* **Recap** $\longrightarrow$ Convex functions

↳ Strongly convex functions

↳ L-smooth functions

↳ Recap:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$$

$$\underbrace{\mu I \preceq}_{\text{when } f \text{ is } \mu\text{-sc}} \nabla^2 f \preceq L I$$

when f is μ-sc $\Longrightarrow$ $L \geq \mu$

$$\boxed{\frac{\mu}{L} \leq 1}$$

* Slater's condition $\Rightarrow$ Strong duality

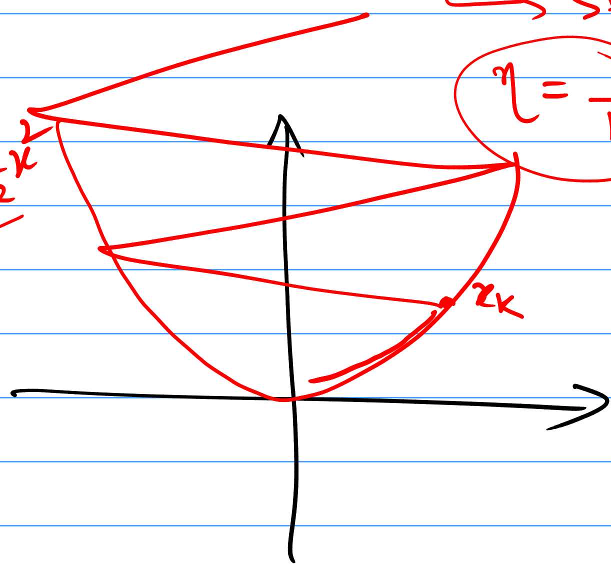
(Dual formulation of primal optimization problems)

GD: $x_{k+1} = x_k - \eta \nabla f(x_k)$ $\longleftarrow$ f is L-smooth

$\eta \longrightarrow$ Step-size?

$\eta = \frac{1}{L}$ $\longleftarrow$ A suitable choice of learning rate

$f(x) = \frac{1}{2}x^2$



$$x_{k+1} = x_k - \eta x_k$$

$$= (1-\eta)x_k$$

↳ $\eta < 1$ convergent

$\eta > 2$ diverges

— x — x — x — x —

* **Ex:** Dual of an LP (Linear Program)

$$\left.\begin{aligned} &\min_{x \in \mathbb{R}^n} \quad c^T x \\ &\text{s.t.} \quad Ax \leq b \end{aligned}\right\} \quad \begin{aligned} &A \in \mathbb{R}^{m \times n} \\ &m << n \end{aligned}$$

$$g(\lambda) := \min_{x \in \mathbb{R}^n} \left\{ f(x) + \lambda^T h(x) \right\}$$

$$= \min_{x \in \mathbb{R}^n} \left\{ c^T x + \lambda^T (Ax - b) \right\}$$

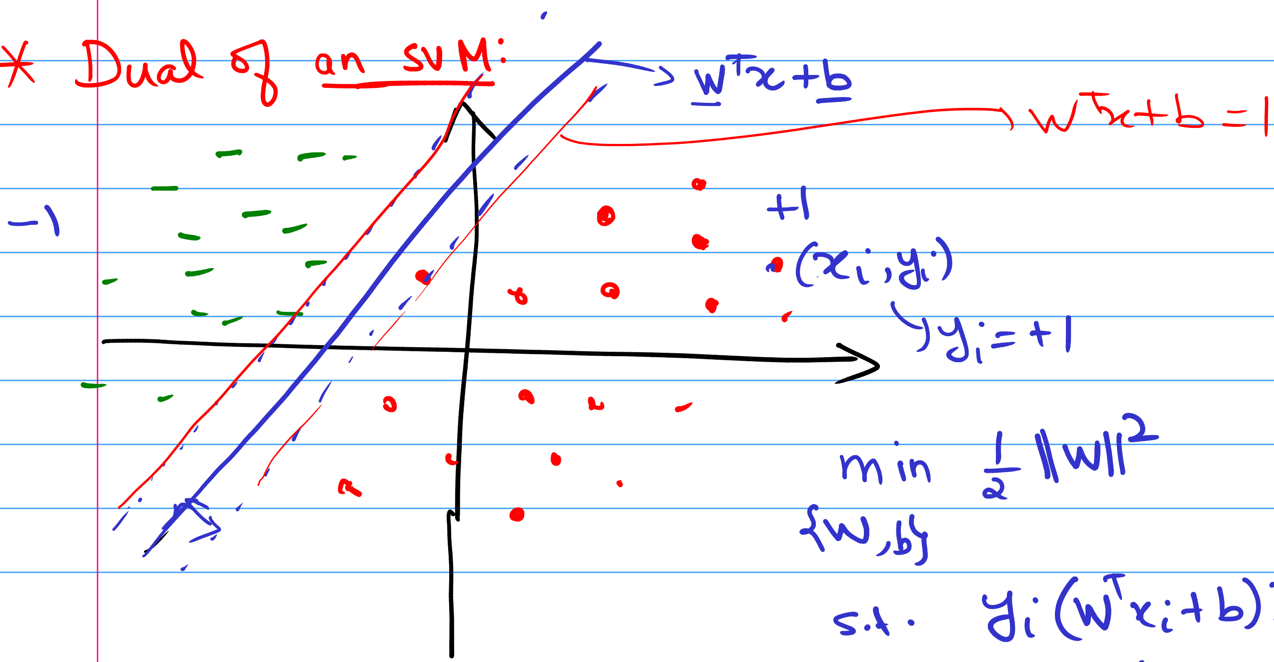
$$= \min_{x \in \mathbb{R}^n} \left\{ -\lambda^T b + \underline{(A^T \lambda + c)}^T x \right\}$$

$$g(\lambda) = \begin{cases} -\lambda^T b & \text{if } A^T \lambda + c = 0 \\ \underline{-\infty} & \text{else} \end{cases}$$

Suppose $\underline{p^*}$ is finite $\Longrightarrow$ Strong duality

$$\boxed{\begin{aligned} &\max_{\lambda \in \mathbb{R}^m} \quad -\lambda^T b \\ &\text{s.t.} \quad A^T \lambda + c = 0, \; \lambda \geq 0 \end{aligned}} \longrightarrow \text{Dual LP}$$

* Dual of <u>an SVM:</u>



$$\min_{\{w, b\}} \quad \frac{1}{2} \|w\|^2$$

$$\text{s.t.} \quad y_i (w^T x_i + b) \geq 1$$

$$\forall i \in \{1, 2, \dots, m\}$$

Dual of SVM:

$$g(\lambda) := \min_{\{w,b\}} \left(\frac{1}{2}\|w\|^2 + \sum_{i=1}^{m} \lambda_i \left(1 - y_i(w^T x_i + b)\right) \right)$$

$\nabla_w L(w,b,\lambda) = 0$
$$\boxed{w^* - \sum_{i=1}^{m} \lambda_i y_i x_i = 0}$$

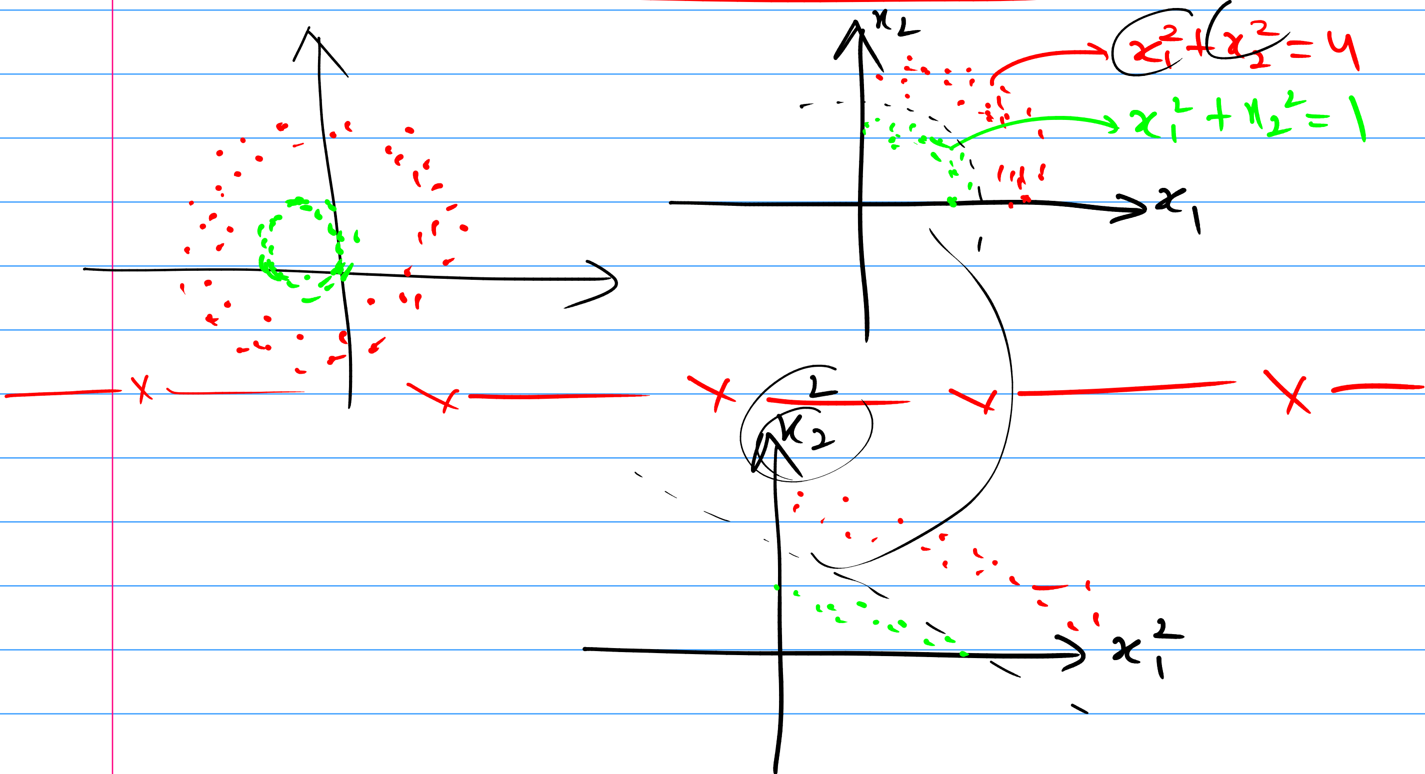
$\nabla_b L = 0$
$$\boxed{\sum_{i=1}^{m} \lambda_i y_i = 0}$$

$$g(\lambda) = \frac{1}{2}\left(\sum_{i=1}^{m} \lambda_i y_i x_i\right)^T \left(\sum_{j=1}^{m} \lambda_j y_j x_j\right) + \sum_{i=1}^{m} \lambda_i$$
$$- \sum_{i=1}^{m} \lambda_i y_i \left(\sum_{j=1}^{m} \lambda_j y_j x_j\right)^T x_i - b \underbrace{\cancel{\sum_{i=1}^{m} \lambda_i y_i}}_{\to 0}$$

$K(x_i, x_j)$

$$\boxed{g(\lambda) = \sum_{i=1}^{m} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{m} \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle}$$

$$\boxed{\begin{aligned} &\max_{\lambda \in \mathbb{R}^m} \; g(\lambda) \\ &\text{s.t.} \; \sum_{i=1}^{m} \lambda_i y_i = 0, \; \lambda_i \geq 0 \end{aligned}} \to \text{Dual of SVM}$$



* KKT Conditions:

(Karush-Kuhn-Tucker conditions:

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t.} \quad h_i(x) \le 0 \quad \forall i \in \{1,2,\dots,m\}$$
$$l_j(x) = 0 \quad \forall j \in \{1,2,\dots,r\}$$

$$L(x,\lambda,\nu) = f(x) + \underbrace{\sum_{i=1}^{m} \underset{\ge 0}{\lambda_i} \underset{\le 0}{h_i(x)}}_{\le 0} + \sum_{j=1}^{r} \nu_j l_j(x)$$

* Stationarity: $0 \in \nabla_x L(x,\lambda,\nu)$
* Complementary slackness: $\lambda_i h_i(x) = 0 \quad \forall i$
* Primal feasibility: $h_i(x) \le 0$, $l_j(x) = 0 \quad \forall i,j$
* Dual feasibility: $\lambda_i \ge 0 \quad \forall i$

Thm: For an optimization problem with strong duality, x^* is a primal optimal sol^n, and (λ^*, ν^*) is a dual optimal sol^n

$$\Updownarrow$$

x^*, λ^*, ν^* satisfy KKT conditions.

Check optimality
Suboptimality

Ex: QP with equality constraint.

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T Q x + c^T x \qquad Q \succ 0$$
$$\text{s.t.} \quad Ax = b$$

$$L(x,\nu) = \frac{1}{2} x^T Q x + c^T x + \nu^T (Ax - b)$$

Stationarity: $\nabla_x L(x,\nu) = 0 \Big|_{x^*,\nu^*}$

$$Qx^* + c + A^T \nu^* = 0$$
$$Ax^* - b = 0$$

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

Ex: Dual of SVM:

Inequality constraints:

$$y_i(w^T x_i + b) \geq 1$$

or $1 - y_i(w^T x_i + b) \leq 0 \quad \forall i \in \{1, \ldots, m\}$

→ complementary slackness

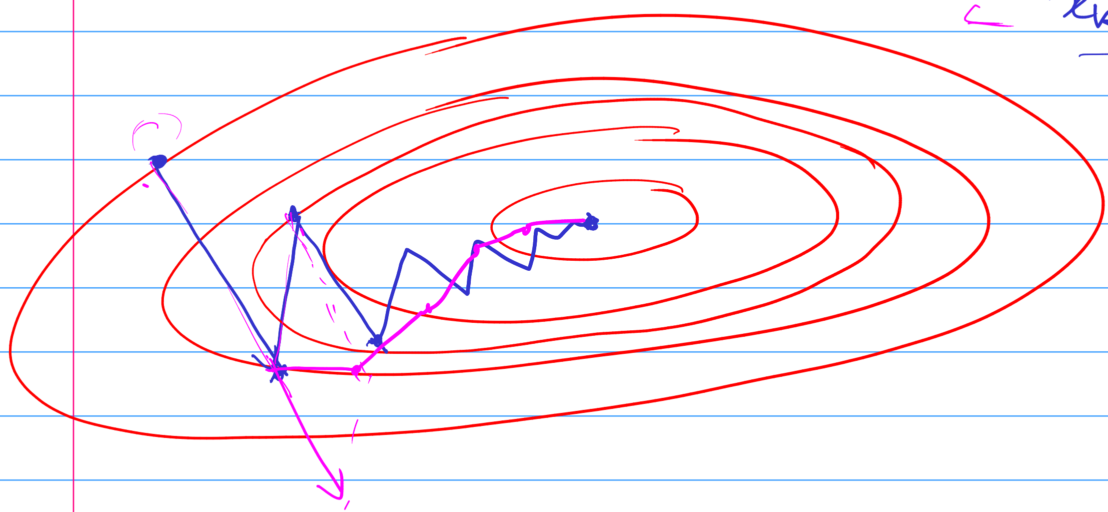
$$\lambda_i^* \left(1 - y_i(w^{*T} x_i + b^*)\right) = 0$$

$\lambda_i = 0$

$\lambda_i \neq 0$ → support vectors

$$1 - y_i(w^{*T} x_i + b^*) = 0$$

$\leftarrow \quad x_{k+1} = x_k - \eta \nabla f(x_k)$



Gradient Descent: f is convex. f is L-smooth. } Assumptions

$$x_{k+1} = x_k - \eta \nabla f(x_k)$$

Current iterate is x_k.

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

[Taylor's expansion]

z lies on the line joining x_k and y.

$$\|\nabla^2(z)\| \le L$$

$$f(y) \le f(x_k) + \nabla f(x_k)^T(y - x_k) + \frac{L}{2}\|y - x_k\|^2$$

$$= f(x_k) + \frac{L}{2}\left\{\|y - x_k\|^2 + \frac{2}{L}\nabla f(x_k)^T(y - x_k) + \frac{1}{L^2}\|\nabla f(x_k)\|^2 - \frac{1}{L^2}\|\nabla f(x_k)\|^2\right\}$$

$$f(y) \le f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 + \frac{L}{2}\underbrace{\|y - x_k + \frac{1}{L}\nabla f(x_k)\|^2}$$

$$x_{k+1} \leftarrow y = x_k - \frac{1}{L}\nabla f(x_k)$$

$$f(x_{k+1}) \le f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2$$

$$\boxed{x_{k+1} \leftarrow x_k - \frac{1}{L}\nabla f(x_k)} \longrightarrow \eta = \frac{1}{L}$$

<u>Thm:</u> Suppose $f(x)$ is convex and L-smooth, then for GD with step-size $1/L$, we have

$$f(x_{k+1}) - f(x^*) \le \frac{L}{2}\frac{\|x_0 - x^*\|^2}{(k+1)} = O\left(\frac{1}{k}\right)$$

$\le \varepsilon$

$O\left(\frac{1}{\varepsilon}\right)$

ε

$\frac{1}{k} = \varepsilon$

$k = \frac{1}{\varepsilon}$

$$g(x) = O(h(x))$$

$$|g(x)| \le M\, h(x) \quad \forall\, x \ge x_0$$

$$h(n) = n^2 - 2n + 3n^3$$

$$|h(n)| \le n^3 + 2n^3 + 3n^3 = 6n^3 \qquad h(n) = O(n^3) \ \forall n \ge 1$$

* Rate of convergence and order of convergence.

$$\lim_{n\to\infty} \frac{|x_{n+1} - x^*|}{|x_n - x^*|^q} = \rho \longrightarrow \text{Rate of convergence}$$

$q \ge 1$, → order of convergence