PS 1 Problem 4

Thursday, April 4, 2019 5:23 PM

$$\min_{x \in \mathbb{R}^n} \quad f(x) := \frac{1}{2} x^T Q x - b^T x$$

$Q \in \mathbb{R}^{n \times n}$ pos def $\lambda_1, \dots, \lambda_n > 0$ (symm)

$b \in \mathbb{R}^n$

a) candidate x^* for local min^m

$\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ pos def

$\Rightarrow Qx^* - b = 0$ $\Rightarrow Q$ pos def

$\Rightarrow \boxed{x^* = Q^{-1} b}$ ✓ given!

$\therefore$ only candidate is $x^* = Q^{-1} b$

$\because Q$ is pos def everywhere

$\Rightarrow \nabla^2 f(x)$ " " " "

$\Rightarrow f$ is strictly convex

$\therefore$ $\boxed{x^* = Q^{-1} b \text{ is both local \& global min}^m}$

b) After $x^{(0)} \in \mathbb{R}^n$, we pick

$$x^{(1)} = x^{(0)} - \alpha^k \left(\nabla^2 f(x^{(0)})\right)^{-1} \nabla f(x^{(0)})$$

If we pick step size $\alpha = 1 = \eta_0$

$$x^{(1)} = x^{(0)} - 1 \; Q^{-1} (Qx^{(0)} - b)$$

$$\Rightarrow x^{(1)} = x^{(0)} - x^{(0)} + Q^{-1} b$$

$$\Rightarrow \boxed{x^{(1)} = Q^{-1} b = x^*}$$

So, we converge in 1 iteration to x^*

If n is large and Q has no particular structure, then computing inverse of $Q \in \mathbb{R}^{n \times n}$ is very computationally expensive, making this method intractable.

c) $S \in \mathbb{R}^{n\times n}$ is symm

$\Rightarrow S = U\Sigma U^T$ $\quad U \in \mathbb{R}^{n\times n}$ orthogonal

$\Sigma = \text{diag}(\mu_1, \dots \mu_n)$

$x \in \mathbb{R}^n$

$$\|Sx\|_2 = \|U\Sigma U^T x\|_2$$

Let $\Sigma U^T x = y$ $\quad \Rightarrow y \in \mathbb{R}^{n\times n}\, \mathbb{R}^{n\times n}\, \mathbb{R}^{n\times 1}$

$\Rightarrow y \in \mathbb{R}^n$

$\therefore \|Uy\|_2 = \|U^T y\|_2 = \|y\|_2$

$\because U$ is orthogonal

$$\therefore \boxed{\|Sx\|_2 = \|U \underbrace{\Sigma U^T x}_{y}\|_2 = \|\Sigma U^T x\|_2}$$

For any $z \in \mathbb{R}^n$

$$\Sigma z = \begin{bmatrix} \mu_1 & & 0 \\ & \mu_2 \ddots & \\ 0 & & \mu_n \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} \mu_1 z_1 \\ \mu_2 z_2 \\ \vdots \\ \mu_n z_n \end{bmatrix}$$

Let $|\mu^*| = \max_{i=1:n} |\mu_i|$

$$\Rightarrow \|\Sigma z\| = \left\| \begin{bmatrix} \mu_1 z_1 \\ \vdots \\ \mu_n z_n \end{bmatrix} \right\| = \left(\sum_{i=1:n} \mu_i^2 z_i^2 \right)^{1/2}$$

Each $\mu_i^2 z_i^2 \le \mu^{*2} z_i^2$ $\quad \because |\mu^*| \geq |\mu_i|$ by defn

$$\Rightarrow \|\Sigma z\|_2 \le \left(\mu^{*2} \sum_{i=1:n} z_i^2\right)^{1/2} = \mu^* \|z\|_2$$

$$\Rightarrow \boxed{\|\Sigma z\|_2 \le \left(\max_{i=1:n} |\mu_i|\right) \|z\|_2}$$

We showed $\|Sx\|_2 = \|\Sigma \underbrace{U^T x}_{z}\|_2$

Let $U^T x = z$

$$\|\Sigma z\|_2 \le \left(\max_{i=1:n} |\mu_i|\right) \|z\|_2$$

$$\Rightarrow \|Sx\|_2 \le \left(\max_{i=1:n} |\mu_i|\right) \|U^T x\|_2$$

$\because U$ orthogonal $\|U^T x\|_2 = \|x\|_2$

$$\Rightarrow \boxed{\|Sx\|_2 \leq \left(\max_{i=1:n} |\mu_i|\right) \|x\|_2}$$

d) $\eta > 0$ eig of $Q \to \lambda_1 \dots \lambda_n$ $\quad \|Q - \lambda I\| = 0$

$$\Rightarrow Q v_i = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} v_i$$

eig of $I - \eta Q$. Choose $v_i =$ eigvectors of Q.

$$(I - \eta Q) v_i = I v_i - \eta Q v_i$$

$$= v_i - \eta \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} v_i$$

$$= \begin{bmatrix} 1-\eta\lambda_1 & & 0 \\ & \ddots & \\ 0 & & 1-\eta\lambda_n \end{bmatrix} v_i$$

$\therefore$ $\boxed{(1-\eta\lambda_1), (1-\eta\lambda_2) \text{ -----} (1-\eta\lambda_n) \text{ are eigenvalues of } (I - \eta Q).}$

e) $x^{(k+1)} = x^{(k)} - \eta \nabla f(x^{(k)})$

$\delta_{k+1} := \|x^{(k+1)} - x^*\|_2 \qquad \delta_k := \|x^{(k)} - x^*\|_2$

$x^* = Q^{-1} b \quad \Rightarrow b = Q x^*$

$\delta_{k+1} = \|x^{(k)} - \eta \nabla f(x^{(k)}) - x^*\|_2$

$\cdot\ f(x) = \frac{1}{2} x^T Q x - b^T x$

$\nabla f(x^{(k)}) = Q x^{(k)} - b$

$$\delta_{k+1} = \|x^{(k)} - \eta Q x^{(k)} + \eta b - x^*\|_2$$

$$= \|x^{(k)} - x^* - \eta Q x^{(k)} + \eta Q x^*\|_2$$

$$= \|\underbrace{(I - \eta Q)}_{S} \underbrace{(x^{(k)} - x^*)}_{y}\|_2$$

We showed $\|Sy\|_2 \leq \left(\max_{i=1:n} |\mu_i|\right) \|y\|_2$

where μ_i are eigenvalues of S

Here eigenvalues of $I-\eta Q$ are $1-\eta \lambda_i$ where λ_i are eigenvalues of Q

$$\therefore \delta_{k+1} = \|(I-\eta Q)(x^{(k)}-x^*)\|_2$$

$$\leq \left(\max_{i=1:n} |1-\eta\lambda_i|\right) \|x^{(k)}-x^*\|_2$$

$$\Rightarrow \boxed{\begin{array}{l} \delta_{k+1} \leq \gamma(\eta)\, \delta_k \\ \text{where } \gamma(\eta) = \max_{i=1:n} |1-\eta\lambda_i| \end{array}}$$

By induction,

$$\delta_1 \leq \gamma(\eta)\,\delta_0$$

$$\delta_k \leq \gamma(\eta)\,\delta_{k-1} \leq \gamma(\eta)^2\,\delta_{k-2} \leq \gamma(\eta)^3\,\delta_{k-3} \ldots$$

$$\Rightarrow \boxed{\delta_k \leq \gamma(\eta)^k\,\delta_0}$$

We want $\lim_{k\to\infty} x^{(k)} = x^*$

$$\Rightarrow \lim_{k\to\infty} \delta_k = 0$$

$$\Rightarrow \text{req'd } 0 < \gamma(\eta) < 1$$

$$\Rightarrow \max_{i=1:n} |1-\eta\lambda_i| < 1$$

$$\Rightarrow \forall\, i=1:n \quad 0 < \eta\lambda_i < 2$$

$$\Rightarrow \forall\, i=1:n \quad 0 < \eta < 2/\lambda_i$$

f) $d_k = -\nabla f(x^{(k)})$

$$\eta_k = \arg\min_{\eta\geq 0} f(x^{(k)}+\eta d_k)$$

$$f(x^{(k)}) = \frac{1}{2} x^{(k)T} Q x^{(k)} - b^T x^{(k)}$$

$$\nabla f(x^{(k)}) = Q x^{(k)} - b$$

$$\Rightarrow d_k = -Q x^{(k)} + b \qquad \Rightarrow x^{(k)} = Q^{-1}(b-d_k)$$

η_k st $f(x^{(k)}+\eta d_k)$ minimized wrt η

$$x^{(k)} + \eta d_k = Q^{-1}b - Q^{-1}d_k + \eta d_k$$

$$\bullet\; f(x^{(k)}+\eta d_k) = \frac{1}{2}(Q^{-1}b - Q^{-1}d_k + \eta d_k)^T Q (Q^{-1}b - Q^{-1}d_k + \eta d_k) - b^T(Q^{-1}b - Q^{-1}d_k + \eta d_k)$$

$$\therefore f(x^k + \eta d_k) = \frac{1}{2}(Q^{-1}b - Q^{-1}d_k + \eta d_k)^T Q (\ldots)$$

$$\frac{\partial f(x^k + \eta d_k)}{\partial \eta} = 0$$

$$\Rightarrow d_k^T Q(Q^{-1}b - Q^{-1}d_k + \eta d_k) \quad - d_k^T b = 0$$

$$\Rightarrow \cancel{d_k^T b} - d_k^T d_k + \eta d_k^T Q \cdot d_k - \cancel{d_k^T b} = 0$$

$$\Rightarrow \quad \eta\, d_k^T Q d_k = \|d_k\|_2^2$$

$$\Rightarrow \boxed{\eta_k = \frac{\|d_k\|_2^2}{d_k^T Q d_k}}$$

g) $x^* = Q^{-1}b \qquad n=2 \quad \gamma=10$

$$f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2) \qquad Q : \mathbb{R}^{2\times 2}$$

$$= \frac{1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} \alpha & \beta \\ \delta & \varepsilon \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} b_1 & b_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} \alpha x_1 + \delta x_2 & \beta x_1 + \varepsilon x_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - (b_1 x_1 + b_2 x_2)$$

$$= \frac{1}{2}\left(\alpha x_1^2 + (\delta + \beta)x_1 x_2 + \varepsilon x_2^2\right) - (b_1 x_1 + b_2 x_2)$$

$$\Rightarrow \alpha = 1 \qquad \varepsilon = \gamma \qquad \beta = \delta = 0$$

$$b_1 = b_2 = 0$$

$$\Rightarrow x^* = \begin{bmatrix} 1 & 0 \\ 0 & \gamma \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\nabla f(x^{(k)}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} x_1 \\ \gamma x_2 \end{bmatrix}$$

$\lambda_i = \text{eig}(Q) = 1, \ \gamma$

reqd for const step size $0 < \eta < 2/\lambda_i \ \forall i$

$$\Rightarrow 0 < \eta < 2 \text{ and}$$

$$0 < \eta < 2/\gamma$$

So we pick $\eta < 2/10$

$\eta = 0.05$ for example

Optimal soln: $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Exact line search:
- finds soln faster
- for lower $\gamma = 2$, it overshoots the solution
- we see a zig zag close to the optimal solution where slight (+)ve or (−)ve gradients cause zig zagging due to sharp change in η

Constant step size:
- takes many iterations to find solution
- takes a smooth path to origin for both $\gamma = 10$ and $\gamma = 2$