Vector norms: assume $x \in \mathbb{R}^n$

1. 1-norm: $\|x\|_1 \triangleq \sum_{i=1}^{n} |x_i|$

2. 2-norm: $\|x\|_2 \triangleq \sqrt{\sum_{i=1}^{n} x_i^2}$

   often denoted as $\|\cdot\|$

3. ∞-norm: $\|x\|_\infty \triangleq \max_{1 \le i \le n} |x_i|$

4. p-norm: $\|x\|_p \triangleq \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$, for $1 \le p < \infty$

In $\mathbb{R}^n$, a function $\phi : \mathbb{R}^n \to \mathbb{R}$ defines a norm if:

(1) $\phi(x) \geq 0, \ \forall x \in \mathbb{R}^n$
and $\phi(x) = 0 \iff x = \vec{0}$

$\Big\}$ $\phi(x) > 0$, if $x \neq \vec{0}$

(2) $\phi(\alpha x) = |\alpha| \phi(x), \ \forall x \in \mathbb{R}^n, \forall \alpha \in \mathbb{R}$

(3) $\phi(x+y) \le \phi(x) + \phi(y), \ \forall x \in \mathbb{R}^n, y \in \mathbb{R}^n$

Examples:

1. verify $\|x\|_1$ satisfies the three conditions above.

2. $\|x\|_p$ is not a norm if $0 < p < 1$

   Take $p = \frac{1}{2}$ for instance.

(1) $\|x\|_{\frac{1}{2}} \geq 0, \quad \forall x \in \mathbb{R}^n$

$\|x\|_{\frac{1}{2}} = 0 \iff x = \vec{0}$

$$\left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} = 0 \iff \sum_{i=1}^{n} |x_i|^p = 0 \iff |x_i|^p = 0, \forall i$$

$$\iff |x_i| = 0 \iff x_i = 0 \quad \forall i.$$

(2) $\|\alpha x\|_{\frac{1}{2}} = |\alpha| \cdot \|x\|_{\frac{1}{2}}$

$$\|\alpha x\|_{\frac{1}{2}} = \left(\sum_{i=1}^{n} |\alpha x_i|^{\frac{1}{2}} \right)^2 = \left(\sum_{i=1}^{n} |\alpha|^{\frac{1}{2}} |x_i|^{\frac{1}{2}} \right)^2$$

$$= \left(|\alpha|^{\frac{1}{2}} \sum_{i=1}^{n} |x_i|^{\frac{1}{2}} \right)^2 = |\alpha| \left(\sum_{i=1}^{n} |x_i|^{\frac{1}{2}} \right)^2$$

$$= |\alpha| \cdot \|x\|_{\frac{1}{2}}$$

(3) Pick $x = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad y = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad x+y = \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

$$\|x\|_{\frac{1}{2}} = 1, \quad \|y\|_{\frac{1}{2}} = 1, \quad \|x+y\|_{\frac{1}{2}} = \left(\sum_{i=1}^{n} |x_i + y_i|^{\frac{1}{2}} \right)^2$$

$$= 2^2 = 4$$

$$\|x+y\|_{\frac{1}{2}} > \|x\|_{\frac{1}{2}} + \|y\|_{\frac{1}{2}}$$

Therefore, $\|\cdot\|_{\frac{1}{2}}$ does not define a norm.

Weighted vector norm: given a matrix $A \in \mathbb{R}^{n\times n}$, if A is symmetric and positive definite, then

$$\phi(x) = \|x\|_A \triangleq \sqrt{x^T A x}, \quad \forall x \in \mathbb{R}^n$$

defines a norm.

Recall: A symmetric if $A^T = A$

A is positive definite if $x^T A x > 0, \ \forall x \neq \vec{0}$

A is positive semi-definite if $x^T A x \geq 0, \ \forall x$.

Matrix norm: assume $X \in \mathbb{R}^{m\times n}$, $m \leq n$

$$X = \begin{bmatrix} | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} - & \tilde{x}_1^T & - \\ - & \tilde{x}_2^T & - \\ & \vdots & \\ - & \tilde{x}_m^T & - \end{bmatrix}$$

1. matrix 1-norm:

$$\|X\|_1 \triangleq \max_{\|y\|_1 = 1} \|Xy\|_1 = \max_{1 \leq j \leq n} \|\vec{x}_j\|_1$$

2. matrix ∞-norm:

$$\|X\|_\infty \triangleq \max_{\|y\|_\infty = 1} \|Xy\|_\infty = \max_{1 \leq i \leq m} \|\tilde{x}_i\|_1$$

3. Matrix 2-norm:

$$\|X\|_2 \triangleq \max_{\|y\|_2=1} \|Xy\|_2 = \underbrace{\sigma_{max}(X)}$$

maximum singular value of X

4. Frobenius norm:

$$\|X\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |x_{ij}|^2}$$

5. nuclear norm:

$$\|X\|_* \triangleq \sum_{i=1}^{m} \underbrace{\sigma_i}(X)$$

$\sigma_i(X)$ is the i-th singular value of X

Example:

$$X = \begin{array}{cc} \begin{bmatrix} -1 & 2 & 0 \\ 3 & -3 & 1 \end{bmatrix} & \begin{array}{c} \leftarrow 3 \\ \leftarrow 7 \end{array} \\ \begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ 4 & 5 & 1 \end{array} & \end{array}$$

① what is $\|X\|_1$? $\quad \|X\|_1 = 5$

② what is $\|X\|_\infty$? $\quad \|X\|_\infty = 7$

③ what is $\|X\|_F$? $\quad \|X\|_F = \sqrt{(-1)^2+2^2+0^2+3^2+(-3)^2+1^2} = \sqrt{24}$

④ What is $\|X\|_2$ and $\|X\|_*$?

Compute eigenvalues of

$$XX^T = \begin{bmatrix} -1 & 2 & 0 \\ 3 & -3 & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 \\ 2 & -3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & -9 \\ -9 & 19 \end{bmatrix}$$

$$\det(\lambda I - XX^T) = \det\left(\begin{bmatrix} \lambda-5 & 9 \\ 9 & \lambda-19 \end{bmatrix} \right) = 0$$

Assume the roots are $\lambda_1 \geq \lambda_2 > 0$

Then $\|X\|_2 = \sqrt{\lambda_1}$, $\|X\|_* = \sqrt{\lambda_1} + \sqrt{\lambda_2}$

Properties about matrix norm

① $\|Ax\|_1 \leq \|A\|_1 \cdot \|x\|_1$, $\forall x \in \mathbb{R}^n$, $\forall A \in \mathbb{R}^{m \times n}$

because $\|A\|_1 = \max\limits_{\|y\|_1 = 1} \|Ay\|_1 = \max\limits_{y \neq 0} \dfrac{\|Ay\|_1}{\|y\|_1} \geq \dfrac{\|Ax\|_1}{\|x\|_1}$

$\left\| \dfrac{y}{\|y\|_1} \right\|_1 = 1$

② $\|Ax\|_\infty \leq \|A\|_\infty \cdot \|x\|_\infty$

③ $\|Ax\| \leq \|A\|_2 \cdot \|x\|$

④ $\|AB\|_F \leq \|A\|_2 \cdot \|B\|_F$

Gradient: Let $f: \mathbb{R}^n \to \mathbb{R}$ be a differentiable function

Its gradient is defined:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix} \in \mathbb{R}^n$$

Example: Let $f(x) = \frac{1}{2} x^T Q x + c^T x$, where $Q = Q^T \in \mathbb{R}^{n \times n}$

What is $\nabla f(x)$?

First, let $Q = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, $c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Then $f(x) = \frac{1}{2} [x_1, x_2] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [1, 1] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$= \frac{1}{2} (2x_1^2 + 2x_1x_2 + 2x_2^2) + (x_1 + x_2)$$

$\frac{\partial f}{\partial x_1}(x) = 2x_1 + x_2 + 1$, $\frac{\partial f}{\partial x_2}(x) = x_1 + 2x_2 + 1$

So $\nabla f(x) = \begin{bmatrix} 2x_1 + x_2 + 1 \\ x_1 + 2x_2 + 1 \end{bmatrix} = Qx + c$

Claim: for general $n \geq 1$, $Q = Q^T \in \mathbb{R}^{n\times n}$, $c \in \mathbb{R}^n$

$$\nabla f(x) = \nabla\left(\tfrac{1}{2}x^T Q x + c^T x\right) = Qx + c$$

Remark: for $f(x) = \frac{1}{2}x^T Q x + c^T x + d$, where $Q \in \mathbb{R}^{n\times n}$

but Q may not be symmetric,

$$\nabla f(x) = \tfrac{1}{2}(Q + Q^T)x + c$$

Pf: $f(x) = \frac{1}{2}x^T Q x + c^T x + d$

$$= \tfrac{1}{4}x^T Q x + \tfrac{1}{4}x^T Q x + c^T x + d$$

$$= \tfrac{1}{4}x^T Q x + \tfrac{1}{4}(x^T Q x)^T + c^T x + d$$

$$= \tfrac{1}{4}x^T Q x + \tfrac{1}{4}x^T Q^T x + c^T x + d$$

$$= \tfrac{1}{2}x^T\left(\tfrac{1}{2}(Q + Q^T)\right)x + c^T x + d$$

Example: Let $f(x) = \frac{1}{2}\|Ax - b\|^2$. What is $\nabla f(x)$?

Solution:

$$f(x) = \tfrac{1}{2}(Ax - b)^T(Ax - b)$$

$$= \tfrac{1}{2}(x^T A^T - b^T)(Ax - b)$$

$$= \tfrac{1}{2}\Big(x^T A^T A x - \underbrace{x^T A^T b}_{\langle A^T b,\, x\rangle} - \underbrace{b^T A x}_{\langle A^T b,\, x\rangle} + b^T b\Big)$$

$$= \frac{1}{2} x^T A^T A x - (A^T b)^T x + \frac{1}{2} b^T b$$

so $\nabla f(x) = A^T A x - A^T b = A^T (Ax - b)$

<u>Hessian Matrix</u>: assume $f: \mathbb{R}^n \to \mathbb{R}$ is 2nd-order differentiable

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_1 \partial x_i} & \frac{\partial^2 f}{\partial x_2 \partial x_i} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_i} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

$\leftarrow$ ith row is $\left(\nabla \frac{\partial f}{\partial x_i}(x)\right)^T$

Example: Let $f(x) = \frac{1}{2} x^T Q x + c^T x + d$, where $Q = Q^T \in \mathbb{R}^{n \times n}$

what is $\nabla^2 f(x)$? $\leftarrow$ $\nabla^2 f(x) = Q = \frac{1}{2}(Q + Q^T)$

First, let $Q = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, $c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Then $\frac{\partial f}{\partial x_1}(x) = 2x_1 + x_2 + 1$, $\frac{\partial f}{\partial x_2}(x) = x_1 + 2x_2 + 1$

$$\nabla \frac{\partial f}{\partial x_1}(x) = \nabla(2x_1 + x_2 + 1) = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

$$\nabla \frac{\partial f}{\partial x_2}(x) = \nabla(x_1 + 2x_2 + 1) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

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

Example: for $f(x) = \frac{1}{2}\|Ax - b\|^2$

Recall $f(x) = \frac{1}{2} x^T A^T A x - (A^T b)^T x + \frac{1}{2} b^T b$

So $\nabla^2 f(x) = A^T A$

Example: Let $f(x) = \log(1 - \exp(a^T x + b))$, $\forall x \in \mathbb{R}^n$

Take $a = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $b = -1$

$$f(x) = \log(1 - \exp(x_1 + 2x_2 - 1))$$

$$\frac{\partial f}{\partial x_1}(x) = \frac{1}{1 - \exp(x_1 + 2x_2 - 1)} \cdot \frac{\partial(1 - \exp(x_1 + 2x_2 - 1))}{\partial x_1}$$

$$= \frac{1}{1 - \exp(x_1 + 2x_2 - 1)} \left(-\exp(x_1 + 2x_2 - 1)\right) \cdot \frac{\partial(x_1 + 2x_2 - 1)}{\partial x_1}$$

$$= \frac{1}{1 - \exp(x_1 + 2x_2 - 1)} \left(-\exp(x_1 + 2x_2 - 1)\right)$$

$$= \frac{-1}{\exp(-x_1 - 2x_2 + 1) - 1} = \frac{1}{1 - \exp(-x_1 - 2x_2 + 1)}$$

$$\frac{\partial f}{\partial x_2}(x) = \frac{1}{1-\exp(x_1+2x_2-1)} \cdot \frac{\partial(1-\exp(x_1+2x_2-1))}{\partial x_2}$$

$$= \frac{1}{1-\exp(x_1+2x_2-1)} \left(-\exp(x_1+2x_2-1)\right) \cdot \frac{\partial(x_1+2x_2-1)}{\partial x_2}$$

$$= \frac{1}{1-\exp(x_1+2x_2-1)} \left(-\exp(x_1+2x_2-1)\right) \cdot 2$$

$$= \frac{-2}{\exp(-x_1-2x_2+1)-1} = \frac{2}{1-\exp(-x_1-2x_2+1)}$$

So

$$\nabla f(x) = \begin{bmatrix} \dfrac{1}{1-\exp(-x_1-2x_2+1)} \\[2ex] \dfrac{2}{1-\exp(-x_1-2x_2+1)} \end{bmatrix} = \frac{1}{1-\exp(-x_1-2x_2+1)} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$= \frac{1}{1-\exp(-a^T x - b)}\, a$$

$$\frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right)$$

$$= -\frac{1}{(1-\exp(-x_1-2x_2+1))^2} \cdot \frac{\partial(1-\exp(-x_1-2x_2+1))}{\partial x_1}$$

$$= -\frac{1}{\left(1-\exp(-x_1-2x_2+1)\right)^2}\left(-\exp(-x_1-2x_2+1)\right)\cdot\frac{\partial(-x_1-2x_2+1)}{\partial x_1}$$

$$= -\frac{1}{\left(1-\exp(-x_1-2x_2+1)\right)^2}\left(-\exp(-x_1-2x_2+1)\right)\cdot(-1)$$

$$= -\frac{\exp(-x_1-2x_2+1)}{\left(1-\exp(-x_1-2x_2+1)\right)^2}$$

$$\frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right)$$

$$= -\frac{1}{\left(1-\exp(-x_1-2x_2+1)\right)^2}\cdot\frac{\partial\left(1-\exp(-x_1-2x_2+1)\right)}{\partial x_2}$$

$$= -\frac{1}{\left(1-\exp(-x_1-2x_2+1)\right)^2}\left(-\exp(-x_1-2x_2+1)\right)\cdot\frac{\partial(-x_1-2x_2+1)}{\partial x_2}$$

$$= -\frac{1}{\left(1-\exp(-x_1-2x_2+1)\right)^2}\left(-\exp(-x_1-2x_2+1)\right)\cdot(-2)$$

$2\cdot\exp(-x_1-2x_2+1)$

$$= - \frac{\ldots}{\left(1-\exp(-x_1-2x_2+1)\right)^2}$$

So
$$\nabla\left(\frac{\partial f}{\partial x_1}\right) = - \frac{\exp(-x_1-2x_2+1)}{\left(1-\exp(-x_1-2x_2+1)\right)^2} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Check after class!

$$\nabla\left(\frac{\partial f}{\partial x_2}\right) = - \frac{\exp(-x_1-2x_2+1)}{\left(1-\exp(-x_1-2x_2+1)\right)^2} \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

$$\nabla^2 f(x) = - \frac{\exp(-x_1-2x_2+1)}{\left(1-\exp(-x_1-2x_2+1)\right)^2} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

$$= - \frac{\exp(-a^T x - b)}{\left(1-\exp(-a^T x - b)\right)^2} \, a a^T$$