Preliminaries

P1. The Jacobian

Let $F : \mathbb{R}^n \to \mathbb{R}^m$,

$$F(x) := \begin{bmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_m(x) \end{bmatrix}, \quad F_1, \ldots, F_m : \mathbb{R}^n \to \mathbb{R}$$

be such that $\dfrac{\partial F_j(x)}{\partial x_k}$ $j = 1, \ldots, m$ $k = 1, \ldots, n$ exist. The Jacobian $F' : \mathbb{R}^n \to \mathbb{R}^{m \times n}$ is defined by

$$F'(x) := \begin{bmatrix} \frac{\partial F_1(x)}{\partial x_1} & \frac{\partial F_1(x)}{\partial x_2} & \cdots & \frac{\partial F_1(x)}{\partial x_n} \\ \vdots & & & \\ \frac{\partial F_m(x)}{\partial x_1} & \frac{\partial F_m(x)}{\partial x_2} & \cdots & \frac{\partial F_m(x)}{\partial x_n} \end{bmatrix}.$$

<u>Ex</u>

$$F(x) = \begin{bmatrix} x_1 e^{x_2} \\ x_1^2 + x_2^2 \end{bmatrix}, \qquad F'(x) = \begin{bmatrix} e^{x_2} & x_1 e^{x_2} \\ 2x_1 & 2x_2 \end{bmatrix}$$

Generalized Product Rule

$F, G : \mathbb{R}^n \to \mathbb{R}^m$ (with well-defined Jacobian)

$$\left[F(x)^T G(x)\right]' = F(x)^T G'(x) + G(x)^T F'(x)$$

Ex : $q: \mathbb{R}^n \to \mathbb{R}$

$$(*) \quad q(x) = x^T A x$$

$$q'(x) = x^T [Ax]' + [Ax]^T [x]'$$

$$= x^T A + x^T A^T I_n$$

$$= x^T (A + A^T)$$

Remarks

Suppose $f: \mathbb{R}^n \to \mathbb{R}$

$$f'(x) = [\nabla f(x)]^T$$

$$[\nabla f(x)]' = \nabla^2 f(x)$$

Consequently, for $q(x)$ as in $(*)$

$$\nabla q(x) = (A + A^T)x$$

$$\nabla^2 q(x) = (A + A^T)$$

Generalized Chain Rule

$F: \mathbb{R}^n \to \mathbb{R}^m$, $G: \mathbb{R}^m \to \mathbb{R}^p$ are differentiable,

and $H: \mathbb{R}^n \to \mathbb{R}^p$, $H(x) = G(F(x))$

$$\underbrace{H'(x)}_{p \times n} = \underbrace{G'(F(x))}_{p \times m} \underbrace{F'(x)}_{m \times n} \quad \left(\text{order is important}\right)$$

Ex

$f: \mathbb{R}^n \to \mathbb{R}$ differentiable, $\underline{x}, \underline{p} \in \mathbb{R}^n$

$$\phi(\alpha) := f(\underline{x} + \alpha \underline{p})$$

$$\phi'(\alpha) = [f(\underline{x}+\alpha\underline{p})]' \frac{d(\underline{x}+\alpha\underline{p})}{d\alpha}$$

$$= \nabla f(\underline{x}+\alpha\underline{p})^T \underline{p}$$

$$\phi'(\alpha) = \nabla f(\underline{x}+\alpha\underline{p})^T \underline{p}$$

$$\phi''(\alpha) = \underline{p}^T [\nabla f(\underline{x}+\alpha\underline{p})]' \frac{d(\underline{x}+\alpha\underline{p})}{d\alpha}$$

$$= \underline{p}^T \nabla^2 f(\underline{x}+\alpha\underline{p}) \underline{p}$$

P2. Taylor's theorem

THM (Univariate Lagrange Remainder)

Let $f: \mathbb{R} \to \mathbb{R}$ be such that $f(x), \dots, f^{(n-1)}(x)$ are continuous on $[a,b]$ and $f^{(n)}(x)$ exists on (a,b). Then

$$f(b) = f(a) + f'(a)(b-a) + \dots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \frac{f^{n}(\gamma)}{n!}(b-a)^n$$

for some $\gamma \in (a,b)$.

<u>THM</u> (Multivariate Taylor's thm, 2nd Order)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, and $x, \underline{x} \in \mathbb{R}^n$.

(i) $$f(x) = f(\underline{x}) + \nabla f(\underline{x})^T p + \frac{1}{2} p^T \nabla^2 f(\underline{x} + tp) p$$

for some $t \in (0,1)$, where $p := x - \underline{x}$.

(ii) $$\nabla f(x) = \nabla f(\underline{x}) + \int_0^1 \nabla^2 f(\underline{x} + tp) p \, dt.$$

<u>PROOF</u>

(i) Apply univariate Taylor's thm to $\phi(\alpha) := f(\underline{x} + \alpha p)$ on $[0,1]$.

$$\phi(1) = \phi(0) + \phi'(0) + \frac{\phi''(t)}{2}$$

$$f(x) = f(\underline{x}) + \nabla f(\underline{x})^T p + \frac{1}{2} p^T \nabla^2 f(\underline{x} + tp) p$$

for some $t \in (0,1)$.

(ii) Define $L(\alpha) := \nabla f(\underline{x} + \alpha p)$ and notice

$$L'(\alpha) = \nabla^2 f(\underline{x} + \alpha p) p$$

by the generalized chain rule.

Thus by the fundamental thm

$$L(1) = L(0) + \int_0^1 L'(t) \, dt$$

$$\nabla f(x) = \nabla f(\underline{x}) + \int_0^1 \nabla^2 f(\underline{x} + tp) p \, dt.$$

$\square$

P3. Rayleigh Characterization of Extreme Eigenvalues

THM

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

$$\lambda_{max}(A) = \max_{\substack{v \in \mathbb{C}^n \\ \|v\|_2 = 1}} v^T A v$$

$$\lambda_{min}(A) = \min_{\substack{v \in \mathbb{C}^n \\ \|v\|_2 = 1}} v^T A v$$

PROOF

Let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ be eigenvalues of A and $\{v_1, v_2, \ldots, v_n\}$ be the corresponding set of orthonormal eigenvectors. Any unit vector $v \in \mathbb{R}^n$ can be written of the form

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

where $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ such that $\alpha_1^2 + \cdots + \alpha_n^2 = 1$.

(i.e. $v = V\alpha$ where $V = [v_1 \; v_2 \cdots v_n]$ is s.t. $V^T V = I$ and $\alpha = [\alpha_1 \cdots \alpha_n]^T$, so $\|\alpha\|_2 = \|V\alpha\|_2 = 1$)

Consequently,

$$\begin{aligned} v^T A v &= (\alpha_1 v_1 + \cdots + \alpha_n v_n)^T A (\alpha_1 v_1 + \cdots + \alpha_n v_n) \\ &= (\alpha_1 v_1 + \cdots + \alpha_n v_n)^T (\alpha_1 \lambda_1 v_1 + \cdots + \alpha_n \lambda_n v_n) \\ &= \lambda_1 \alpha_1^2 + \cdots \lambda_n \alpha_n^2 \end{aligned}$$

meaning

$$\lambda_n = \lambda_n(\alpha_1^2 + \cdots + \alpha_n^2) \leqslant v^T A v \leqslant \lambda_1 (\alpha_1^2 + \cdots + \alpha_n^2) = \lambda_1$$

Furthermore, $v_n^T A v_n = \lambda_n$ and $v_1^T A v_1 = \lambda_1$, so the result follows.

$\square$

It follows from these characterizations that, if $A \in \mathbb{R}^{n \times n}$ is symmetric, for each $v \in \mathbb{R}^n$ we have,

$$\|v\|^2 \lambda_{min}(A) \leqslant v^T A v = \|v\|^2 \left(\left[\frac{v}{\|v\|}\right]^T A \left[\frac{v}{\|v\|}\right] \right)$$
$$\leqslant \|v\|^2 \lambda_{max}(A).$$

P4. 2-norm and Frobenius norm for Matrices

<u>DEFN</u>

Let $A \in \mathbb{R}^{m \times n}$.

$$\|A\|_2 := \text{maximize} \{ \|Av\|_* \mid v \in \mathbb{R}^n, \|v\|_* = 1 \}$$
$$\|A\|_F := \sqrt{\sum_{j=1}^{m} \sum_{k=1}^{n} a_{jk}^2}$$

Notice the property

$$\|Av\|_* = \|v\| \left\| A \frac{v}{\|v\|} \right\|$$
$$\leqslant \|v\| \cdot \|A\|_2$$

for every $v \in \mathbb{R}^n$.

THM

Suppose $A \in \mathbb{R}^{n \times n}$ is symmetric and positive definite.

$$\|A\|_2 = \lambda_{max}(A)$$

PROOF

Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of A and $\{v_1, \dots, v_n\}$ be the corresponding set of orthonormal eigenvectors. For every $v \in \mathbb{R}^n$ of unit length with expansion

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

such that $\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2 = 1$, we have

$$\begin{aligned}
\|Av\| &= \|A(\alpha_1 v_1 + \cdots + \alpha_n v_n)\| \\
&= \|\alpha_1 \lambda_1 v_1 + \cdots + \alpha_n \lambda_n v_n\| \\
&= \Bigg\| \underbrace{\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}}_{V} \begin{bmatrix} \alpha_1 \lambda_1 \\ \vdots \\ \alpha_n \lambda_n \end{bmatrix} \Bigg\| \\
&\underset{\substack{\text{since} \\ V^T V = I_n}}{=} \Bigg\| \begin{bmatrix} \alpha_1 \lambda_1 \\ \vdots \\ \alpha_n \lambda_n \end{bmatrix} \Bigg\| = \sqrt{(\alpha_1 \lambda_1)^2 + \cdots + (\alpha_n \lambda_n)^2} \\
&\leq \sqrt{\lambda_1^2 (\alpha_1^2 + \cdots + \alpha_n^2)} = \lambda_1
\end{aligned}$$

Moreover $\|Av_1\| = \|\lambda_1 v_1\| = \lambda_1$ completing the proof. $\square$

If $D \in \mathbb{R}^{m \times n}$
$\mathrm{Trace}(D) := \sum_{j=1}^{p} d_{jj}$
$p := \min(m,n)$

THM

For every $A \in \mathbb{R}^{m \times n}$

$$\|A\|_F = \sqrt{\mathrm{Trace}\,(A^T A)}\,.$$

PROOF

$$\|A\|_F = \sqrt{\sum_{k=1}^{n} \sum_{j=1}^{m} a_{jk}^2}$$

$$= \sqrt{\sum_{k=1}^{n} \|a_k\|^2}$$

$$= \sqrt{\sum_{k=1}^{n} \underbrace{a_k^T a_k}_{d_{kk}}} \quad \text{where } D = A^T A$$

$$= \sqrt{\mathrm{Trace}\,(A^T A)}\,. \qquad \square$$

P5. Order of convergence

Let $\{x^{(k)}\}$ be a sequence in $\mathbb{R}^n$, such that $\lim_{k \to \infty} x^{(k)} = x_*$.

We say that the order of convergence is p (a positive integer) if there exists a constant $c > 0$ (and < 1 if $p = 1$) such that

$$\frac{\| x_{k+1} - x_* \|}{\| x_k - x_* \|^p} \leq c$$

for all k sufficiently large.

linear convergence — $p = 1$ (and $c < 1$)

quadratic convergence — $p = 2$

cubic convergence — $p = 3$

We say that the sequence $\{x^{(k)}\}$ converges to x_* superlinearly if

$$\lim_{k\to\infty} \frac{\|x^{(k+1)} - x_*\|}{\|x^{(k)} - x_*\|} = 0 .$$

Examples

$$\{2^{-k}\} = \{1, 1/2, 1/4, 1/8, \ldots\}$$

$$\lim_{k\to\infty} 2^{-k} = 0, \qquad \frac{|2^{-(k+1)} - 0|}{|2^{-k} - 0|} = \frac{1}{2}$$

converges linearly

$$\{10^{-2^k}\} = \left\{\frac{1}{10}, \frac{1}{100}, 10^{-4}, 10^{-8}, 10^{-16}, \ldots\right\}$$

$$\lim_{k\to\infty} 10^{-2^k} = 0, \qquad \frac{|10^{-2^{(k+1)}} - 0|}{|10^{-2^k} - 0|^2} = 1$$

converges quadratically

$$\{1/k!\} = \left\{1, 1, 1/2, 1/6, \frac{1}{24}, \frac{1}{120}, \ldots\right\}$$

$$\lim_{k\to\infty} \frac{1}{k!} = 0, \quad \lim_{k\to\infty} \frac{|1/(k+1)! - 0|}{|1/k! - 0|} = \lim_{k\to\infty} \frac{1}{k+1} = 0$$

converges superlinearly