Chapter 6

最优化计算方法

6.1 最优化计算方法知识点总结

最优化计算方法

Chapter I. Introduction.

Convexity.

convex set: S, $\forall x,y \in S \Rightarrow \alpha x+(1-\alpha)y \in S$, $\forall \alpha \in [0,1]$.

convex function: f. its domain S is convex.

$\forall x,y \in S \Rightarrow f(\alpha x+(1-\alpha)y) \le \alpha f(x)+(1-\alpha)f(y)$.

convex programming.

$\min f$ — convex

s.t. $c_i(x)=0$, $i\in\mathcal{E}$ — linear

$c_i(x)\geq 0$, $i\in\mathcal{I}$ — concave

Rates of convergence

Q- Linear $\dfrac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}\le r$, $r\in(0,1)$

Q- superlinear $\lim\limits_{k\to\infty}\dfrac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}=0$

Q- Quadratic $\dfrac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|^2}\le M$, $M>0$.

Chapter II. Unconstrained Optimization.

$\min\limits_{x\in\mathbb{R}^n} f(x)$.

Taylor's theorem.

If $f\in C^1$, $p\in\mathbb{R}^n$, $f(x+p)=f(x)+\nabla f(x+tp)^Tp$, $t\in(0,1)$

If $f\in C^2$, $\begin{cases}\nabla f(x+p)=\nabla f(x)+\int_0^1\nabla^2 f(x+tp)p\,dt \\ f(x+p)=f(x)+\nabla f(x)^Tp+\frac{1}{2}p^T\nabla^2 f(x+tp)p\end{cases}$, $t\in[0,1]$

First-order necessary conditions:

x^*: a local minimizer, $f \in C^1$ in an open neighborhood.

$\nabla f(x^*) = 0$.

Second-order necessary conditions

x^*: a local minimizer. $\nabla^2 f(x)$ exists and continous in neighborhood of x^*.

$\nabla f(x^*) = 0$, $\nabla^2 f(x^*)$ is positive semidefinite.

Stationary point. if $\nabla f(x^*) = 0$.

Local minimizer is stationary point.

Sufficient conditions

$\nabla^2 f(x)$ is continous, $\nabla f(x^*) = 0$, $\nabla^2 f(x^*)$ positive definite.

$\Rightarrow$ x^* is a strict local minimizer.

Convex optimization

f is convex $\Rightarrow$ any local minimizer is a global minimizer

In addition, f is continuous $\Rightarrow$ any stationery point is global.

Chapter 3. Line search methods.

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

step length search direction

- Steepest descent direction

$$f(x_k+\alpha p) = f(x_k) + \alpha p^T \nabla f_k + \frac{\alpha^2}{2} p^T \nabla^2 f(x_k+tp)p$$

$$\Rightarrow p = -\frac{\nabla f_k}{\|\nabla f_k\|} . \quad (p^T \nabla f_k < 0)$$

- Newton direction.

$$p_k^N = -(\nabla^2 f_k)^{-1} \nabla f_k \qquad \text{(more accurate)}$$

Quadratic convergence rate (only local)

modification: $p^N = -(\nabla^2 f_k + \sigma I)^{-1} \nabla f_k$

- Quasi-Newton direction

$$B_{k+1} s_k = y_k$$

Different conditions on B_{k+1} $\Rightarrow$ different quasi-Newton.

BFGS: $$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} .$$

$$p_k = -B_k^{-1} \nabla f_k .$$

- exact line search

$$\alpha_k = \arg\min_{\alpha>0} f(x_k+\alpha p_k) = \phi(\alpha)$$

$$\Rightarrow \phi'(\alpha) = 0 \Rightarrow p_k^T \nabla f(x_k+\alpha_k p_k) = 0 .$$

- inexact line search

sufficient decrease condition

(1) $f(x_k+\alpha_k p_k) \le f(x_k) + c_1 \alpha_k \nabla f_k^T p_k$, $c_1 \in (0,1)$.

Curvature condition (avoid short steplength)

② $\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k$, $c_2 \in (c_1, 1)$

① + ② $\Rightarrow$ Wolfe Conditions.

- Strong Wolfe Conditions:

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k$$

$$|\nabla f(x_k + \alpha_k p_k)^T p_k| \leq c_2 |\nabla f_k^T p_k| \qquad 0 < c_1 < c_2 < 1$$

(close to stationary point of ϕ).

- Goldstein Conditions.

$$f(x_k) + (1-c)\alpha_k \nabla f_k^T p_k \leq f(x_k + \alpha_k p_k) \leq f(x_k) + c\alpha_k \nabla f_k^T p_k.$$

$(0 < c < \frac{1}{2})$

- Convergence of line search methods

f bounded below, $f \in C^1$, α_k: Wolfe Conditions,

∇f: Lipschitz continous, $\|\nabla f(x) - \nabla f(\tilde{x})\| \leq L\|x - \tilde{x}\|$.

$\Rightarrow \sum_{k \geq 0} \cos^2\theta_k \cdot \|\nabla f_k\|^2 < \infty$, Zoutendijk condition

Chapter 4. Trust-Region Methods

model function $f(x_k + p) = f_k + g_k^T p + \frac{1}{2} p^T \nabla^2 f(x_k + tp) p$

$$m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p.$$

$m_k(p) - f(x_k + p) = O(\|p\|^2) = O(\|p\|^3)$ if $B_k = \nabla^2 f(x_k)$.

· The cauchy point s_k^c.

$$s_k^c = \arg\min m_k(p)$$
$$\text{s.t. } p = -\tau \nabla f_k$$
$$\|p\| \le \Delta_k, \ \tau \ge 0.$$

$$\Rightarrow s_k^c = -\tau_k \frac{\Delta_k}{\|\nabla f_k\|} \nabla f_k.$$

where
$$\tau_k = \begin{cases} 1 & \text{if } \nabla f_k^T B_k \nabla f_k < 0 \\ \min\left(\frac{\|\nabla f_k\|^3}{\nabla f_k^T B_k \nabla f_k}, 1\right), & \text{o.w} \end{cases}$$

· Dogleg method:

$$p^B = -B^{-1} g, \ \text{if}$$

$$\tilde{p}(\tau) = \begin{cases} \tau p^U & 0 \le \tau \le 1 \\ p^U + (\tau - 1)(p^B - p^U) & , 1 \le \tau \le 2 \end{cases}$$

where $p^U = -\frac{g^T g}{g^T B g} g$

Chapter 5. Conjugate Gradient Methods

· Conjugate Direction Method.

n-dim problem $\Rightarrow$ n 1-dim problems

Theorem 1. $\{x_k\}$ is generated by the Conjugate Direction algorithm, $r_k^T p_i = 0$, $i = 0, \cdots, k-1$.

x_k minimizes $\phi(x)$ over $\{x : x = x_0 + \text{span}\{p_0, \cdots, p_{k-1}\}\}$

How to choose $\{p_0, \cdots, p_{n-1}\}$:

① eigenvectors, $v_1, \cdots, v_n$. $\begin{cases} v_i^T v_j = 0 \\ v_i^T A v_j = 0 \end{cases}$ $i \neq j$

② Gram - Schmidt orthognalization process

、Conjugate Gradient Method.

Alg. Given x_0

set $r_0 = Ax_0 - b$, $p_0 = -r_0$, $k=0$

While $r_k \neq 0$, do:

$\alpha_k = -\dfrac{r_k^T p_k}{p_k^T A p_k}$, $r_k^T p_k = r_k^T(-r_k + \beta_k p_{k-1}) = -r_k^T r_k$

$r_k^T p_{k-1} = 0$ from Theorem

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

$r_{k+1} = Ax_{k+1} - b$, $r_{k+1} = A(x_k + \alpha_k p_k) - b = r_k + \alpha_k A p_k$.

$\beta_{k-1} = \dfrac{r_{k+1}^T A p_k}{p_k^T A p_k}$. $\beta_{k+1} = \dfrac{r_{k+1}^T (r_{k+1} - r_k)}{p_k^T (r_{k+1} - r_k)} = \dfrac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$.

$p_{k+1} = -r_{k+1} + \beta_{k+1} p_k$.

$k := k+1$.

end.

Theorem 2: $\{x_k\}$ generated by CG method, $x_k \neq x^*$.

$r_k^T r_i = 0$, $\forall i = 0, \cdots, k-1$.

$\text{span}\{r_0, \cdots, r_k\} = \text{span}\{r_0, Ar_0, \cdots, A^k r_0\}$

$\text{span}\{p_0, \cdots, p_k\} = \text{span}\{r_0, Ar_0, \cdots, A^k r_0\}$

$p_k^T A p_i = 0$, $\forall i = 0, \cdots, k-1$.

$\{x_k\} \to x^*$ in at most n steps.

- Convergence Rate
- Preconditioning
- Nonlinear Conjugate Gradient Method

β_{k+1}

	$\nabla f_{k+1}^T \nabla f_{k+1}$	$\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)$
$\nabla f_k^T \nabla f_k$	FR	PR
$p_k^T (\nabla f_{k+1} - \nabla f_k)$	DY	HS

Chapter 6. Quasi-Newton Method

(require only gradient)

$p_k = -B_k^{-1} \nabla f_k$

- DFP:
$$\min \|B - B_k\| \quad \text{s.t. } B = B^T,\ B s_k = y_k$$

$$B_{k+1} = (I - \gamma_k y_k s_k^T) B_k (I - \gamma_k y_k s_k^T) + \gamma_k y_k y_k^T$$
$$\gamma_k = \frac{1}{y_k s_k^T}$$

- SMW:
$$H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k}.$$

- BFGS:
$$\min \|H - H_k\|_W \quad \text{s.t. } H = H^T \quad H y_k = s_k.$$
$$H_{k+1} = (I - \rho_k s_k y_k^T) H_k (I - \rho_k s_k y_k^T) + \rho_k s_k s_k^T, \quad \rho_k = \frac{1}{y_k^T s_k}.$$

- Symmetric rank-1 method

$$B_{k+1} = B_k + \sigma v v^T, \quad \sigma = \pm 1$$

s.t. $y_k = B_{k+1} s_k$.

- prevent the breakdown and numerical instabilities skipping rule.

- Nonsymmetric rank-1 method.

$$B_{k+1} = B_k + u v^T.$$

s.t. $B_{k+1} s_k = y_k$

- Broyden class

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} + \phi_k (s_k^T B_k s_k) v_k v_k^T.$$

$$v_k = \frac{y_k}{y_k^T s_k} - \frac{B_k s_k}{s_k^T B_k s_k}.$$

$$\begin{cases} \text{BFGS}: & \phi_k = 0 \\ \text{DFP}: & \phi_k = 1. \end{cases}$$

$$B_{k+1} = (1-\phi_k) B_{k+1}^{BFGS} + \phi_k B_{k+1}^{DFP}.$$

Chapter 8. Least-Squares Problems

LS problem: $\min f(x) = \frac{1}{2} \sum_{j=1}^{m} r_j^2(x)$.

$$f = \frac{1}{2} \sum_{j=1}^{m} r_j^2(x)$$

$$\nabla f(x) = \sum_{j=1}^{m} r_j(x) \nabla r_j(x) = J(x)^T r(x)$$

$$\nabla^2 f(x) = \sum_{j=1}^{m} \nabla r_j(x) \nabla r_j(x)^T + \sum_{j=1}^{m} r_j(x) \nabla^2 r_j(x)$$

$$= J(x)^T J(x) + \sum_{j=1}^{m} r_j(x) \nabla^2 r_j(x).$$

* Linear Least-Squares Problems (LLS)

$$f(x) = \frac{1}{2}\|Jx - y\|_2^2$$

$$\nabla f(x) = J^T(Jx-y), \quad \nabla^2 f(x) = J^T J.$$

$$\boxed{\nabla f(x^*) = 0 \iff J^T J x^* = J^T y.} \longrightarrow \text{Normal Equations}$$

$m \geq n$, J full column rank

Direct approaches:

- Cholesky-based algorithm $\begin{cases} \text{compute } J^TJ \text{ and } J^Ty \\ \text{compute } J^TJ = LL^T \\ LL^T x = Jy \Rightarrow x^* \end{cases}$

- QR-based algorithm

$$J\Pi = Q\begin{bmatrix} R \\ 0 \end{bmatrix} = [Q_1\ Q_2]\begin{bmatrix} R \\ 0 \end{bmatrix} = Q_1 R$$

$$x^* = \pi R^{-1} Q^T y, \quad x^* = \pi z, \ z \text{ solve } Rz = Q^T y$$

- SVD algorithm:
- Iterative approaches.

* Nonlinear LS problems

$$\min f(x) = \frac{1}{2}\|r(x)\|^2$$

$$\nabla f(x) = J(x)^T r(x)$$

$$\nabla^2 f(x) = J(x)^T J(x) + \sum_{j=1}^{m} r_j(x) \nabla^2 r_j(x).$$

- Newton step: $\nabla^2 f(x) p = -\nabla f(x_k)$
- Gauss-Newton step: $J_k^T J_k p^{GN} = -J_k^T r_k$

  (p_k^{GN} solves $\min \frac{1}{2}\|J_k p + r_k\|^2$)

Nonlinear LS → Linear LS

$\min \frac{1}{2}\|r(x)\|^2 \to \min \frac{1}{2}\|r_k + J_k p\|^2$

$\downarrow$

p_k^{GN}

(Similar to Newton method if close to x^* or r_i is close to linear.)

- Levenberg-Marquardt method
  trust-region framework.
  works if $J(x)$ is rank-deficient or nearly so.

$$\min \frac{1}{2}\|r_k + J_k p\|^2 \quad \text{s.t. } \|p\| \le \Delta_k.$$

Solution p^{LM}:
$$\begin{cases} (J^T J + \lambda I)p^{LM} = -J^T r \\ \lambda(\Delta - \|p^{LM}\|) = 0 \\ \|p^{LM}\| \le \Delta, \quad \lambda \ge 0. \end{cases}$$

- Truncated CG algorithm
- Large-residual problems.

Chapter 9. Nonlinear Equations

- $r(x) = 0$
  $r(x) = (r_1(x), \dots, r_n(x))^T$, $r_i : \mathbb{R}^n \to \mathbb{R}$

- Newton's Method
  $r(x+p) = r(x) + \int_0^1 J(x+tp)p\,dt$.

Chapter 10. Theory of Constrained Optimization

$$\min f(x)$$
$$\text{s.t. } c_i(x)=0, \quad i\in\mathcal{E}.$$
$$c_i(x)\geq 0, \quad i\in\mathcal{I}.$$

feasible set

$$\Omega=\{x: c_i(x)=0,\ i\in\mathcal{E};\ c_i(x)\geq 0,\ i\in\mathcal{I}\}$$

- local solution : $f(x^*)\leq f(x)$. $\forall x\in\mathcal{N}\cap\Omega$.
- strict local solution : $f(x^*)<f(x)$. $\forall x\in\mathcal{N}\cap\Omega$.
- isolated local solution : x^* is the only local solution.

Active set

$$\mathcal{A}(x)=\mathcal{E}\cup\{i\in\mathcal{I}: c_i(x)=0\}$$

index i is active, if $\mathcal{E}_i(x)=0$.

linearized feasible directions

$$\mathcal{F}(x)=\left\{d: \begin{array}{ll} d^T c_i(x)=0, & i\in\mathcal{E} \\ d^T\nabla c_i(x)\geq 0, & i\in\mathcal{A}(x)\cap\mathcal{I}\end{array}\right\}$$

Constraints Qualifications: $\mathcal{F}(x)\approx T_\Omega(x)$.

LICQ: Linear Independence Constraint Qualification.

$\{\nabla c_i(x);\ i\in\mathcal{A}(x)\}$ linearly independent

Lagrange function:

$$L(x;\lambda)=f(x)-\sum_{i\in\mathcal{E}\cup\mathcal{I}}\lambda_i c_i(x).$$

First order necessary conditions (KKT conditions)

x^* : local solution,

LICQ holds at x^*, $\exists \lambda^*$.

(optimality) (complementary) (Feasibility)

$$\begin{cases} \nabla_x L(x^*, \lambda^*) = 0 \\ \lambda_i^* c_i(x^*) = 0 \\ c_i(x^*) = 0,\ i \in \mathcal{E};\ c_i(x^*) \geq 0,\ i \in \mathcal{I},\ \lambda_i^* \geq 0,\ \forall i \in \mathcal{I}. \end{cases}$$

Critical cone

$$C(x^*, \lambda^*) = \{ w \in \mathcal{F}(x^*) : \nabla c_i(x^*)^T w = 0,\ \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0 \}$$

$w \in C(x^*, \lambda^*)$

$$\Leftrightarrow \begin{cases} \nabla c_i(x^*)^T w = 0, & \forall i \in \mathcal{E} \\ \nabla c_i(x^*)^T w = 0, & \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0 \\ \nabla c_i(x^*)^T w \geq 0, & \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* = 0 \end{cases}$$

$$w \in C(x^*, \lambda^*) \Rightarrow \lambda_i^* \nabla c_i(x^*)^T w = 0,\ \forall i \in \mathcal{E} \cup \mathcal{I}$$
$$\Rightarrow w^T \nabla f(x^*) = 0.$$

Example

$$\min x_1$$
$$\text{s.t. } x_2 \geq 0$$
$$1 - (x_1 - 1)^2 - x_2^2 \geq 0$$

KKT: $\begin{pmatrix} 1 \\ 0 \end{pmatrix} - \lambda_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \lambda_2 \begin{pmatrix} -2(x_1-1) \\ -2x_2 \end{pmatrix} = 0 = \begin{pmatrix} 1 + 2\lambda_2(x_1 - 1) \\ -\lambda_1 + 2\lambda_2 x_2 \end{pmatrix}$

$$\lambda_1 x_2 = 0, \quad \lambda_2 (1 - (x_1-1)^2 - x_2^2) = 0$$

$$x_2 \geq 0, \quad 1 - (x_1 - 1)^2 - x_2^2 \geq 0, \quad \lambda_1, \lambda_2 \geq 0.$$

$$\lambda^* = \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix}, \quad x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$\mathcal{F}(x^*) = \left\{ d : \begin{array}{l} d^T \begin{pmatrix} 0 \\ 1 \end{pmatrix} \geq 0 \\ d^T \begin{pmatrix} 2 \\ 0 \end{pmatrix} \geq 0 \end{array} \right\} = \{ d : d \geq 0 \}$$

$$C(x^*, \lambda^*) = \left\{ w \geq 0,\ w^T \begin{pmatrix} 2 \\ 0 \end{pmatrix} = 0 \right\} = \{ (0, w_2)^T : w_2 \geq 0 \}$$

- Second-order necessary conditions

  x^*: local solution, LICQ holds at x^*, KKT conditions hold.

  $\Rightarrow w^T \nabla^2_{xx} L(x^*, \lambda^*) w \geq 0, \quad \forall w \in C(x^*, \lambda^*)$

- Second-order sufficient conditions.

  x^*, λ^* : KKT conditions hold, $w^T \nabla^2_{xx} L(x^*, \lambda^*) w > 0$

  $\Rightarrow x^*$ is a strict local solution.

- Second-order necessary condition:

  $z^T \nabla^2_{xx} L(x^*, \lambda^*) z \geq 0$

- Second-order sufficient condition:

  $z^T \nabla^2_{xx} L(x^*, \lambda^*) z > 0$.

- Duality.

  (P) $\min f(x)$ s.t. $c(x) \geq 0$. Convex prob

  (D) $\max q(\lambda) = \inf_x L(x, \lambda)$ s.t. $\lambda \geq 0$.

- Quadratic Programming | Wolfe Dual:

  (P) $\min \frac{1}{2} x^T G x + c^T x$ s.t. $Ax - b \geq 0$ | $\max_{(\lambda, x)} \frac{1}{2} x^T G x + c^T x - \lambda^T (Ax - b)$ s.t. $Gx + c - A^T \lambda = 0,\ \lambda \geq 0$.

Chapter 11. Fundamentals of Algorithms for Nonlinear Constrained Opt.

$$\min f(x)$$
$$\text{s.t. } c_i(x) = 0,\ i \in \mathcal{E}$$
$$c_i(x) \geq 0,\ i \in \mathcal{I}.$$

- LP, QP, NLP, LCO, BCO, CP.

KKT:
$$\nabla_x L(x^*, \lambda^*) = 0$$
$$c_i(x^*) = 0,\ i \in \mathcal{E}$$
$$c_i(x^*) \geq 0,\ i \in \mathcal{I}$$
$$\lambda_i \geq 0,\ i \in \mathcal{I}$$
$$\lambda_i^* c_i(x^*) = 0,\ i \in \mathcal{E} \cup \mathcal{I}.$$

active constraints:

hard constraints: must be satisfied

soft constraints: can be penalized

Merit function

$\begin{cases} \text{reduce objective } f \\ \text{satisfy constraints } c \end{cases}$

- ℓ_1-penalty function
$$\phi_1(x;\mu) = f(x) + \mu \sum_{i \in \mathcal{E}} |c_i(x)| + \mu \sum_{i \in \mathcal{I}} |c_i(x)|^-$$
- exact merit function

$\exists \mu^*$, for any $\mu > \mu^*$, only local solution of original problem is a local minimizer of $\phi(x;\mu)$.

ℓ_1 merit function is exact for $\mu \geq \mu^*$, with
$$\mu^* = \max\{|\lambda_i^*|\ \ i \in \mathcal{E} \cup \mathcal{I}\}$$
λ_i^*: optimal multiplier.

、 l_2 merit function (equality case)

$$\phi_2(x;\mu) = f(x) + \mu \| c(x) \|_2$$

、 Fletcher's augmented Lagrangian (equality case)

$$\phi_F(x;\mu) = f(x) - \lambda(x)^T c(x) + \frac{1}{2}\mu \sum_{i\in\mathcal{E}} c_i(x)^2 .$$

$\mu > 0$, $\lambda(x) = [A(x)A(x)^T]^{-1} A(x) \nabla f(x)$. $A^T(x)\lambda(x) = \nabla f(x)$.

、 augmented Lagrangian function

$$L_A(x,\lambda;\mu) = f(x) - \lambda^T c(x) + \frac{1}{2}\mu \|c(x)\|^2$$

Chapter 12. Quadratic Programming

(QP). $\min \; q(x) = \frac{1}{2} x^T G x + x^T c$

s.t. $a_i^T x = b_i, \quad i \in \mathcal{E}$

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

if $G \succeq 0$, convex QP.

、 Equality-constrained QP.

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

$$\text{s.t. } Ax = b.$$

$A \in \mathbb{R}^{m\times n}$ $(m \le n)$. A: full row rank.

KKT: $\begin{cases} Gx + c = A^T\lambda \\ Ax = b. \end{cases}$

、 if $Z^T G Z$ has zero eigenvalue, x^* is local minimizer, but not strict.

、 if $Z^T G Z$ has negative eigenvalue, x^* is stationary point, not local minimizer.

Inequality - constrained QP

- active - set method : small - medieum - scale probs.
- gradient - projection method : bounded optimization.
- Interior - point method : large convex QP.

Bound - constrained Optimization

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

$$\text{s.t.} \quad l \leq x \leq u . \quad \text{box}$$

- box constrained
  - $l_i = -\infty$, or $u_j = +\infty$
  - Do not require $G > 0$.
  - KKT: $Gx + c = \lambda_1 - \lambda_2$ $\quad l \leq x \leq u$

$$\lambda_1 \circ (x - l) = 0 \qquad \Longleftrightarrow \qquad (Gx+c)_i = \begin{cases} \geq 0, & x_i = l_i \\ 0, & x_i \in (l_i, u_i) \\ \leq 0, & x_i = u_i . \end{cases}$$

$$\lambda_2 \circ (x - u) = 0$$

$$l \leq x \leq u, \quad \lambda_1 \geq 0, \quad \lambda_2 \geq 0.$$

Projected operator:

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

$$\text{KKT} \Rightarrow x = p(x - (Gx + c), l, u)$$

$$l \leq x \leq u .$$

- Gradient Projection method

① first stage

② second stage

- Interior-point method

Chapter 13. Penalty methods

$$\min f(x) \quad \text{s.t. } x \in X$$

$\Rightarrow$ unconstrained optimization.

$$\min f(x) + \delta(x|X), \text{ where } \delta(x|X) = \begin{cases} 0, & x \in X \\ +\infty, & x \notin X \end{cases}$$

- Quadratic penalty method

$$\min f(x) \quad \text{s.t. } c_i(x) = 0, \ i \in \mathcal{E}.$$

$$p(x;\mu) = f(x) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} c_i^2(x) = f(x) + \frac{\mu}{2} \|c(x)\|_2^2$$

$\mu > 0$: penalty parameter, $\mu \to +\infty$.

- Alg: Quadratic Penalty Method – Equality case.

Barrier function

- $p(x;\mu) = f(x) + \mu^{-1} \sum_{i=1}^{m} \frac{1}{c_i(x)}$

$p(x;\mu) = f(x) - \mu^{-1} \sum_{i=1}^{m} \log c_i(x)$ $\Leftarrow$ $\min f(x)$ s.t. $c_i(x) \geq 0$.

Lemma: $\mu_2 > \mu_1 > 0 \Rightarrow f(x(\mu_2)) \leq f(x(\mu_1))$

$$-\sum_{i \in \mathcal{E}} \log(c_i(x(\mu_2))) \geq -\sum_{i \in \mathcal{E}} \log(c_i(x(\mu_1)))$$

Chapter 14. Sequential Quadratic Programming

Lagrange-Newton Method.

$$\min f(x)$$
$$\text{s.t. } c(x) = 0$$

$x_k \to x_{k+1}$?

- Construct a quadratic programming whose solution is used to define x_{k+1}.

- How to construct such QP with good properties?
- Newton's method to KKT conditions.

$A(x)^T = [\nabla c_1(x), \cdots, \nabla c_m(x)]$, A: full row rank.

KKT: $$F(x,\lambda) = \begin{bmatrix} \nabla f(x) - A(x)^T \lambda \\ c(x) \end{bmatrix} = 0$$

Newton's method

$$J_F = \begin{bmatrix} \nabla_{xx}^2 L(x,\lambda) & -A(x)^T \\ A(x) & 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} + \alpha_k \begin{bmatrix} p_k \\ p_\lambda \end{bmatrix}$$

$$\begin{bmatrix} \nabla_{xx}^2 L(x,\lambda) & -A(x)^T \\ A(x) & 0 \end{bmatrix} \begin{bmatrix} p_k \\ p_\lambda \end{bmatrix} = \begin{bmatrix} -\nabla f_k^T + A_k^T \lambda_k \\ -c_k \end{bmatrix}.$$

Lagrange-Newton Method!

$$\text{QP: } \min f_k + \nabla f_k^T p + \frac{1}{2} p^T \nabla_{xx}^2 L_k p \qquad (*).$$
$$\text{s.t. } c_k + A_k p = 0$$

(p_k, l_k) :
$$\begin{cases} \nabla_{xx}^2 L_k p_k + \nabla f_k - A_k^T l_k = 0 \\ A_k p_k + c_k = 0 \end{cases}$$

$$\Rightarrow \begin{bmatrix} \nabla_{xx}^2 L_k & -A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} p_k \\ l_k \end{bmatrix} = \begin{bmatrix} -\nabla f_k \\ -c_k \end{bmatrix}$$

$$\Rightarrow l_k = \lambda_k + p_\lambda.$$

$$\begin{cases} x_{k+1} = x_k + \alpha_k p_k \\ \lambda_{k+1} = \lambda_k + \alpha_k (l_k - \lambda_k) \end{cases}$$

Maratos effect.
Superlinear step $\frac{\|x_k + d_k - x_*\|}{\|x_k - x_*\|} \to 0$ couldn't be accepted.

Example. $\min\ 3v^2 - 2u$
s.t. $u = v^2$

KKT: $\begin{pmatrix} -2 \\ 6v \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ -2v \end{pmatrix} \Rightarrow \lambda^* = -2,\ v_* = 0,\ u_* = 0,\ x_* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$B = \nabla_{xx}^2 (x_*, \lambda_*) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

$x_k = \begin{pmatrix} \varepsilon^2 \\ \varepsilon \end{pmatrix}$, $\|x_k - x_*\| = \varepsilon$ $\quad \begin{cases} f_k = \varepsilon^2 \\ c_k = 0 \end{cases}$

$x_k + d_k = \begin{pmatrix} -\varepsilon^2 \\ 0 \end{pmatrix}$, $\|x_k + d_k - x_*\| = \varepsilon^2$. $\quad \begin{cases} f(x_k + d_k) = 2\varepsilon^2 \\ c(x_k + d_k) = -\varepsilon^2 \end{cases}$

$\Rightarrow p(x_k + d_k; \mu) > p(x_k; \mu)$ — l_1-penalty fun.

Chapter 15. Interior-point methods.

(works well for large probs, particularly when the number of free variables is large)

$$\min f(x)$$
$$\text{s.t. } c_E(x) = 0,\ y$$
$$c_I(x) \geq 0,\ z$$

KKT: $\nabla f(x) - A_E^T y - A_I^T z = 0$

$$F(x,s,y,z) = \begin{cases} z \cdot C_I(x) = 0 \\ C_E(x) = 0 \\ C_I(x) - s = 0 \\ s \geq 0,\ z \geq 0 \end{cases}$$

Barrier approach:

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } C_E(x) = 0 \\ &\quad\ C_I(x) - s = 0 \\ &\quad\ s \geq 0 \end{aligned}$$

$$\Rightarrow \begin{aligned} &\min\ f(x) - \mu \sum_{i=1}^{m} \log s_i \\ &\text{s.t. } C_E(x) = 0 \qquad y \\ &\quad\ C_I(x) - s = 0 \qquad z \end{aligned}$$

$\mu_k \to 0$, solve above problem w.r.t. fixed μ_k.

KKT:

$$\begin{pmatrix} \nabla f(x) \\ -\mu S^{-1} e \end{pmatrix} = \begin{pmatrix} A_E^T(x) y + A_I^T(x) z \\ -z \end{pmatrix}$$

$$C_E(x) = 0$$

$$C_I(x) - s = 0$$

$$\Leftrightarrow \begin{aligned} &\nabla f(x) - A_E^T(x) y - A_I^T(x) z = 0 \\ &-\mu S^{-1} e + z = 0 \\ &C_E(x) = 0 \\ &C_I(x) - s = 0. \end{aligned}$$

- Newton's method to perturbed KKT conditions.

6.2 最优化计算方法往年期末考试题

最优化历年期末考试题

7.3 试卷3

问题1. 计算下列问题的最优解

$$\min f(x) = u^2+4v^2-4u-8v$$
$$x=(u,v)^T\in\mathbb{R}^2$$

(a) 设初值为 $x_0=(0,0)^T$，应用FR共轭梯度法（结合精确线搜索）计算 x_1 和 x_2.

(b) 证明当 f 为二次函数时，从同一初始点出发，精确线搜索下的共轭梯度法采用 β^{FR}、β^{PRP}、β^{HS}、β^{DY} 产生相同的点列.

解：(a) $f(x)=\frac{1}{2}x^TAx-b^Tx$，$x=\begin{pmatrix}u\\v\end{pmatrix}$

$$A=\begin{pmatrix}2&0\\0&8\end{pmatrix},\quad b=\begin{pmatrix}4\\8\end{pmatrix},\quad \nabla f=\begin{pmatrix}\frac{\partial f}{\partial u}\\\frac{\partial f}{\partial v}\end{pmatrix}=\begin{pmatrix}2u-4\\8v-8\end{pmatrix}$$

① $f_0=(x_0)=0$，$\nabla f_0=\nabla f(x_0)=\begin{bmatrix}-4\\-8\end{bmatrix}$，$p_0=-\nabla f_0=\begin{bmatrix}4\\8\end{bmatrix}$

$$\alpha_0=\min_{\alpha} f(x_0+\alpha_0p_0)=\min_{\alpha}16\alpha^2+4\cdot8^2\alpha^2-4\cdot4\alpha-8\cdot8\alpha=\frac{5}{34}$$

$$x_1=x_0+\alpha_0p_0=\begin{bmatrix}0\\0\end{bmatrix}+\frac{5}{34}\begin{bmatrix}4\\8\end{bmatrix}=\begin{bmatrix}\frac{10}{17}\\\frac{20}{17}\end{bmatrix}$$

$$\nabla f_1=\nabla f(x_1)=\begin{bmatrix}-\frac{48}{17}\\\frac{24}{17}\end{bmatrix}$$

$$\beta_1^{FR}=\frac{\nabla f_1^T\nabla f_1}{\nabla f_0^T\nabla f_0}=\frac{(-\frac{48}{17})^2+(\frac{24}{17})^2}{16+64}=\frac{36}{289}$$

$$p_1=-\nabla f_1+\beta_1^{FR}p_0=-\begin{bmatrix}-\frac{48}{17}\\\frac{24}{17}\end{bmatrix}+\frac{36}{289}\begin{bmatrix}4\\8\end{bmatrix}=\begin{bmatrix}\frac{960}{289}\\-\frac{120}{289}\end{bmatrix}$$

② $$\alpha_1=\min_{\alpha}f(x_1+\alpha p_1)=\min_{\alpha}f\begin{pmatrix}\frac{10}{17}+\frac{960}{289}\alpha\\\frac{20}{17}+\frac{-120}{289}\alpha\end{pmatrix}=0.424.$$

$$x_2=x_1+\alpha_1p_1=\begin{bmatrix}\frac{10}{17}\\\frac{20}{17}\end{bmatrix}+0.424\begin{bmatrix}\frac{960}{289}\\\frac{-120}{289}\end{bmatrix}=\begin{bmatrix}2\\1\end{bmatrix}.$$

(b) 由线性 CG 方法的 $r_k^T p_i=0,\ i=0,\cdots,k-1$, $r_k^T r_i=0,\ i=0,\cdots,k-1$ 可知:

$$\nabla f_k^T p_i = 0 \quad (i=0,1,\cdots,k-1)$$

$$\nabla f_k^T \nabla f_i = 0 \quad (i=0,1,\cdots,k-1)$$

而

$$\beta_{k+1}^{FR} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k}$$

$$\beta^{PRP} = \frac{\nabla f_{k+1}^T(\nabla f_{k+1}-\nabla f_k)}{\nabla f_k^T \nabla f_k} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k} = \beta_{k+1}^{FR}$$

$$\beta_{k+1}^{HS} = \frac{\nabla f_{k+1}^T(\nabla f_{k+1}-\nabla f_k)}{p_k^T(\nabla f_{k+1}-\nabla f_k)} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{-p_k^T \nabla f_k}$$

$$= \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{-(-\nabla f_k^T + \beta_k^{HS} p_{k-1})\nabla f_k} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k} = \beta_{k+1}^{FR}$$

$$\beta_{k+1}^{DY} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}^T}{(\nabla f_{k+1}-\nabla f_k)^T p_k} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{-\nabla f_k^T p_k}$$

$$= \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{-\nabla f_k^T(-\nabla f_k + \beta_k^{HS} p_{k-1})} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k} = \beta_{k+1}^{FR}.$$

故 $\beta_{k+1}^{FR} = \beta_{k+1}^{PRP} = \beta_{k+1}^{HS} = \beta_{k+1}^{DY}$

而共轭梯度法算法框架相同, 仅 β_{k+1} 的计算不同, 故由以上证明可知 4 种方法产生相同的序列.

问题2. 简单讨论非线性共轭梯度法与有限内存BFGS方法的相同之处和差别.

解：相同之处：算法中仅需要计算梯度，不需要计算Hessian矩阵，也不需要进行求逆操作，同时均为迭代更新计算.

不同之处：非线性共轭梯度法具有二次终止性，能够在不超过n步内收敛，所需迭代次数较少.

有限内存BFGS的计算量很小，所需存储的数据较小.

问题3：考虑下列三个问题是否有解，请分别给予解释。

这里，$(u,v)^T \in \mathbb{R}^2$.

$$\min\ u+v,\quad s.t.\ u^2+v^2=2,\ 0\le u\le 1,\ 0\le v\le 1$$

$$\min\ u+v\quad s.t.\ u^2+v^2\le 1,\ u+v=3$$

$$\min\ uv\quad s.t.\ u+v=2.$$

解：① 由约束条件知：可行集合为 $W=\{(1,1)\}$，有解

② 由约束条件知：可行集合为 ϕ，故无解

③ 由约束条件知：可行集合为 $W=\{(u,v):u+v=2\}$，有解.

问题4：考虑问题：

$$\min_{x=(u,v)\in\mathbb{R}^2}\ -2u+v$$

$$s.t.\ (1-u)^3-v\ge 0$$

$$v+0.25u^2-1\ge 0.$$

最优解为 $x^*=(0,1)^T$.

(a) 验证LICQ条件在该点处是否成立？

(b) 验证KKT条件是否满足？

(c) 写出 $F(x^*)$ 及 $C(x^*,\lambda^*)$

(d) 分别验证二阶必要条件，二阶充分性条件是否成立？

解：$\nabla c_1 = \begin{bmatrix} \frac{\partial c_1}{\partial u} \\ \frac{\partial c_1}{\partial v} \end{bmatrix} = \begin{bmatrix} -3(1-u)^2 \\ -1 \end{bmatrix}$，$\nabla c_2 = \begin{bmatrix} \frac{\partial c_2}{\partial u} \\ \frac{\partial c_2}{\partial v} \end{bmatrix} = \begin{bmatrix} 0.5u \\ 1 \end{bmatrix}$

(a) $x^* = (0,1)^T$ 时，$\nabla c_1 = \begin{bmatrix} -3 \\ -1 \end{bmatrix}$，$\nabla c_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$。而 ∇c_1 和 ∇c_2 线性无关，故 LICQ 条件在该点处成立。

(b) $L = -2u + v - \lambda_1 [(1-u)^3 - v] - \lambda_2 (v + 0.25u^2 - 1)$

KKT 条件：$\nabla L = \begin{pmatrix} -2 \\ 1 \end{pmatrix} + \lambda_1 \begin{pmatrix} 3(1-u)^2 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} -0.5u \\ -1 \end{pmatrix} = 0$

$\lambda_1 [(1-u)^3 - v] = 0$，$\lambda_2 [v + 0.25u^2 - 1] = 0$.

$(1-u)^3 - v \geq 0$，$v + 0.25u^2 - 1 \geq 0$，$\lambda_1, \lambda_2 \geq 0$.

代入 $(0,1)^T$ 可得：$\lambda_1 = \frac{2}{3}$，$\lambda_2 = \frac{5}{3}$，$\lambda_1, \lambda_2 \geq 0$，故点 $x^* = (0,1)$ 满足 KKT 条件.

(c) $\mathcal{F}(x^*) = \{d : d^T \nabla c_i(x) \geq 0\}$

$= \{d : d^T \begin{pmatrix} -3 \\ -1 \end{pmatrix} \geq 0,\ d^T \begin{pmatrix} 0 \\ 1 \end{pmatrix} \geq 0\}$

$= \{d : 3d_1 + d_2 \leq 0,\ d_2 \geq 0\}$

$C(x^*, \lambda^*) = \{w : \nabla c_i(x^*)^T w = 0\}$

$= \{w : (-3,-1)w = 0,\ (0,1)w = 0\}$

$= \{w : w = 0\}$

(d) 二阶条件为：

二阶必要性条件：对于 (x^*, λ^*) 需满足 $d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \geq 0$，$\forall d \in \bar{G}$

其中 $\bar{G} = \left\{ d \;\middle|\; \begin{array}{l} d \in \mathbb{R}^n \\ \nabla c_i(x)^T d = 0, \quad i \in \mathcal{I} \text{ 且 } \lambda_i > 0 \text{ 或 } i \in \mathcal{E} \\ \nabla c_i(x)^T d \geq 0, \quad i \in \mathcal{I} \text{ 且 } \lambda_i = 0 \end{array} \right\}$

二阶充分性条件：对于符合一阶必要性条件的 (x^*, λ^*)，对任意的 $d \in G$，满足：$d^T \nabla_{xx}^2 L(x^*, \lambda^*) d > 0$.

其中：$G = \left\{ d \;\middle|\; \begin{array}{l} d \neq 0 \\ \nabla c_i(x)^T d = 0 \quad i \in \mathcal{I} \text{ 且 } \lambda_i > 0 \text{ 或 } i \in \mathcal{E} \\ \nabla c_i(x)^T d \geq 0 \quad i \in \mathcal{I} \text{ 且 } \lambda_i = 0 \end{array} \right\}$

② Lagrage 函数为: $L=-2u+v-\lambda_1[(1-u)\beta-v]-\lambda_2(v+0.25u^2-1)$

$\frac{\partial L}{\partial u}=-2+\lambda_1\beta-0.5\lambda_2 u$

$\frac{\partial^2 L}{\partial u^2}=-0.5\lambda_2$

$\frac{\partial L}{\partial v}=1+\lambda_1-\lambda_2$

$\frac{\partial^2 L}{\partial v\partial u}=0$

$\frac{\partial^2 L}{\partial v^2}=0$

$$L_{xx}^2(x^*,\lambda^*)=\begin{pmatrix}-0.5\lambda_2 & 0\\ 0 & 0\end{pmatrix}=\begin{pmatrix}-\frac{5}{6} & 0\\ 0 & 0\end{pmatrix}$$

因为 $\lambda_1=\frac{2}{3}$, $\lambda_2=\frac{5}{3}$; $\lambda_1,\lambda_2>0$. 故

$$\begin{cases}\nabla C_1^T d=0\\ \nabla C_2^T d=0\end{cases}\Rightarrow\begin{cases}(-3,-1)^T d=0\\ (0,1)^T d=0\end{cases}\Rightarrow d=\begin{pmatrix}d_1\\ d_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$$

故可得 $G=0$. 这表明在充分条件中对曲率的要求自然满足, 因此 $x^*=(0,1)$ 满足二阶必要条件和二阶充分条件, 是局部极小点.

问题5. 利用罚函数法 (可以任意挑选一种罚函数) 求解:

$$\min_{x=(u,v)^T\in\mathbb{R}^2} \quad 3u^2-2v$$

$$\text{s.t.} \quad u^2-v=0$$

请给出算法框架, 并挑选一初始点 x_0 出发计算 x_1, x_2.

解: 罚函数可设为 $L=Q(x,\mu)=3u^2-2v+\frac{\mu}{2}(u^2-v)^2$

那么 $\nabla L=\begin{pmatrix}6u\\ -2\end{pmatrix}+\mu\begin{pmatrix}(u^2-v)\cdot 2u\\ -(u^2-v)\end{pmatrix}=\begin{pmatrix}6u+2u\mu(u^2-v)\\ -2-\mu(u^2-v)\end{pmatrix}$

$$\nabla^2 L=\begin{bmatrix}6+2\mu(3u^2-v) & -2\mu u\\ -2\mu u & \mu\end{bmatrix}$$

取 $x_0=\begin{pmatrix}0\\ 1\end{pmatrix}$, $\mu_0=1$, $\mu_{k+1}=2\mu_k$, $\tau_k=\frac{1}{1+k}$.

那么：① $\nabla L_0 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$，$\nabla^2 L_0 = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$，采用牛顿法求步长：

$$d_0 = -(\nabla^2 L_0)^{-1} \nabla L_0 = -\frac{1}{4}\begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$x_1 = x_0 + d_0 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$，$\nabla Q(x,\mu) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$，即：$\|\nabla Q(x,\mu)\| \le \tau_0 = 1$.

故可更新

② $\mu_1 = 2\mu_0 = 2$

$\nabla L_1 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$，$\nabla L_1^2 = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$，

$$d_1 = -(\nabla^2 L_1)^{-1} \nabla L_1 = -\begin{pmatrix} -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

$x_2 = x_1 + d_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$，$\nabla Q(x,\mu) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. 即：$\|Q(x,\mu)\| \le \tau_1 = \frac{1}{2}$.

③ $\mu_2 = 2\mu_1$

$\nabla L_2 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$，$\nabla^2 L_2 = \begin{pmatrix} -2 & 0 \\ 0 & 4 \end{pmatrix}$，$d_2 = -(\nabla^2 L_2)^{-1} \nabla L_2 = \begin{pmatrix} 0 \\ -\frac{1}{2} \end{pmatrix}$

$x_3 = x_2 + d_2 = \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix}$，$\nabla Q(x,\mu) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

即：$\|Q(x,\mu)\| \le \tau_2 = \frac{1}{3}$.

故可更新.

问题6：利用SQP方法求解约束优化问题

$$\min_{x=(u,v)^T \in \mathbb{R}^2} \quad 2u - v$$
$$\text{s.t.} \quad v + u^2 + 2u = 0$$

设初值为 $x_0 = (0,0)^T$，请给出前两次迭代的计算过程

解：使用Lagrange-Newton方法：

$x_0 = (0,0)^T$，$\lambda_0 = 1$，$\beta = \frac{1}{2}$，

$$P(x,\lambda) = \|\nabla f(x) - \lambda A(x)\|_2^2 + \|c(x)\|_2^2$$

其中 $A(x) = \nabla c(x)$.

$$\nabla f(x) = \begin{pmatrix} \frac{\partial(2u-v)}{\partial u} \\ \frac{\partial(2u-v)}{\partial v} \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \quad \nabla^2 f(x) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$c(x) = v + u^2 + 2u$,

$$A(x) = \nabla c(x) = \begin{pmatrix} \frac{\partial(v+u^2+2u)}{\partial u} \\ \frac{\partial(v+u^2+2u)}{\partial v} \end{pmatrix} = \begin{pmatrix} 2u+2 \\ 1 \end{pmatrix}, \quad \nabla^2 c(x) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

$$P(x_0,\lambda_0) = \left\| \begin{pmatrix} 2 \\ -1 \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\|_2^2 + \|0+0^2+2\cdot 0\|_2^2 = 4$$

下面求解：

$$\begin{pmatrix} W(x_k,\lambda_k) & -A(x_k) \\ -A(x_k)^T & 0 \end{pmatrix} \begin{pmatrix} (\delta x)_k \\ (\delta\lambda)_k \end{pmatrix} = -\begin{pmatrix} \nabla f(x_k) - A(x_k)\lambda_k \\ -c(x_k) \end{pmatrix}$$

其中 $W(x,\lambda) = \nabla^2 f(x) - \sum_{i=1}^{m} (\lambda_k)_i \nabla^2 c_i(x_k)$.

又知：$W(x_0,\lambda_0) = \nabla^2 f(x_0) - \lambda_0 \nabla^2 c(x_0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix}$

则要求解

$$\begin{pmatrix} \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} & -\begin{pmatrix} 2 \\ 1 \end{pmatrix} \\ -(2\ 1) & 0 \end{pmatrix} \begin{pmatrix} (\delta x)_0 \\ (\delta\lambda)_0 \end{pmatrix} = -\begin{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \end{pmatrix}\cdot 1 \\ -0 \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} -2 & 0 & -2 \\ 0 & 0 & -1 \\ -2 & -1 & 0 \end{pmatrix} \begin{pmatrix} (\delta x)_0 \\ (\delta\lambda)_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$$

则：$\begin{cases} (\delta x)_0 = \begin{pmatrix} 2 \\ -4 \end{pmatrix} \\ (\delta\lambda)_0 = -2 \end{cases}$

$\alpha = \frac{1}{4}$, $\quad x_0 + \alpha(\delta x)_0 = \begin{pmatrix} \frac{1}{2} \\ -1 \end{pmatrix} \quad A\begin{pmatrix} \frac{1}{2} \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$

$\lambda_0 + \alpha(\delta\lambda)_0 = \frac{1}{2}$

$$P\left(\begin{pmatrix} \frac{1}{2} \\ -1 \end{pmatrix}, \frac{1}{2}\right) = \left\| \begin{pmatrix} 2 \\ 4 \end{pmatrix} - \frac{1}{2}\cdot\begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\|_2^2 + \left\| -1 + \left(\tfrac{1}{2}\right)^2 + 2\cdot\tfrac{1}{2} \right\|_2^2 = \frac{41}{16} < \left(1 - \frac{1}{4}\cdot\frac{1}{2}\right)\cdot 4 = \frac{7}{2}$$

$\nabla f(x_1) = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$, $c(x_1) = \frac{1}{4}$

故: $x_1 = \begin{pmatrix} \frac{1}{2} \\ -1 \end{pmatrix}$, $\lambda_1 = \frac{1}{2}$, $P(x_1, \lambda_1) = \frac{41}{16}$, $A(x_1) = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$,

继续求解:

$$\begin{pmatrix} W(x_1, \lambda_1) & -A(x_1) \\ -A(x_1)^T & 0 \end{pmatrix} \begin{pmatrix} (\delta x)_1 \\ (\delta \lambda)_1 \end{pmatrix} = -\begin{pmatrix} \nabla f(x_1) - A(x_1)\lambda_1 \\ -c(x_1) \end{pmatrix}$$

其中 $W(x_1, \lambda_1) = \nabla^2 f(x_1) - \lambda_1 \nabla^2 c(x_1) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$

即:
$$\begin{pmatrix} -1 & 0 & -3 \\ 0 & 0 & -1 \\ -3 & -1 & 0 \end{pmatrix} \begin{pmatrix} (\delta x)_1 \\ (\delta \lambda)_1 \end{pmatrix} = -\begin{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 3 \\ 1 \end{pmatrix} \\ -\frac{1}{4} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ \frac{3}{2} \\ \frac{1}{4} \end{pmatrix}$$

解得: $\begin{cases} (\delta x)_1 = \begin{pmatrix} 5 \\ -\frac{61}{4} \end{pmatrix} \\ (\delta \lambda)_1 = -\frac{3}{2} \end{cases}$

$x_2 = \begin{pmatrix} \frac{1}{2} \\ -1 \end{pmatrix} + \frac{1}{16}\begin{pmatrix} 5 \\ -\frac{61}{4} \end{pmatrix} = \begin{pmatrix} \frac{13}{16} \\ -\frac{64+61}{64} \end{pmatrix} = \begin{pmatrix} \frac{13}{16} \\ -\frac{125}{64} \end{pmatrix}$

$\lambda_2 = \frac{1}{2} - \frac{1}{16}\frac{3}{2} = \frac{13}{32}$

$$P\left(\begin{pmatrix} \frac{11}{2} \\ -\frac{65}{4} \end{pmatrix}, -1\right) = \left\| \begin{pmatrix} 2 \\ -1 \end{pmatrix} - \frac{1}{8}\begin{pmatrix} \frac{29}{8} \\ 1 \end{pmatrix} \right\|_2^2 + \left\| -\frac{125}{64} + \frac{13^2}{16^2} + \frac{13}{8} \right\|_2^2$$

$$= \left\| \begin{matrix} \frac{99}{64} \\ -\frac{9}{8} \end{matrix} \right\|_2^2 + \left\| \frac{-500+169+13\times 32}{256} \right\|_2^2$$

$$= \left(\frac{99}{64}\right)^2 + \left(\frac{9}{8}\right)^2 + \left(\frac{85}{256}\right)^2$$

$=$

有待继续检查

$= 5.9985 > \frac{117}{256}$. 不取. $\alpha = \frac{1}{4}$.

问题7. 考虑非线性共轭梯度法：$d_k = -g_k + \beta_k d_{k-1}$，其中：

$$\beta_k = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}} \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}}$$

计算 $d_k^T g_k$，分析 d_k 是否为下降方向.

解：

$$\begin{aligned}
d_k^T g_k &= (-g_k + \beta_k d_{k-1})^T g_k \\
&= -g_k^T g_k + \beta_k d_{k-1}^T g_k \\
&= -g_k^T g_k + \left(\frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}} \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} \right) d_{k-1}^T g_k \\
&= -g_k^T g_k + \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} d_{k-1}^T g_k \\
&\quad - \frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}} \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} d_{k-1}^T g_k \\
&= -g_k^T g_k + \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} d_{k-1}^T g_k - \frac{\|y_{k-1}\|^2 (g_k^T d_{k-1})^2}{(d_{k-1}^T y_{k-1})^2}
\end{aligned}$$

该步可有可无 $\left(= -\left\| \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} y_{k-1} - g_k \right\|_2^2 - \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} y_{k-1}^T g_k . \right)$

又由共轭梯度精确选取步长可知 $d_{k-1}^T g_k = 0$.

故可直接得到：$d_k^T g_k < 0$.

问题 8、如果 $\sigma > \sum_{i=1}^{m} |\lambda_i^*|$，$\lambda_i^*$ 是等式约束优化问题

$$\min f(x)$$
$$s.t.\ c_i(x)=0,\quad i\in\{1,\cdots,m\}$$

的 Lagrange 乘子，则 $f(x)+\sigma\max|c_i(x)|$ 为精确罚函数.

证明：往证 $f(x)+\sigma\|c(x)\|_\infty$ 在 $\sigma>\|\lambda^*\|_1$ 时为精确罚函数：

设 x^* 为等式约束问题的局部极小点.

由于 $\sigma > \|\lambda^*\|_1$，$\|c(x)\|=0$. 存在 $\delta>0$，使得当 $\|x-x^*\|<\delta$ 时，有：

$$\sigma_1\|c(x)\| < \sigma-\|\lambda^*\|_1$$

从而当 $\|x-x^*\|\le\delta$ 时，有：

$$\begin{aligned} P(x,\sigma) &= f(x)+\|\lambda^*\|_1\|c(x)\|_\infty+(\sigma-\|\lambda^*\|_1)\|c(x)\|_\infty \\ &\geq L(x^*,\lambda^*)+\sigma_1\|c(x)\|_\infty\|c(x)\|_\infty \\ &\geq L(x^*,\lambda^*)+\tfrac{1}{2}\sigma_1\|c(x)\|_2^2 \\ &> L(x^*,\lambda^*) = P(x^*,\sigma) \end{aligned}$$

（注：$L(x^*,\lambda^*) = f(x^*)-(\lambda^*)^T c(x^*)$）

则 x^* 是 $P(x,\sigma)=f(x)+\sigma\|c(x)\|_\infty=f(x)+\sigma\max|c_i(x)|$ 的局部（严格）极小点.

则罚函数 $f(x)+\sigma\max|c_i(x)|$ 的极小点和原问题的解吻合，

故证得 $f(x)+\sigma\max|c_i(x)|$ 为精确罚函数.

6.3 最优化计算方法本学期作业题：第一次

Question 1. Compute the gradient and Hessian of the function $q(x) = \frac{1}{2}x^\top Ax + b^\top x$, where A is symmetric.

Solution.

Let $q_1(x) = b^\top$, $q_2(x) = \frac{1}{2}x^\top Ax$, and hence $q(x) := q_1(x) + q_2(x)$. We have that for $q_1(x)$, $q_1(x) = b^\top x = \sum_{i=1}^{n} b_i x_i$,

$$\nabla q_1(x) = \begin{bmatrix} \frac{\partial q_1}{\partial x_1} \\ \cdots \\ \frac{\partial q_1}{\partial x_1} \end{bmatrix} = \begin{bmatrix} b_1 \\ \cdots \\ b_n \end{bmatrix} = b,$$

$$\nabla^2 q_1(x) = \begin{bmatrix} \frac{\partial^2 q_1}{\partial x_1^2} & \frac{\partial^2 q_1}{\partial x_2 \partial x_1} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} = \left[\frac{\partial^2 \sum_i b_i x_i}{\partial x_2 \partial x_2} \right]_{s,t=1\cdots n} = 0.$$

For $q_2(x) = x^\top Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j$, since A is symmetric we have

$$\begin{aligned}
\nabla q_2(x) &= \left[\frac{\partial q_2}{\partial x_s}\right]_{s=1\cdots n} = \frac{1}{2}\left[\sum_j A_{sj}x_j + \sum_i A_{is}x_i\right]_{s=1\cdots n} \\
&= \left[\sum_{j=1}^{n} A_{sj}x_j\right]_{s=1\cdots n} \\
&= Ax. \\
\nabla^2 q_2(x) &= \left[\frac{\partial^2 q_2}{\partial x_s \partial x_t}\right] = \frac{1}{2}\left[\frac{\partial^2 \sum_i \sum_j A_{ij}x_i x_j}{\partial x_s \partial x_t}\right] = \frac{1}{2}\left[A_{st} + A_{ts}\right] = A. \\
\nabla q(x) &= \nabla q_1(x) + \nabla q_2(x) = b + Ax \\
\nabla^2 q(x) &= \nabla^2 q_1(x) + \nabla^2 q_2(x) = 0 + A = A.
\end{aligned}$$

□

Question 2. Compute the gradient and Hessian of

$$f(x) = 100\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2.$$

Show that $x^* = (1,1)^\top$ is the only local minimizer of this function, and that the Hessian at this point is positive definite. Write a program on trust region method with subproblems solved by the Dogleg method. Apply it to minimize this function. Choose $B_k = \nabla^2 f(x_k)$. Experiment with the update rule of trust region. Give the first two iterates.

Solution.

From our target function, we can get the following by calculation that for $x = (x_1, x_2)^\top$

$$\begin{aligned}
\frac{\partial f}{\partial x_1} &= 100 \cdot 2\left(x_2 - x_1^2\right)(-2x_1) + 2(1 - x_1)(-1) \\
&= -400x_1\left(x_2 - x_1^2\right) - 2(1 - x_1). \\
\frac{\partial f}{\partial x_2} &= 200\left(x_2 - x_1^2\right).
\end{aligned}$$

Thus the gradient of f at x is

$$\begin{aligned}
\nabla f(x) &= \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right)^\top \\
&= \left(100 \cdot 2\left(x_2 - x_1^2\right)(-2x_1) + 2(1 - x_1)(-1), 200\left(x_2 - x_1^2\right)\right)^\top.
\end{aligned}$$

$$\frac{\partial^2 f}{\partial x_1^2} = -400\left[x_1(-2x_1) + \left(x_2 - x_1^2\right)(1)\right] + 2 = -400\left(x_2 - 3x_1^2\right) + 2.$$
$$\frac{\partial^2 f}{\partial x_2 \partial x_1} = \frac{\partial^2 f}{\partial x_1 \partial x_2} = -400x_1 \quad \text{and} \quad \frac{\partial^2 f}{\partial x_2^2} = 200.$$

Thus the Hessian of f at x is

$$\nabla^2 f(x) = \begin{bmatrix} -400\left(x_2 - 3x_1^2\right) + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}.$$

Therefore we know that

1. $x^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is the only solution to $\nabla f(x) = 0$.

2. $\nabla^2 f(x^*) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}$ is positive definite since $802 > 0$, and $\det \nabla^2 f(x^*) = 802 \cdot 200 - 400 \cdot 400 = 400 > 0$.

3. $\nabla f(x)$ is continuous.

Coding: See the detail of codes in the Appendix.

In my codes, the initial condition is set as the following:

$$x_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \triangle_0 = 0.1,\ \eta = 0.1,\ \hat{\Delta} = 1,\ \|\nabla f(x)\|_2 \leqslant \varepsilon = 1.0^{-7}.$$

The iteration stops after 15 steps, terminating at the optimal point, which is x^* and the optimal value is $f^* = 0$. Figure 1 shows the function value $f(x_k)$ versus iteration number k: The first two iteration points are

$$x_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \to x_1 = \begin{bmatrix} 0.1 \\ 0 \end{bmatrix} \to x_2 = \begin{bmatrix} 0.2933 \\ 0.0513 \end{bmatrix}.$$

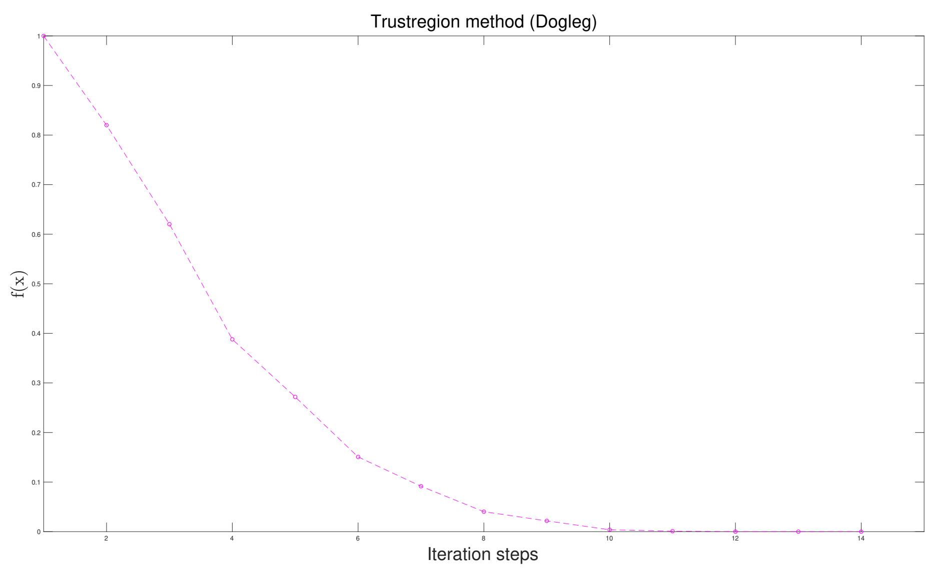

Fig.1 Trust region method with Dogleg stategy

□

Question 3. Apply Steepest Descent method with exact line search to the problem:

$$\min f(x) = 5x_1^2 + \frac{1}{2}x_2^2.$$

Carry out two iterations, starting from $x^0 = (0.1, 1)^\top$. Think about how $\{x^k\}$ will behave.

Solution.

The gradient of f at x is $\nabla f(x) = \begin{bmatrix} 10x_1 \\ x_2 \end{bmatrix}$, and the Hessian of f at x is $\nabla^2 f(x) = \begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix}$
The searching direction of Steepest Descent method is $p^k = -\nabla f(x^k)$, and the step length with exact line search for the steepest descent method is as $\alpha^k = -\frac{{p^k}^\top p^k}{{p^k}^\top A p^k}$, where $A = \nabla f(x)$.
Our starting point is $x^0 = (0.1, 1)$. By computation, we know that

$$p^0 = -\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \alpha^0 = \frac{2}{11}, \quad x^1 = x^0 + \alpha^0 p^0 = \frac{9}{11}\left(-\frac{1}{10}, 1\right)^\top$$

and we can get that

k	x^k	p^k	α^k
0	$\left(\frac{1}{10}, 1\right)^\top$	$(-1,-1)^\top$	$\frac{2}{11}$
1	$\frac{9}{11}\left(-\frac{1}{10}, 1\right)^\top$	$\frac{9}{11}(1,-1)^\top$	$\frac{2}{11}$
2	$\frac{81}{121}\left(\frac{1}{10}, 1\right)^\top$	$\frac{81}{121}(-1,-1)^\top$	$\frac{2}{11}$

.

Thus x^k has the form as

$$x_k = \left(\frac{9}{11}\right)^k \left(\frac{(-1)^k}{10}, 1\right)^\top$$

Coding: See the detail of codes in the Appendix.

The codes show that

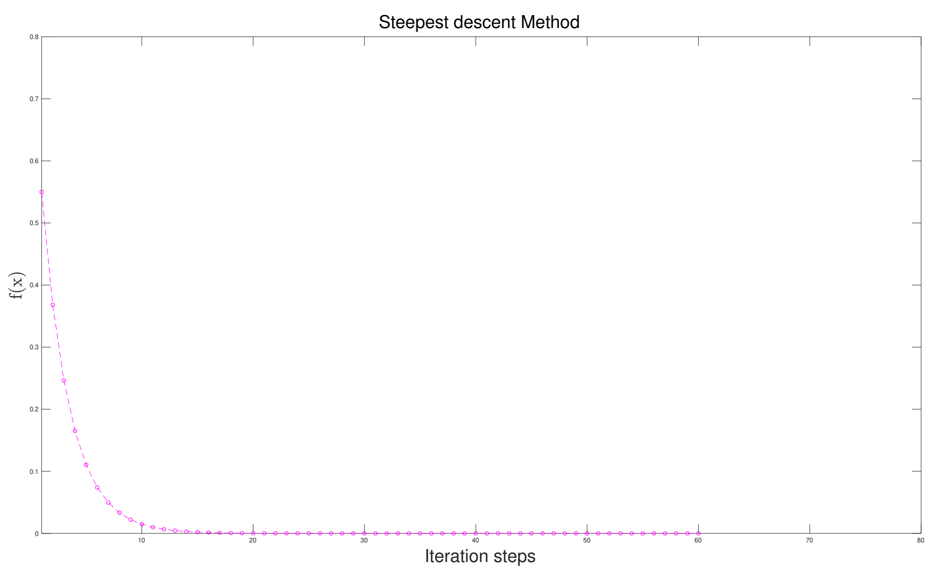

Fig.2 Steepest descent method with Exact line search

□

Question 4. Apply CG method with exact line search to solve

$$\min \frac{1}{2}x^\top Ax + b^\top x,$$

starting from $x_0 = (2,1)^T$. Here $A = [4,1;1,3]$ and $b = -(1,2)$.

Solution.

We choose initial searching direction is chosen to be $-\nabla f(x_0)$ at x_0, that is

$$x_0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad r_0 = Ax_0 + b = \begin{bmatrix} 8 \\ 3 \end{bmatrix}, \quad p_0 = -r_0,$$

Then we have that

$$\alpha_0 = \frac{r_0^\top r_0}{p_0^\top A p_0} = 0.2205, \quad x_1 = x_0 + \alpha_0 p_0 = \begin{bmatrix} 0.2356 \\ 0.3384 \end{bmatrix}$$

$$r_1 = r_0 + \alpha_0 A p_0 = \begin{bmatrix} 0.2810 \\ -0.7492 \end{bmatrix}, \quad \beta_0 = \frac{r_1^\top r_1}{r_0^\top r_0} = 0.0088$$

$$p_1 = -r_0 + \beta_0 p_0 = \begin{bmatrix} -0.3511 \\ 0.7229 \end{bmatrix}, \quad \alpha_1 = 0.4122,$$

$$x_2 = x_1 + \alpha_1 p_1 = \begin{bmatrix} 0.0909 \\ 0.6364 \end{bmatrix}, \quad r_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Then it terminates.

Coding: See the detail of codes in the Appendix.

The codes show the same thing as well.

□

6.4 最优化计算方法本学期作业题：第二次

Question 1. Apply the Newton's method with step size 1 to the following problem:

$$\min_{x=(t_1,t_2)^T\in\mathbb{R}^2} \quad t_1^2+t_2^2+t_1^4$$

with starting point $x_0=(\varepsilon,\varepsilon)^T$ where $\varepsilon>0$ is very small, calculate the next iterate and you should find that $\|x_1\|_2=O(\varepsilon^3)$.

Solution.

Notice that the gradient is $\nabla f(x)=(2t_1+4t_1^3,2t_2)^T$, and the Hessian is $\nabla^2 f(x)=[2+12t_1^2,0;0,2]$, at the starting point $x_0=(\varepsilon,\varepsilon)^T$, we can get that

$$\nabla f(x_0)=\left(2\varepsilon+4\varepsilon^3,2\varepsilon\right)^T,\quad \nabla^2 f(x_0)=\begin{bmatrix}2+12\varepsilon^2 & 0\\ 0 & 2\end{bmatrix}$$

Applying Newton's method with step size 1, we can get $x_1=x_0-\left(\nabla^2 f(x_0)\right)^{-1}g(x_0)=\left(\dfrac{4\varepsilon^3}{1+6\varepsilon^2},0\right)^T$.

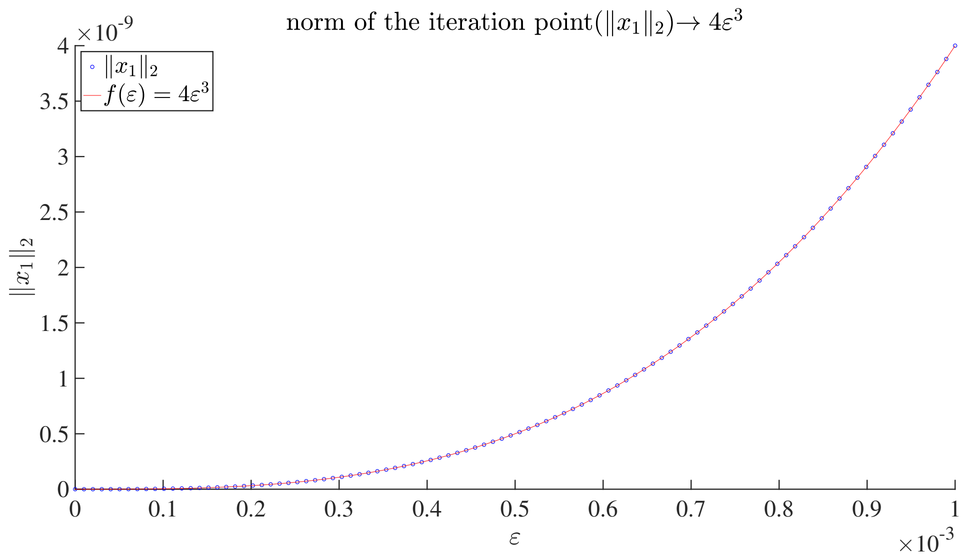

Fig.1 norm of x_1

Since $\varepsilon>0$ is very small, we have $\dfrac{\|x_1\|_2}{\varepsilon^3}\to 4$, as $\varepsilon\to 0$. Therefore, $\|x_1\|_2=O(\varepsilon^3)$. Besides, writing a program on Matlab shows us the same conclusion that $\|x_1\|_2\to 4\varepsilon^3$. See Fig.1 for the detail.

□

Question 2. Write a program on quasi-Newton method with exact line search to solve the problem:

$$\min_{x=(t_1,t_2)^T\in\mathbb{R}^2} \quad t_1^2+2t_2^2$$

starting form $x_0=(1,1)^T$. Use BFGS and DFP update formula, respectively. Set the initial $H_0=I$ for both methods.

Solution.

We show the program result by Fig.2 by writing programs on Matlab (See the detail of the codes in the Appendix).

Fig.2 shows the function value $f(x_k)$ versus iteration number k with DFP and BFGS method respectively (with different color lines).

It shows that both methods terminate the iterate at the optimal point $x=(0,0)^T$ in only two iterations, which accord with the quadratic termination property. The iterations are shown as the following:

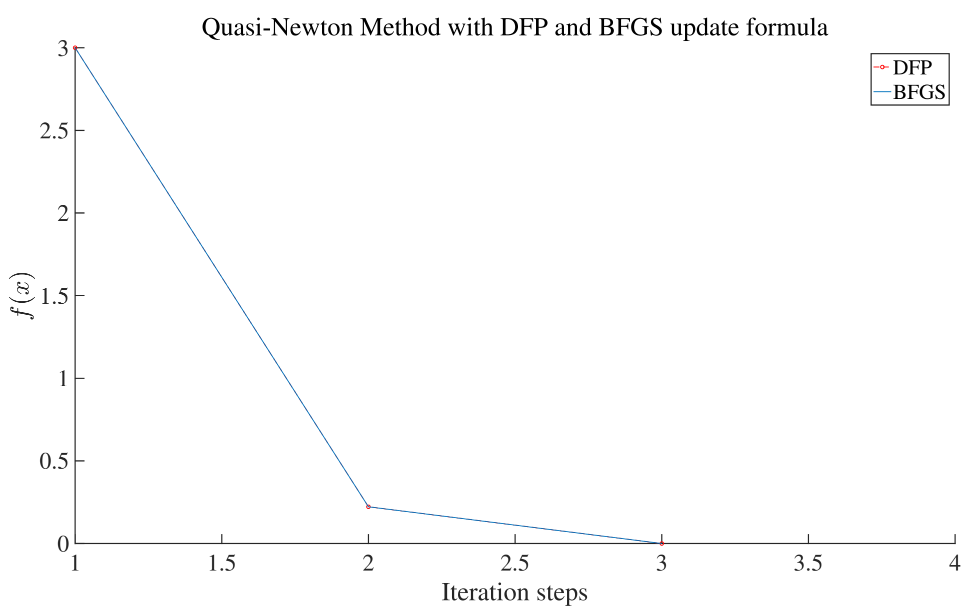

Fig.2 Quasi-Newton Method with DFP and BFGS update formula

The iteration points of both methods are the same:

$$x_0=\begin{bmatrix}1\\1\end{bmatrix},\ x_1=\begin{bmatrix}0.4444444444444444\\-0.1111111111111111\end{bmatrix},\ x_2=\begin{bmatrix}0\\0\end{bmatrix}$$

$\square$

Question 3. Prove that if $J_k^T r_k \neq 0$, the Cauchy point of the trust region subproblem for nonlinear least-squares problem

$$\min q_k(d) = \frac{1}{2}\left\|J_k d + r_k\right\|_2^2 \quad \text{ s.t. } \|d\|_2 \leqslant \Delta_k$$

satisfies

$$q_k(0) - q_k\left(s_k^c\right) \geqslant \frac{1}{2}\left\|J_k^T r_k\right\| \min\left\{\frac{\left\|J_k^T r_k\right\|}{\left\|J_k^T J_k\right\|}, \Delta_k\right\}.$$

Proof.

The model function is

$$\begin{aligned} q_k(d) &= \frac{1}{2}\left\|J_k d + r_k\right\|_2^2 \\ &= \frac{1}{2} d^T J_k^T J_k d + \left(J_k^T r_k\right)^T d + \frac{1}{2} r_k^T r_k. \end{aligned}$$

According to the definition of the Cauchy point, we know that

$$s_k^c = \arg\min q_k(d) \quad \text{ s.t. } d = -\tau J_k^T r_k, \|d\| \leqslant \Delta_k, \tau \geqslant 0,$$

the parameter τ_k satisfies

$$\tau_k = \arg\min \phi(\tau) = \arg\min -\tau\left\|J_k^T r_k\right\|^2 + \frac{\tau^2}{2}\left(J_k^T r_k\right)^T J_k^T J_k\left(J_k^T r_k\right), \quad \text{ s.t. } 0 \leq \tau \leq \frac{\Delta_k}{\left\|J_k^T r_k\right\|}$$

We get that if $J_k J_k^T r_k = 0$, then

$$\tau_k = \Delta_k / \left\|J_k^T r_k\right\|, \text{ then } q_k(0) - q_k\left(s_k^c\right) = \Delta_k\left\|J_k^T r_k\right\|,$$

otherwise, we have

$$\tau_k = \min\left\{\frac{\left\|J_k^T r_k\right\|^2}{\left(J_k^T r_k\right)^T J_k^T J_k\left(J_k^T r_k\right)}, \frac{\Delta_k}{\left\|J_k^T r_k\right\|}\right\}.$$

Case (i): $\tau_k = \dfrac{\left\|J_k^T r_k\right\|^2}{\left(J_k^T r_k\right)^T J_k^T J_k\left(J_k^T r_k\right)}$: In this case, we have

$$q_k(0) - q_k\left(s_k^c\right) = \frac{\left\|J_k^T r_k\right\|^4}{2\left(J_k^T r_k\right)^T J_k^T J_k\left(J_k^T r_k\right)} \geqslant \frac{\left\|J_k^T r_k\right\|^2}{2\left\|J_k^T J_k\right\|}.$$

Case (ii): $\tau_k = \dfrac{\Delta_k}{\|J_k^T r_k\|}$: In this case, we have

$$\begin{aligned} q_k(0) - q_k\left(s_k^c\right) = & \Delta_k \left\|J_k^T r_k\right\| - \frac{\Delta_k^2 \left(J_k^T r_k\right)^T J_k^T J_k \left(J_k^T r_k\right)}{2\left\|J_k^T r_k\right\|^2} \\ \geqslant & \Delta_k \left\|J_k^T r_k\right\| - \frac{\Delta_k \left\|J_k^T r_k\right\|}{2} \\ \geqslant & \frac{\Delta_k \left\|J_k^T r_k\right\|}{2}. \end{aligned}$$

Therefore, we proved that

$$q_k(0) - q_k\left(s_k^c\right) \geqslant \frac{1}{2}\left\|J_k^T r_k\right\| \min\left\{\frac{\left\|J_k^T r_k\right\|}{\left\|J_k^T J_k\right\|}, \Delta_k\right\}.$$

□

Question 4. Find the KKT point(s) of the following problem:

$$\begin{aligned} \min_{x=(t_1,t_2)^T\in\mathbb{R}^2} \quad & t_1 + t_2 \\ \text{s. t.} \quad & 2 - 2t_1^2 - t_2^2 \geqslant 0, \quad t_2 \geqslant 0 \end{aligned}$$

Solution.

We know that the KKT conditions are

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \lambda_1 \begin{bmatrix} -4t_1 \\ -2t_2 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\begin{aligned} \lambda_1\left(2 - 2t_1^2 - t_2^2\right) = 0, & \quad \lambda_2 t_2 = 0 \\ 2 - 2t_1^2 - t_2^2 \geqslant 0, & \quad t_2 \geqslant 0 \\ \lambda_1 \geqslant 0, & \quad \lambda_2 \geqslant 0. \end{aligned}$$

We have the following two cases since $\lambda_2 t_2 = 0$:

Case (i): $t_2 = 0$. Since $1 = -2\lambda_1 t_2 + \lambda_2, \lambda_2 = 1$. Then we can get that $\lambda_1\left(2 - 2t_1^2 - t_2^2\right) = 0, 1 = -4\lambda_1 t_1, \lambda_1 \geqslant 0$, which yields $t_1 = -1, \lambda_1 = \dfrac{1}{4}$. Therefore, we know that $(-1, 0)^T$ is a KKT point.

Case (ii): $\lambda_2 = 0$. Then $1 = -2\lambda_1 t_2$ which contradicts with $t_2 \geqslant 0, \lambda_1 \geqslant 0$.

Therefore, there is only one KKT point, which is $(-1, 0)^T$.

□

Question 5. Write a program to apply a penalty function method to solve

$$\min_{x=(t_1,t_2)^T\in\mathbb{R}^2} \quad t_1+t_2$$
$$\text{s.t.} \quad t_1^2+t_2^2-2=0$$

Calculate the iterates generated in the first two iterations, namely, x_1 and x_2.

Solution.

In the coding, we set the penalty parameter $\mu_0 = 1$ for Courant Penalty Method and Augmented Lagrange Method, and the Lagrange multiplier $\lambda_0 = 0.5$ for Augmented Lagrange Method. See Fig.3 for the detail and see codes in the Appendix.

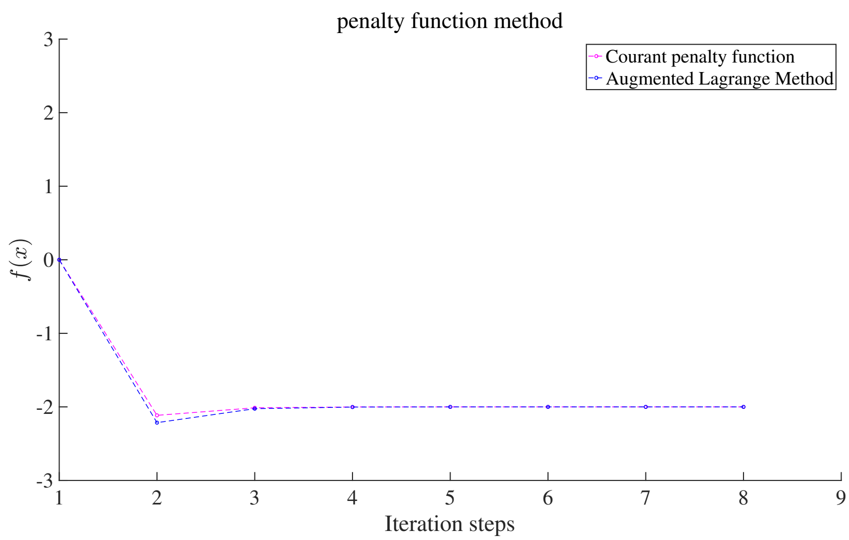

Fig.3 Penalty function method

The first two iteration points of Courant Penalty Method:

$$x_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ x_1 = \begin{bmatrix} -1.05748178442269 \\ -1.05743647909116 \end{bmatrix}, \ x_2 = \begin{bmatrix} -1.00616043865704 \\ -1.00622042415767 \end{bmatrix}.$$

The first two iteration points of Augmented Lagrange Method:

$$x_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ x_1 = \begin{bmatrix} -1.107168992903381 \\ -1.107128887189611 \end{bmatrix}, \ x_2 = \begin{bmatrix} -1.012291549452587 \\ -1.012253996286513 \end{bmatrix}.$$

□

6.5 最优化计算方法 2020 期末考试理论题

Consider the following problem

$$\begin{aligned}\min_{x\in\mathbb{R}^n} \quad & c^T x \\ \text{s.t.} \quad & Ax = Az \\ & \left\|D^{-1}(x-z)\right\|_2 \le \beta\end{aligned}$$

where $A \in R^{m\times n}$ has full row rank, $z \in \mathbb{R}^n$ is a given vector with all entries being positive, D is a diagonal matrix with positive diagonal elements $D_{ii} := z_i, i = 1\cdots,n$, and $\beta \in (0,1)$.

(i) Give the KKT optimality conditions for this problem.

(ii) From the optimality conditions express the optimal solution x^* as $x^* = z + p$; i.e., what is p?

Solution.

(i) The problem's constraint $\left\|D^{-1}(x-z)\right\|_2 \le \beta$ is equivalent to $(x-z)^T D^{-2}(x-z) \le \beta^2$. The Lagrangian function is $L(x,\lambda) = c^T x - \lambda_1^T A(x-z) - \lambda_2\left(\beta^2 - (x-z)^T D^{-2}(x-z)\right)$, where $\lambda_1 \in \mathbb{R}^m, \lambda_2 \in \mathbb{R}$. Suppose that x^* is a local solution, $\lambda^* = (\lambda_1^*, \lambda_2^*)^T$ is a Lagrangian multiplier vector. The KKT optimality conditions for this problem is

$$\begin{aligned}\nabla_x L\left(x^*,\lambda^*\right) &= 0 \\ A\left(x^*-z\right) &= 0 \\ \beta - \left\|D^{-1}(x^*-z)\right\|_2 &\ge 0 \\ \lambda_2^* &\ge 0 \\ \lambda_2^*\left(\beta - \left\|D^{-1}\left(x^*-z\right)\right\|_2\right) &= 0,\end{aligned}$$

which means

$$\begin{aligned}c - A^T\lambda_1^* + 2\lambda_2^* D^{-2}(x^*-z) =&0 \\ Ax^* =& Az \\ \left\|D^{-1}(x^*-z)\right\|_2 \le& \beta \\ \lambda_2^* \ge& 0\end{aligned}$$

$$\left(\beta-\left\|D^{-1}(x^*-z)\right\|_2\right)\lambda_2^*=0.$$

The KKT optmiality conditions for this problem is above.

(ii) From (i), we get

$$\begin{aligned}
c-A^T\lambda_1^*+2\lambda_2^*D^{-2}p=&0\\
Ap=&0\\
\left\|D^{-1}p\right\|_2\leq&\beta\\
\lambda_2^*\geq&0\\
\left(\beta-\left\|D^{-1}p\right\|_2\right)\lambda_2^*=&0.
\end{aligned}$$

Multiplying AD^2 on both sides of the first condition in KKT optmiality conditions leads to the equation

$$AD^2c-AD^2A^T\lambda_1^*=0.$$

Then $\lambda_1^*=\left(AD^2A^T\right)^{-1}AD^2c$, and we know that the first condition of KKT optmiality conditions becomes to

$$2\lambda_2^*p=-D^2\left(c-A^T\lambda_1^*\right)=-D^2\left(I-A^T\left(AD^2A^T\right)^{-1}AD^2\right)c.$$

There are two cases according to $(I-A^T(AD^2A^T)^{-1}AD^2)c=0$ or not.

CASE 1: $(I-A^T(AD^2A^T)^{-1}AD^2)c=0$, which means $\lambda_2^*p=0$. Besides, we know that $\lambda_2^*=0$. Otherwise, if $\lambda_2^*>0$, we have $\|D^{-1}p\|_2=\beta$ according to theorem of complementary slackness. Then we get that $\lambda_2^*=0$, which is contradict. Thus we get that $\lambda_2^*=0$, and then $\|D^{-1}p\|<\beta$, which only needs $Ap=0$. Thus $Ap=0$ in CASE 1. For all p that satisfies $Ap=0$, and $\|D^{-1}p\|_2\leq\beta$, $x^*=z+p$ is the optimal solution.

CASE 2: $\left(I-A^T\left(AD^2A^T\right)^{-1}AD^2\right)c\neq0$, which means $\lambda_2^*p\neq0$, and $\lambda_2^*\neq0$. Then $\|D^{-1}p\|_2=\beta$.

Recall that $D^{-1}p=\frac{1}{2\lambda_2^*}D\left(I-A^T\left(AD^2A^T\right)^{-1}AD^2\right)c$, thus

$$D^{-1}p=-\beta\frac{D\left(I-A^T\left(AD^2A^T\right)^{-1}AD^2\right)c}{\left\|D\left(I-A^T\left(AD^2A^T\right)^{-1}AD^2\right)c\right\|_2},$$

then p is

$$p = -\beta \frac{D^2 \left(I - A^T \left(AD^2 A^T \right)^{-1} AD^2 \right) c}{\left\| D \left(I - A^T \left(AD^2 A^T \right)^{-1} AD^2 \right) c \right\|_2}$$

in CASE 2.

All in all, we get the solution of p and the optimal solution $x^* = z + p$. □