Lecture 1

E1 260

Mathematical background # 1

Roadmap :

- Vector spaces :
  - Define
  - Norms , inner products, Cauchy - Schwarz
  - Subsets of $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$
  - Linear transformation
  - Linear / Affine spaces
- Calculus :
  - vector derivatives
  - Mean value theorem
  - Subgradients and subdifferentials
- Convergence of a sequence : linear, sublinear, superlinear
- level sets, contour plots, α - sublevel sets

Vector Spaces

Definition: Set of elements called "vectors"

For any two vectors $\underline{x}, \underline{y} \in \mathbb{E}$, the following holds.

1. $\underline{x} + \underline{y} = \underline{y} + \underline{x}$.   2. $x + (y + z) = (x + y) + z$

3. $\underline{x} + \underline{0} = \underline{x}$; $\underline{0}$ is the zero vector

4. $\underline{x} + (-\underline{x}) = \underline{0}$

$$\left.\begin{array}{ll}
5. & \alpha(\beta \underline{x}) = (\alpha\beta)\underline{x} \\
 & \quad \Rightarrow \; 1\,\underline{x} = \underline{x} \\
6. & \alpha(\underline{x} + \underline{y}) = \alpha\underline{x} + \alpha\underline{y} \\
7. & (\alpha + \beta)\underline{x} = \alpha\underline{x} + \beta\underline{y}
\end{array}\right\} \text{ for any } \alpha, \beta \in \mathbb{R}$$

Dimension :

- **Linear independent**

$\{ \underline{v}_1, \underline{v}_2, \ldots \underline{v}_n \}$ in a vector space $\mathbb{E}$ is linearly independent

$$\sum_{i=1}^{n} \alpha_i \underline{v}_i = 0 \iff \alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$$

- **Span**

$\{ \underline{v}_1, \underline{v}_2, \ldots \underline{v}_n \}$ is said to span $\mathbb{E}$

if for any $\underline{x} \in \mathbb{E}$ $\exists\ \beta_1, \beta_2 \ldots \beta_n \in \mathbb{R}$ such that

$$\underline{x} = \sum_{i=1}^{n} \beta_i \underline{v}_i$$

- A **Basis** of $\mathbb{E}$: independent set of vectors that spans $\mathbb{E}$
- $\dim(\mathbb{E})$: number of vectors

Norms: measure vectors

$\|\cdot\| : \mathbb{E} \to \mathbb{R}$ satisfying

1. $\|\underline{x}\| \geq 0$ and $\|\underline{x}\| = 0$ iff $\underline{x} = 0$
2. $\|\lambda \underline{x}\| = |\lambda| \cdot \|\underline{x}\|$
3. $\|\underline{x} + \underline{y}\| \leq \|\underline{x}\| + \|\underline{y}\|$

Inner products:

fn. that associates to each pair of vectors $\underline{x}, \underline{y} \in \mathbb{E}$ a real number

$$\boxed{\langle \underline{x}, \underline{y} \rangle = \underline{x}^T \underline{y}}$$

1. $\langle \underline{x}, \underline{y} \rangle = \langle \underline{y}, \underline{x} \rangle$
2. $\langle \alpha_1 \underline{x}_1 + \alpha_2 \underline{x}_2, \underline{y} \rangle = \alpha_1 \langle \underline{x}_1, \underline{y} \rangle + \alpha_2 \langle \underline{x}_2, \underline{y} \rangle$
3. $\langle \underline{x}, \underline{x} \rangle \geq 0$ and $\langle \underline{x}, \underline{x} \rangle = 0$ iff $\underline{x} = 0$

#3

Finite - dimensional vectors:

- Euclidean Space : finite - dimensional vector space equipped with $\langle \cdot , \cdot \rangle$

$\mathbb{R}^n$: $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$; $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$

(n is a +ve integer)

$$\lambda \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{bmatrix}$$

- $\langle \underline{x} , \underline{y} \rangle = \sum_{i=1}^{n} x_i y_i$
- $\sqrt{\langle \underline{x} , \underline{x} \rangle} = \| \underline{x} \| = \sqrt{\sum_{i=1}^{n} x_i^2}$: Euclidean norm
- $\underbrace{\langle \underline{x} , \underline{y} \rangle_Q}_{Q\text{-inner product}} = \underline{x}^T Q \, \underline{y}$ and $\underbrace{\| \underline{x} \|_Q}_{Q\text{-norm}} = \underline{x}^T Q \, \underline{x}$

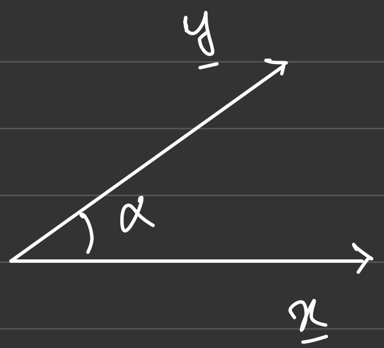
Cauchy - Schwarz inequality:

$$|\underline{x}^T \underline{y}| \leq \|\underline{x}\| \, \|\underline{y}\|$$

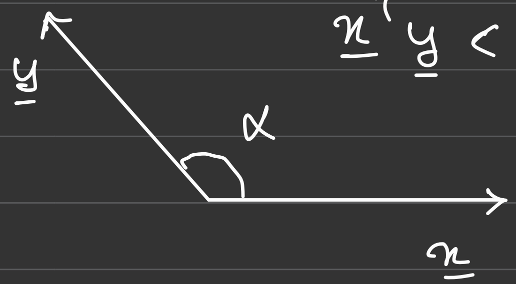
For non-zero vectors:

$$-1 \leq \frac{\underline{x}^T \underline{y}}{\|\underline{x}\| \, \|\underline{y}\|} \leq 1$$

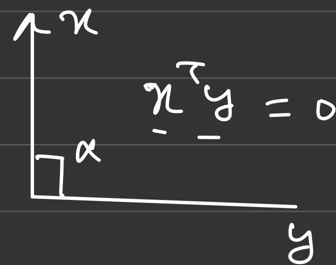
$$\cos(\alpha) = \frac{\underline{x}^T \underline{y}}{\|\underline{x}\| \, \|\underline{y}\|}$$



$x^T y > 0$



$\underline{x}^T \underline{y} < 0$



$\underline{x}^T \underline{y} = 0$

$\underline{x} = \underline{y}$

$\underline{x}^T \underline{y} = 1$

$\alpha = \pi$

$\underline{x} = -\underline{y}$

$\underline{x}^T \underline{y} = -1$

- l_p - norm : For any $p \geq 1$

$$\|\underline{x}\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$$

→ $p = 2$, l_2 - norm or Euclidean norm
→ $p = 1$, l_1 - norm

$p = 0$

$$\|\underline{x}\|_0 = nnz\{x_1, \ldots, x_n\}$$

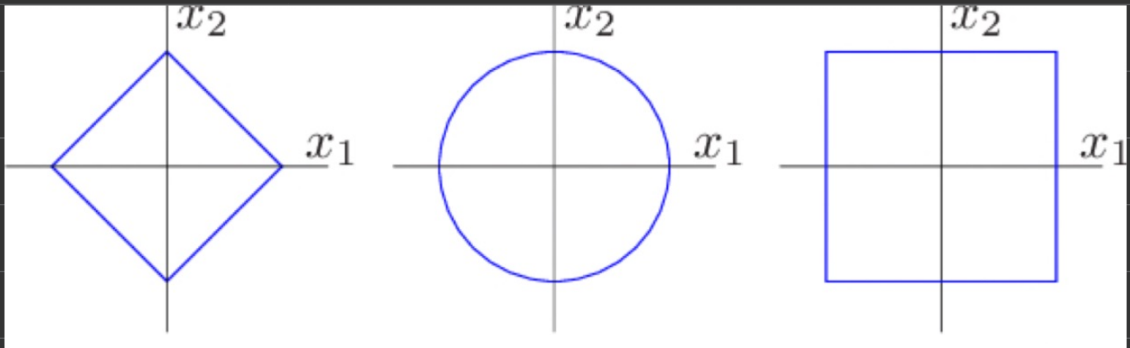
$p = \infty$

$$\|\underline{x}\|_\infty = \max_{i=1,2,\ldots,n} |x_i|$$

Subsets of $\mathbb{R}^n$:

norm ball :

$$B(c, r) = \{\underline{x} \in \mathbb{R}^n : \|\underline{x} - \underline{c}\| \leq r\}$$

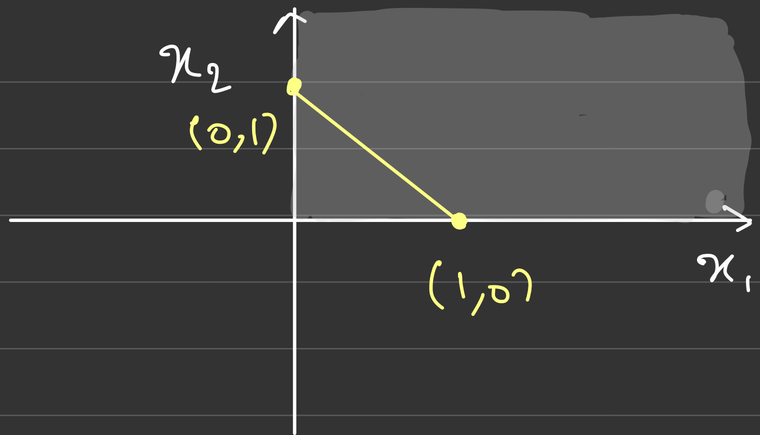


(a) ℓ_1-ball (b) ℓ_2-ball (c) ℓ_∞-ball

#6

- non-negative orthant:

$$\mathbb{R}^n_+ = \{ \underline{x} \in \mathbb{R}^n : \underline{x} \succeq 0 \}$$



- unit Simplex:

$$\Delta_n = \{ \underline{x} \in \mathbb{R}^n : \underline{x} \succeq 0 , \underline{1}^T \underline{x} = 1 \}$$

- Box:

$$\text{Box}[\underline{l}, \underline{u}] = \{ \underline{x} \in \mathbb{R}^n : \underline{l} \preceq \underline{x} \preceq \underline{u} \}$$

$\Rightarrow$ The Space $\mathbb{R}^{m \times n}$:

Set of all real-valued $m \times n$ matrices

$$\langle \underset{m \times n}{A} , \underset{m \times n}{B} \rangle = \mathrm{Tr}(\underbrace{A^T B}_{n \times n}) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ij}$$

Subspaces associated with $\mathbb{R}^{m \times n}$

- Null space :
$$\mathcal{N}(A) = \{ \underline{x} \in \mathbb{R}^n : A\underline{x} = \underline{0} \}$$

- column space :
(range space)
$$\mathcal{R}(A) = \{ \underline{x} \in \mathbb{R}^m : \underline{x} = A\underline{y} , \underline{y} \in \mathbb{R}^n \}$$

For $m = n$:

- Set of all symmetric matrices:
$$\mathbb{S}^n = \{ A \in \mathbb{R}^{n \times n} : A = A^T \}$$

- Set of positive Semi definite matrices
$$\mathbb{S}^n_+ = \{ A \in \mathbb{R}^{n \times n} : A \succeq 0 \}$$

$$\mathbb{S}^n_{++} = \{ A \in \mathbb{R}^{n \times n} : A \succ 0 \}$$

Norms in $\mathbb{R}^{m\times n}$:

- **Frobenius norm**: $\|A\|_F = \sqrt{Tr(A^TA)} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2}$

- **Induced norm**: $\|A\|_{a,b} = \max_{\underline{x}} \{ \|A\underline{x}\|_b : \|\underline{x}\| \leq 1 \}$

Examples are:

→ Spectral norm:

$$\|A\|_2 = \|A\|_{2,2} = \sqrt{\lambda_{max}(A^TA)} = \sigma_{max}(A)$$

$$\|A\underline{x}\|_2 \leq \|A\|_2 \|\underline{x}\|_2$$

→ 1-norm: max. absolute column sum norm

$$\|A\|_1 = \max_{j=1,2,\ldots,n} \sum_{i=1}^{m} |A_{ij}|$$

→ ∞-norm: max. absolute row sum norm

$$\|A\|_\infty = \max_{i=1,\ldots m} \sum_{j=1}^{n} |A_{ij}|$$

Dual Space and norm:

Linear transformation : $\mathcal{A} : \mathbb{E} \to \mathbb{V}$ if it satisfies

$$\mathcal{A}(\alpha \underline{x} + \beta \underline{y}) = \alpha \mathcal{A}(\underline{x}) + \beta \mathcal{A}(\underline{y}) \; ; \; x, y \in \mathbb{E}$$

- linear functional on a vector space $\mathbb{E}$ is a linear transformation form $\mathbb{E}$ to $\mathbb{R}$
- Set of "all" linear functionals on $\mathbb{E}$ is called the dual space $\mathbb{E}^*$

For inner-product spaces:

$f \in \mathbb{E}^*$, there always exists $\underline{v} \in \mathbb{E}$ such that

$$f(\underline{x}) = \langle \underline{v}, \underline{x} \rangle$$

Dual norm :

$$\|y\|_* = \max_{x} \{ \underbrace{\langle \underline{y}, \underline{x} \rangle}_{f(\underline{x}) : \mathbb{E} \to \mathbb{R}} : \|\underline{x}\| \le 1 \}$$

$\underline{y} \in \mathbb{E}^*$

- Tells us how big $\underline{y}$ is relative to the norm of $\underline{x}$

Generalized Cauchy - Schwarz:

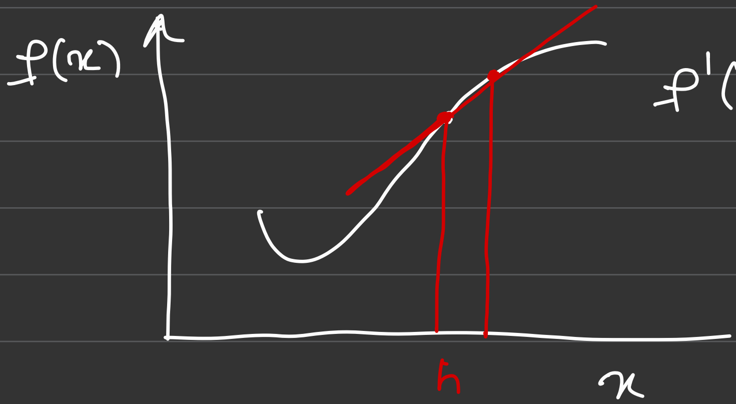
- $|\langle \underline{y}, \underline{x} \rangle| \le \|\underline{y}\|_* \|\underline{x}\| \quad \forall\ \underline{y} \in \mathbb{E}^*, \underline{x} \in \mathbb{E}$

Try these :

① Show that $\|\underline{y}\|_*$ is a valid norm

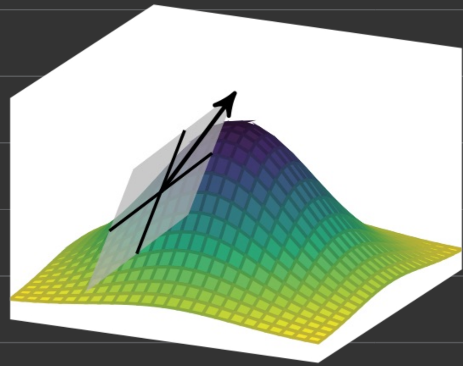
② Show that the dual norm of l_∞ is the l_1-norm and vice-versa.

Derivatives:



$$f'(x) = \frac{df(x)}{dx}$$

$$= \underset{h\to 0}{Lt} \quad \frac{f(x) - f(x-h)}{h}$$



$f : \mathbb{R}^n \to \mathbb{R}$ and $\underline{x} = [x_1, x_2, \ldots x_n]^T$

first- and second-order partials :

$$D_i f(\underline{x}) = \frac{\partial f(\underline{x})}{\partial x_i} \; ; \qquad D_{ij} f(\underline{x}) = \frac{\partial^2 f(\underline{x})}{\partial x_i \partial x_j}$$

$$= \frac{\partial}{\partial x_i} \left[\frac{\partial f(\underline{x})}{\partial x_j} \right]$$

Gradient:

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

<u>Suppose</u> $\underline{f} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$

$$\underline{f}(\underline{x}) = [f_1(\underline{x}), f_2(\underline{x}), \ldots f_m(\underline{x})]^T$$

Gradient matrix }

$$\nabla \underline{f}(\underline{x}) = [\nabla f_1(x), \nabla f_2(x), \ldots \nabla f_m(\underline{x})] : n \times m$$

$$[\nabla \underline{f}(\underline{x})]_{ij} = \frac{\partial f_i(\underline{x})}{\partial x_j}$$

Jacobian matrix }

$$J_{\underline{f}}(\underline{x}) = \nabla^T \underline{f}(\underline{x}) \quad \text{with } [J_{\underline{f}}(\underline{x})]_{ij} = \frac{\partial f_j(\underline{x})}{\partial x_i}$$

Hessian : "Jacobian of the gradient"

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

$n \times n$ matrix

$$\left[\nabla^2 f(\underline{x}) \right]_{ij} = D_{ij} f(\underline{x})$$

$$= \frac{\partial}{\partial x_j} \left[\frac{\partial}{\partial x_i} f(\underline{x}) \right]$$

Symmetric when 2nd order partials are continuous

Chain rule:

$$\underline{f}(\underline{x}) = \underline{g}(\underline{v}(\underline{x}))$$

$$J_f(\underline{x}) = J_{\underline{g}}(\underline{v}) \Big|_{\underline{v} = v(x)} J_{\underline{v}}(x)$$

Example:

$$f(\underline{x}) = \| \underline{y} - A x \|_2^2 \; . \quad \text{Set} \quad \underline{v}(\underline{x}) = \underline{y} - A\underline{x}$$

$$g(\underline{v}) = \| \underline{v} \|_2^2$$

$$\nabla g(\underline{v}) = 2\underline{v} = 2(y - A\underline{x})$$

$$J_{\underline{v}}(x) = -A$$

$$\nabla f(x) = -2A^T(\underline{y} - A\underline{x})$$

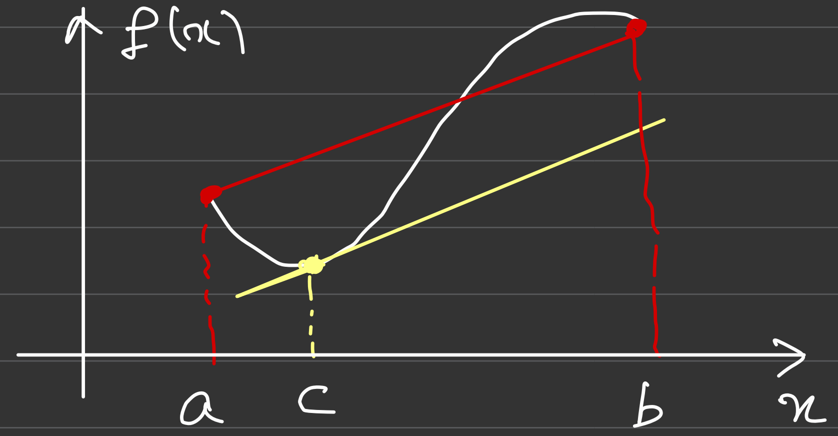
#15

Mean value theorem:

$f: \mathbb{R} \to \mathbb{R}$ is continuously differentiable in (a,b)

then $\exists$ $c \in [a,b]$ such that

$$\frac{f(b) - f(a)}{b - a} = f'(c)$$



For $f: \mathbb{R}^n \to \mathbb{R}$, then

for every $\underline{a}$ and $\underline{b}$ $\exists$ $\alpha \in [0,1]$ such that

$$f(\underline{b}) - f(\underline{a}) = \nabla f^T (\alpha \underline{a} + (1-\alpha)\underline{b})(\underline{b} - \underline{a})$$

Rate of convergence:

Let $\{\underline{x}_k\}$ converge to $\underline{x}^*$. We say that the convergence is of order $p\ (\geq 1)$ and with factor $r\ (>0)$, if $\exists\ k_0$ such that $\forall k \geq k_0$

$$\|\underline{x}_{k+1} - \underline{x}^*\| \leq r\, \|\underline{x}_k - \underline{x}^*\|^p$$

Some points:

- Larger power p, faster convergence
- For the same p, smaller r, faster convergence

- if $\{x_k\}$ converges with order p and factor r, if also converges with order $p' \leq p$ and $r' \geq r$
- So we seek for the largest p and smallest r

Linear convergence:

$p = 1$ and $r < 1$

$$\|\underline{x}_{k+1} - \underline{x}^*\| \le r\,\|\underline{x}_k - \underline{x}^*\|$$

$$\|\underline{x}_{k+2} - \underline{x}^*\| \le r\,\|\underline{x}_{k+1} - \underline{x}^*\|$$
$$\le r^2\,\|\underline{x}_k - \underline{x}^*\|$$
$$\vdots$$

For large enough k,

$$\|\underline{x}_{k+\tau} - \underline{x}^*\| \le r^\tau\,\|\underline{x}_k - \underline{x}^*\|$$

$$\log\|\underline{x}_{k+\tau} - \underline{x}^*\| \le \tau \log r + \log\|\underline{x}_k - \underline{x}^*\|$$

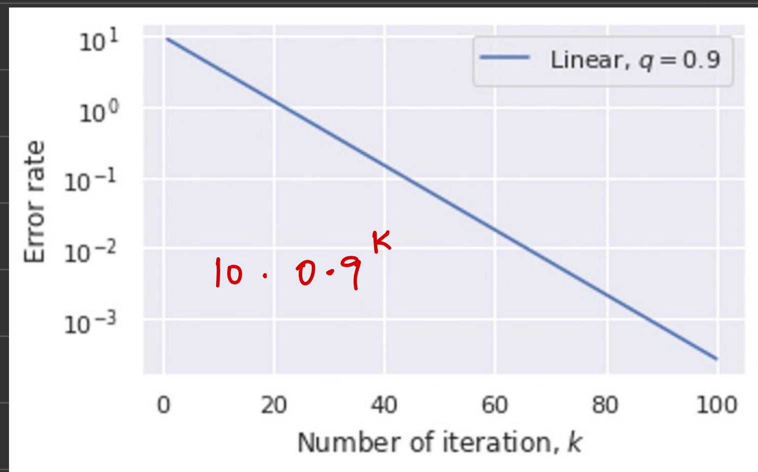
$-\log\|\underline{x}_{k+\tau} - \underline{x}^*\|$ grows "linearly" with τ

$-\log_{10} 10^{-1} = 1$, $-\log_{10} 10^{-2} = 2$, ... so on

Example:

$$\|\underline{x}_k - \underline{x}^*\| \approx q^k$$

$$0 < q < 1$$

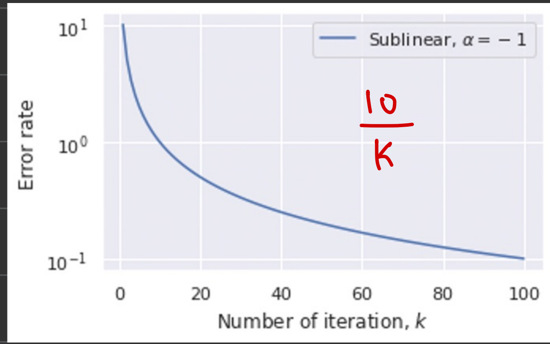


#17

Sublinear convergence :

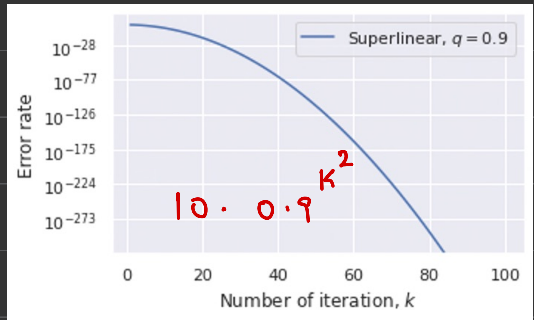
$p = 1$ and $r = 1$

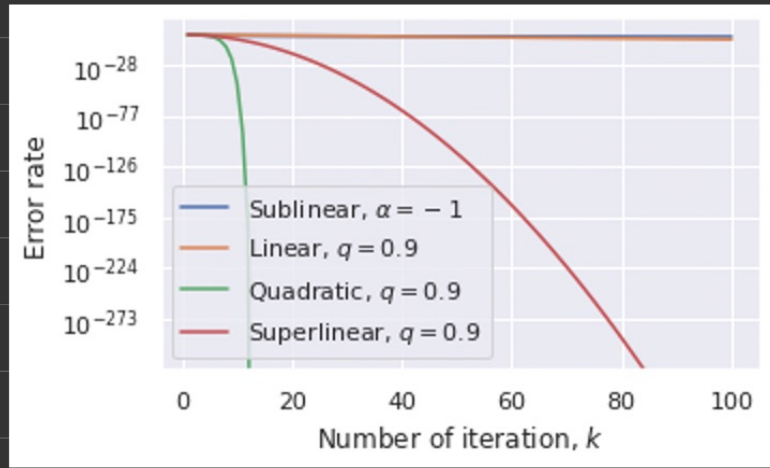
$$\| x_k - x^* \| \approx 1/k$$



Super linear :

$p > 1$



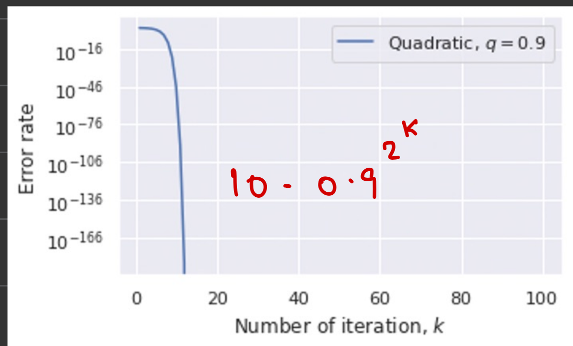


Quadratic :

$p = 2$

$$\| x_k - x^* \| \approx q^{2^k}$$

$0 < q < 1$



Number of iterations required

$f(k) = e^{-k/t}$

Time constant "t" for a function is the time required for the value of the function to decay by a factor ϵ as

$$f(k+\tau) = f(k)/\epsilon$$

- Linear rates:

$$q^k \approx \frac{1}{\epsilon} \Rightarrow k \log q = \log(1/\epsilon)$$

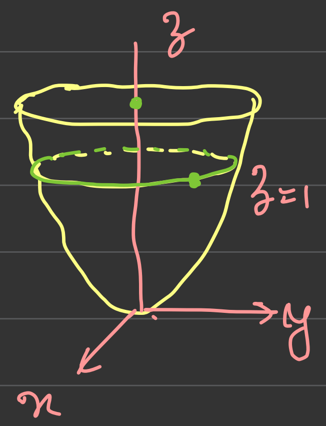
error being $O(q^k)$ needs $O(\log(1/\epsilon))$ iterations

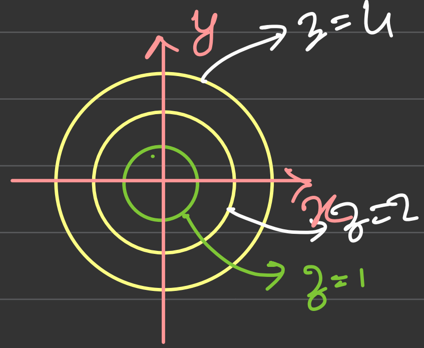
- Sublinear rates error being $O(1/k)$ need $O(1/\epsilon)$ iterations

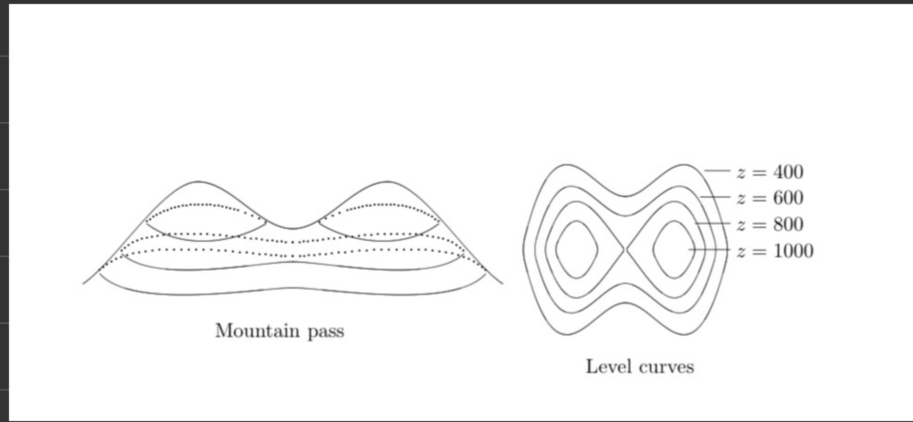
- Quadratic rates error being $O(q^{2^k})$ need $O(\log \log (1/\epsilon))$ iterations.

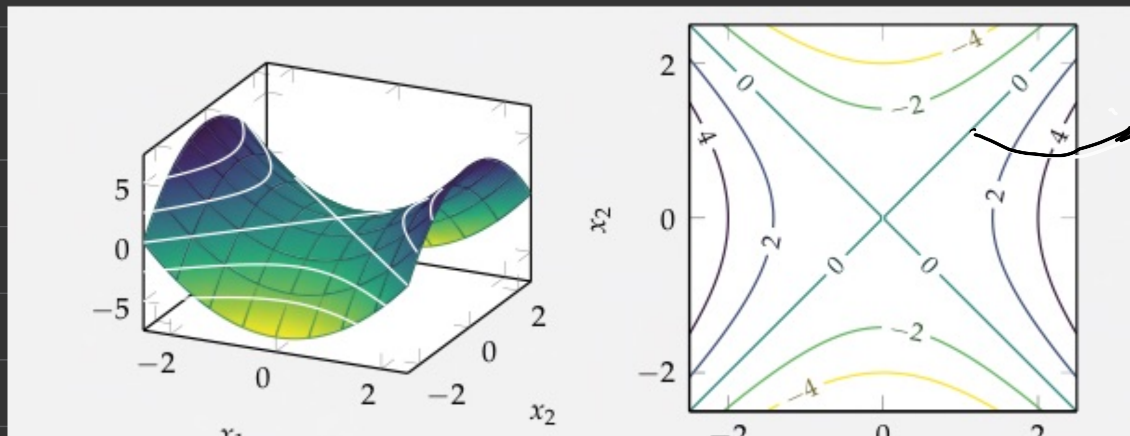
Level curves, contour plots and sublevel sets.

$$f(x, y) = x^2 + y^2 = z$$









$$f(x_1, x_2) = x_1^2 - x_2^2$$

- $x_1^2 - x_2^2 = 0$

  $x_1 = \pm x_2$ (line)

- $x_1^2 - x_2^2 = 1$

  (hyperbola)

α-Sublevel set: $C_\alpha = \{ x \in \operatorname{dom} f : f(x) \leq \alpha \}$