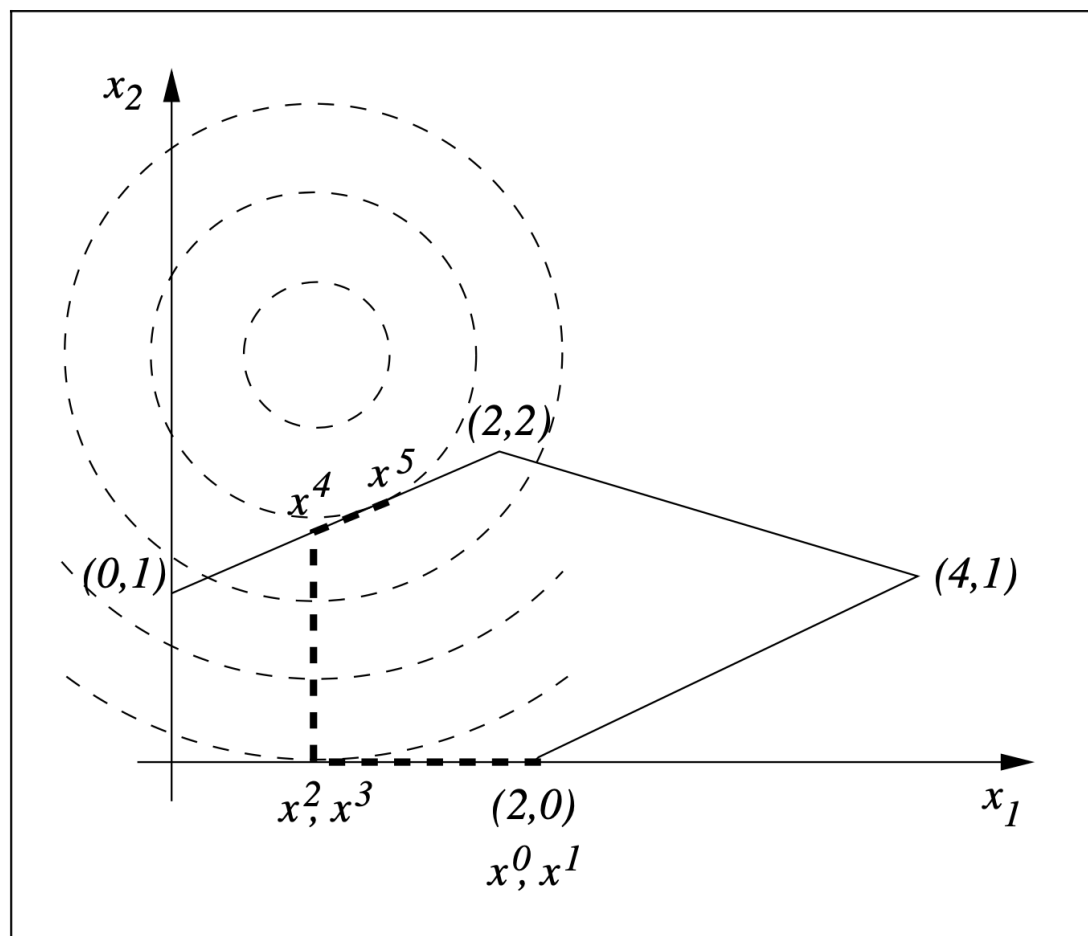$$\min_{x} \ q(x) = (x_1 - 1)^2 + (x_2 - 2.5)^2$$

$$\text{subject to} \quad x_1 - 2x_2 + 2 \geq 0,$$
$$-x_1 - 2x_2 + 6 \geq 0,$$
$$-x_1 + 2x_2 + 2 \geq 0,$$
$$x_1 \geq 0,$$
$$x_2 \geq 0.$$

$$x^{(0)} = (2, 0)^{\mathrm{T}}$$
$$\mathcal{A} = \{3, 5\}$$

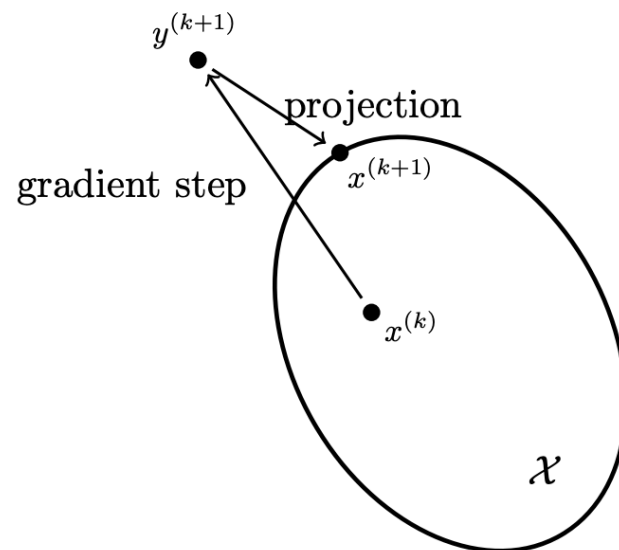**Figure 16.3** Iterates of the active-set method.

2、Gradient Projection Methods

Projected Gradient Descent

Idea: make sure that points are feasible by projecting onto $\mathcal{X}$

Algorithm:

- $y^{(k+1)} = x^{(k)} - t^{(k)} g^{(k)}$
  where $g^{(k)} \in \partial f(x^{(k)})$

- $x^{(k+1)} = \Pi_{\mathcal{X}}(y^{(k+1)})$

The projection operator $\Pi_{\mathcal{X}}$ onto $\mathcal{X}$:

$$\Pi_{\mathcal{X}}(x) = \min_{z \in \mathcal{X}} \|x - z\|$$

Notice: subgradient instead of gradient (even for differentiable functions)

投影算子（Projection Operator）

$$\min_x \quad q(x) = \tfrac{1}{2}x^T G x + x^T c$$
$$\text{subject to} \quad l \le x \le u,$$

- The projection operator:

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

条件梯度法(Conditional Gradient Method)
Frank-Wolfe Algorithm

A projection-free algorithm!
Introduced for QP by Marguerite **Frank** and Philip **Wolfe** (1956)

Algorithm

- Initialize: $x^{(0)} \in \mathcal{X}$

- $s^{(k)} = \underset{s \in \mathcal{X}}{\operatorname{argmin}} \langle \nabla f(x^{(k)}), s \rangle$

- $x^{(k+1)} = x^{(k)} + t^{(k)}(s^{(k)} - x^{(k)})$