LECTURE 11

MODIFIED [redacted] NEWTON'S METHODS

Suppose $\nabla^2 f(x_k) \not\succ 0$

* Quadratic model

$$Q(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k)$$

used by Newton's method has no local minimizer.

* Newton search direction is not necessarily a descent direction.

APPROACH: Modify $\nabla^2 f(x_k)$ so that it becomes positive definite.

MODIFICATIONS BASED ON IDENTITY SHIFTS

λ is an eigenvalue of $\nabla^2 f(x_k)$

$$\Longleftrightarrow$$

$$\nabla^2 f(x_k) v = \lambda v \quad \exists v \neq 0$$

$$\Longleftrightarrow$$

$$(\nabla^2 f(x_k) + \eta I) v = (\lambda + \eta) v \quad \exists v \neq 0$$ (1)

$$\Longleftrightarrow$$

$(\lambda + n)$ is an eigenvalue of $\nabla^2 f(x_k)$

Eigenvalues of $\nabla^2 f(x_k)$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

Eigenvalues of $\nabla^2 f(x_k) + nI$

$$\lambda_1 + n \geq \lambda_2 + n \geq \cdots \geq \lambda_n + n$$

OBSERVATION

Let λ_{min} be the smallest eigenvalue of $\nabla^2 f(x_k)$. The matrix $\nabla^2 f(x) + nI \succ 0$ for all $n > |\lambda_{min}|$.

EXAMPLE

$$f(x) = x_1^4 + x_1 x_2 + (1 + x_2)^2$$

$$\nabla^2 f(x) = \begin{bmatrix} 12x_1^2 & 1 \\ 1 & 2 \end{bmatrix}$$

Recall that at $x_k = (0,0)$ the Newton direction $p_k^N = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$ is not a descent direction.

Hessian at $x_k = (0,0)$

$$\nabla^2 f(x_k) = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} \not\succ 0$$

with eigenvalues $\lambda_{min} = 1-\sqrt{2}$, $\lambda_{max} = 1+\sqrt{2}$.

Shift Hessian by $\eta = 1 > |\lambda_{min}|$

$$\widetilde{\nabla^2 f(x_k)} = \nabla^2 f(x_k) + I$$

$$= \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix} \succ 0$$

(Note: $\lambda_{min} = 2-\sqrt{2}$ for $\widetilde{\nabla^2 f(x_k)}$)

Modified Newton direction

$$\widetilde{\nabla^2 f(x_k)}\, p_k^{NM} = -\nabla f(x_k)$$

$$\Longrightarrow \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix} p_k^{NM} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

$$\Longrightarrow p_k^{NM} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

p_k^{NM} is a descent direction at $x_k = (0,0)$.

$$\nabla f(x_k)^T p_k^{NM} = -2 < 0$$

ESTIMATION OF THE SMALLEST EIGENVALUE

Loose estimates for λ_{min} are given by the Frobenius norm and Gersgorin's thm.

THM

Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Then

$$|\lambda_{min}(A)| \leqslant \|A\|_2 \leqslant \|A\|_F$$

EXAMPLE

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$$

$$|\lambda_{min}(A)| = \sqrt{2} - 1 \approx 0.41 < \|A\|_F = \sqrt{0^2 + 1^2 + 1^2 + 2^2} = \sqrt{6}$$

THM (Gersgorin)

Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Define the Gersgorin intervals

$$\Gamma_i = [a_{ii} - r_i, \, a_{ii} + r_i] \text{ where } r_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|$$

for $i = 1, \ldots, n$. Then

$$\Lambda(A) = \{\lambda \in \mathbb{R} : Av = \lambda v \ \exists v \neq 0\} \subseteq \Gamma_1 \cup \Gamma_2 \cup \ldots \cup \Gamma_n$$

EXAMPLE

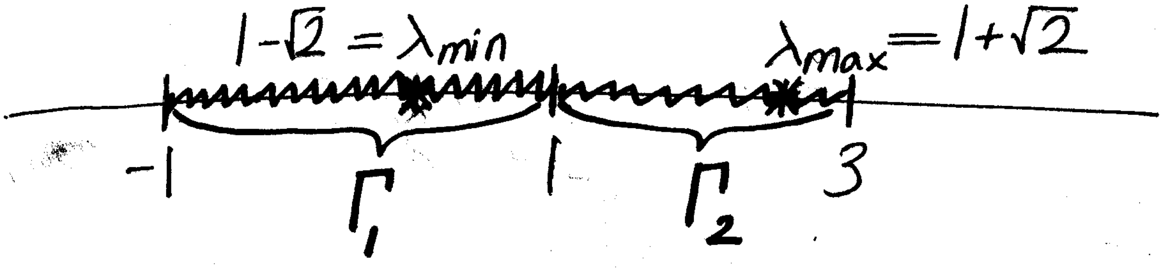
Gersgorin intervals for $A = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$

$$\Gamma_1 = [0-1,\ 0+1] = [-1,\ 1]$$

$$\Gamma_2 = [2-1,\ 2+1] = [1,\ 3]$$

Lower bound by Gersgorin's thm

$$\lambda_{min} \geqslant -1$$



OBSERVATION

$$\lambda_{min} \geqslant \min_{1 \leqslant i \leqslant n} a_{ii} - r_i$$

ALGORITHM (Newton's Method with Identity Shifts)

Given $x_0 \in \mathbb{R}^n$, $k=0$

While $\|\nabla f(x_k)\| > \epsilon$

Compute $\nabla^2 f(x_k)$

Find an n s.t. $n < \lambda_{min}(\nabla^2 f(x_k$
using Gersgorin's thm

If $n < 0$

Solve the linear system

$$(\nabla^2 f(x_k) + nI)\, p_k = -\nabla f(x_k)$$

for p_k

else

Solve

$$\nabla^2 f(x_k) p_k = -\nabla f(x_k)$$

for p_k

end

Calculate the step-length α_k using the Goldstein-Armijo backtracking line search

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

end

MODIFICATIONS BASED ON EIGENVALUE DECOMPOSITION

Let $A \in \mathbb{R}^{n \times n}$ be symmetric with

* real eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$
* the associated set of ~~orthogonal~~ orthonormal eigenvectors $v_1, v_2, \dots, v_n$

i.e.

(i) $A v_i = \lambda_i v_i \qquad i = 1, \dots, n$

(ii) $v_i^T v_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

Combine equalities in (i)

$$[A v_1 \;\; A v_2 \; \dots \; A v_n] = [\lambda_1 v_1 \;\; \lambda_2 v_2 \; \dots \; \lambda_n v_n]$$

$$\Longrightarrow A \underbrace{[v_1 \;\; v_2 \; \dots \; v_n]}_{V} = [v_1 \dots v_n] \underbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \; \lambda_n \end{bmatrix}}_{\Lambda}$$

Note that $V V^T = V^T V = I$ (that is V is an orthogonal matrix) implying

$$A V V^T = V \Lambda V^T \Longrightarrow A = V \Lambda V^T$$

THM (Orthogonal Eigenvalue Decomposition)

Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Then A has the decomposition

$$A = V \Lambda V^T$$

where $V \in \mathbb{R}^{n \times n}$ is orthogonal and $\Lambda \in \mathbb{R}^{n \times n}$ is diagonal.

EXAMPLE

$$A = \begin{bmatrix} -1 & 5 \\ 5 & -1 \end{bmatrix}$$

eigenvalues

$$\lambda_1 = +4, \quad \lambda_2 = -6$$

eigenvectors

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

eigenvalue decomposition

$$A = \underbrace{\left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right)}_{V} \underbrace{\begin{bmatrix} 4 & 0 \\ 0 & -6 \end{bmatrix}}_{\Lambda} \underbrace{\left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right)}_{V^T}$$

- Suppose $\nabla^2 f(x_k) \nsucc 0$. Then the orthogonal eigenvalue decomposition is of the form

$$\nabla^2 f(x_k) = V \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} V$$

where one of $\lambda_j \leqslant 0$.

Modified Hessian matrix

$$\widetilde{\nabla^2 f(x_k)} = V \tilde{\Lambda} V^T \succ 0$$

where

$$\tilde{\Lambda} = \begin{bmatrix} \tilde{\lambda}_1 & & 0 \\ & \ddots & \\ 0 & & \tilde{\lambda}_n \end{bmatrix} \quad \text{with} \quad \tilde{\lambda}_j = \begin{cases} \lambda_j & \text{if } \lambda_j > \\ \delta & \text{if } |\lambda_j| \\ -\lambda_j & \text{if } \lambda_j < \end{cases}$$

EXAMPLE

$$\nabla^2 f(x_k) = \begin{bmatrix} -1 & 5 \\ 5 & -1 \end{bmatrix}$$

$$= \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right) \begin{bmatrix} 4 & 0 \\ 0 & -6 \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right)$$

- Modified Hessian

$$\widetilde{\nabla^2 f(x_k)} = \left(\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\right)\begin{bmatrix}4 & 0\\ 0 & 6\end{bmatrix}\left(\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\right.$$

$$= \begin{bmatrix}5 & -1\\ -1 & 5\end{bmatrix}$$