1. Intro. to Convex Optimization.

For a optimization problem to be convex:

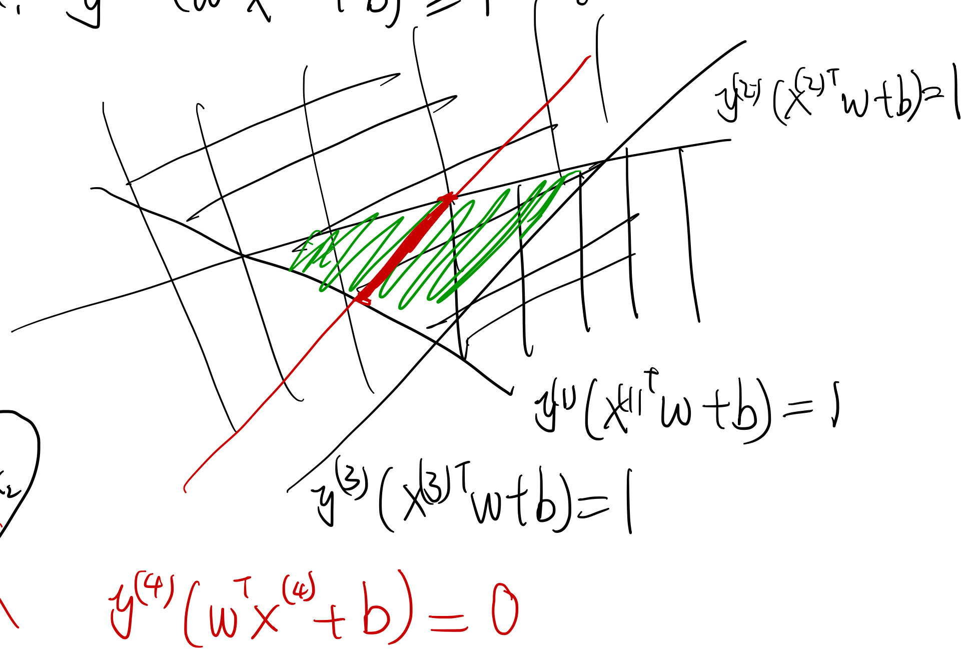
① feasible region has to be a convex set

② objective function is a convex function

Feasible Region

E.g. SVM $\min_{w} \frac{1}{2} \|w\|^2$

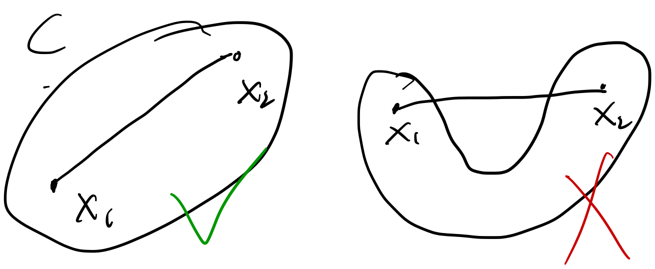
s.t. $y^{(i)} (w^T x^{(i)} + b) \geq 1$, $i = 1, \ldots, m$



Convex Set. C

$x_1, x_2 \in C$ $0 \leq \theta \leq 1$

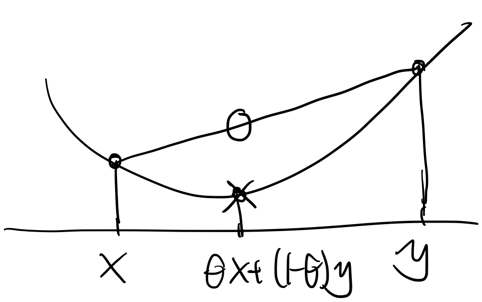
$\Rightarrow \theta x_1 + (1-\theta) x_2 \in C$



Convex Function

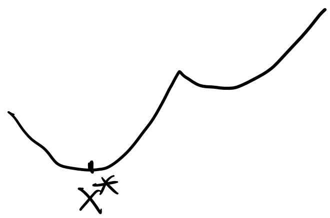
① Domain has ~~to~~ to be convexset.

② $f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$



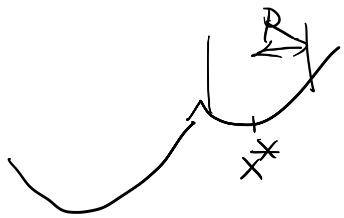
Globally and Locally Optimal

x^* is the globally optimal solution $f(x^*) \le f(x) \quad \forall x \in \text{dom} f$



x^* is the locally optimal solution $f(x^*) \le f(x)$

$\forall x \in \{x \mid \|x - x^*\| \le R\}$



Standard form of convex optimization

Primal Problem

$$\min_x f_0(x) = \min_x \max_{\lambda, v} L$$

$$\text{s.t. } f_i(x) \le 0 \quad i = 1, \dots m \leftarrow \lambda_i$$

$$h_i(x) = 0 \quad i = 1, \dots, p \leftarrow v_i$$

If feasible $\lambda \ge 0$

$$\underline{L(x, \lambda, v) = f_0(x) + \boxed{\sum_{i=1}^{m} \lambda_i f_i(x)}^{\le 0} + \sum_{i=1}^{p} v_i h_i(x) \le f_0(x)}$$

Lagrange dual function

$$g(\lambda, v) = \inf_{x \in D} L(x, \lambda, v)$$

$\rightarrow \min$

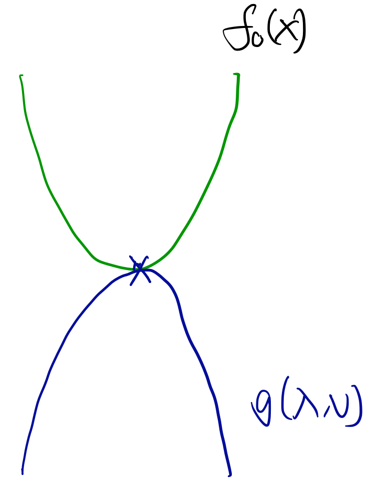
x is feasible $\lambda \ge 0$

$$g(\lambda, v) \le L(x, \lambda, v) \le f_0(x)$$ Weak Duality

Dual Problem

$$\max_{\lambda, \nu} \; g(\lambda, \nu) = \max_{\lambda, \nu} \min_{x} L$$

s.t. $\lambda \geq 0$



Strong Duality:

$$g(\lambda^*, \nu^*) = L(x^*, \lambda^*, \nu^*) = f_0(x^*)$$

KKT Conditions:

1. Primal constraints. $f_i(x) \leq 0 \quad h_i(x) = 0$
2. Dual constraints $\quad \lambda \geq 0$
3. Complementary slackness
$$\lambda_i f_i(x) = 0 \; , \quad i = 1, \ldots, m$$
4. Gradient of $L(x, \lambda, \nu)$ w.r.t x vanishes
$$\nabla_x L(x, \lambda, \nu) = 0$$

2. Find a dual problem

$$\min \; x^T P x$$

$$\text{s.t.} \quad Ax \le b \qquad \lambda$$

$$L(x,\lambda) = x^T P x + \lambda^T (Ax - b)$$

$$g(\lambda) = \overset{\min}{\inf_x} L(x,\lambda)$$

$$\nabla_x L(x,\lambda) = 2Px + A^T\lambda = 0 \implies \underline{x = -\frac{1}{2} P^{-1} A^T \lambda}$$

$$g(\lambda) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

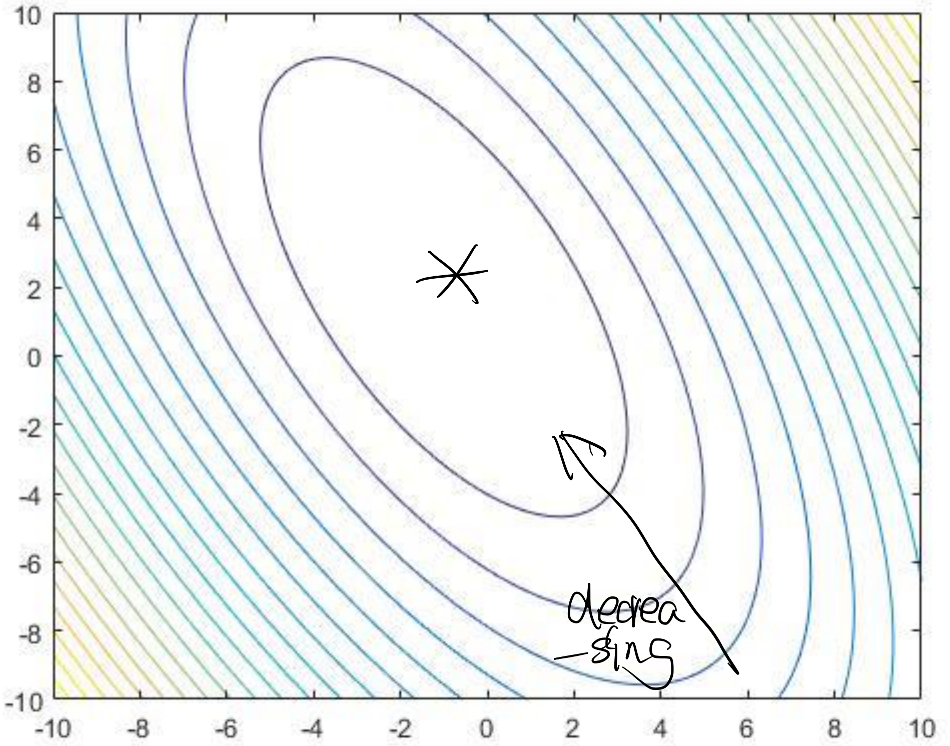
Dual Problem

$$\max_{\lambda} \; g(\lambda)$$

$$\text{s.t.} \quad \lambda \ge 0$$

Unconstrained QP

We want to minimize: $$J(x_1, x_2) = 5x_1^2 + 4x_1x_2 + 2x_2^2 + 2x_1 - 4x_2.$$

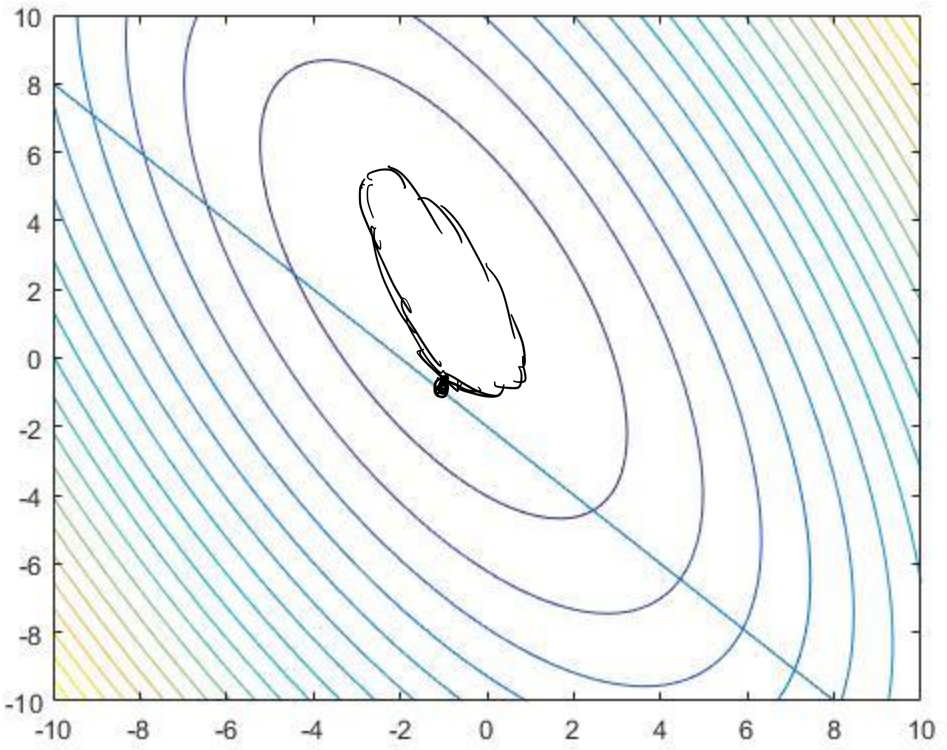



```
A = [5 2;2 2];
b = [2 -4];
cvx_begin
    variable x(2)
    minimize quad_form(x,A)+b*x
cvx_end
```

$x = [-1; 2].$

Linear equality constraint

We add a linear equality constraint: $x_1 + x_2 + 2 = 0$. $\surd$




```
A = [5 2;2 2];
b = [2 -4];
cvx_begin
    variable x(2)
    minimize quad_form(x,A)+b*x
    subject to
        x1+x2+2 == 0;
cvx_end
```

$x = [-1; -1]$.

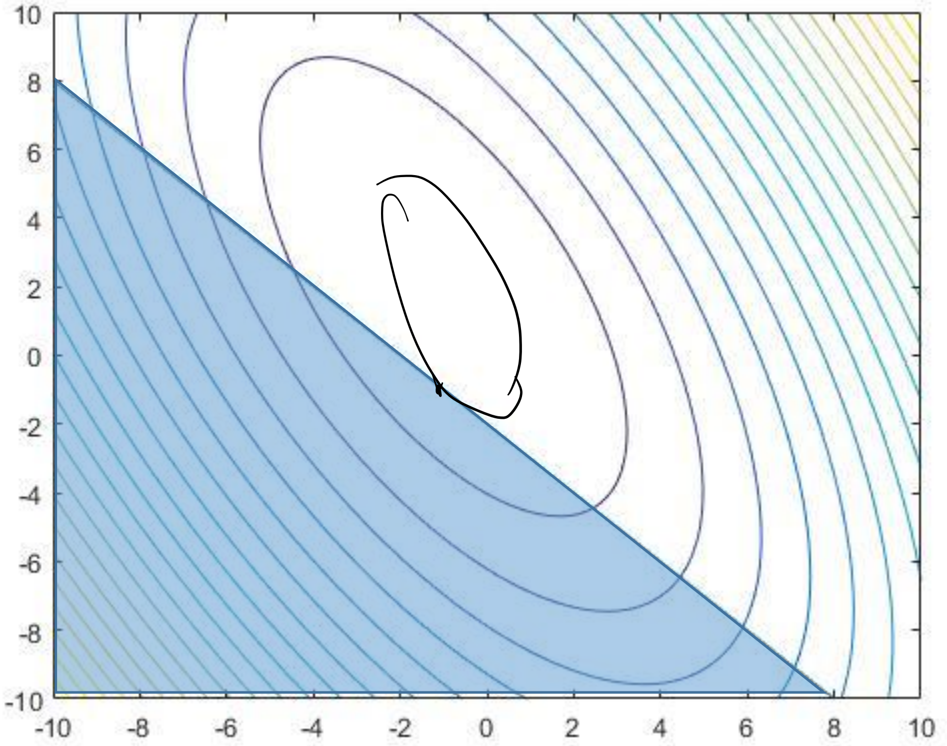
$$L(x,\lambda) = 5x_1^2 + 4x_1x_2 + 2x_2^2 + 2x_1 - 4x_2 + \lambda(x_1 + x_2 + 2)$$

$$\nabla_x L(x,\lambda) \Rightarrow \begin{cases} x_1 + 4x_2 + 2 + \lambda = 0 \\ 4x_1 + 4x_2 - 4 + \lambda = 0 \end{cases}$$

$$\lambda(x_1 + x_2 + 2) = 0$$

Linear inequality constraint

We add a linear inequality constraint: $x_1 + x_2 + 2 \leq 0$.

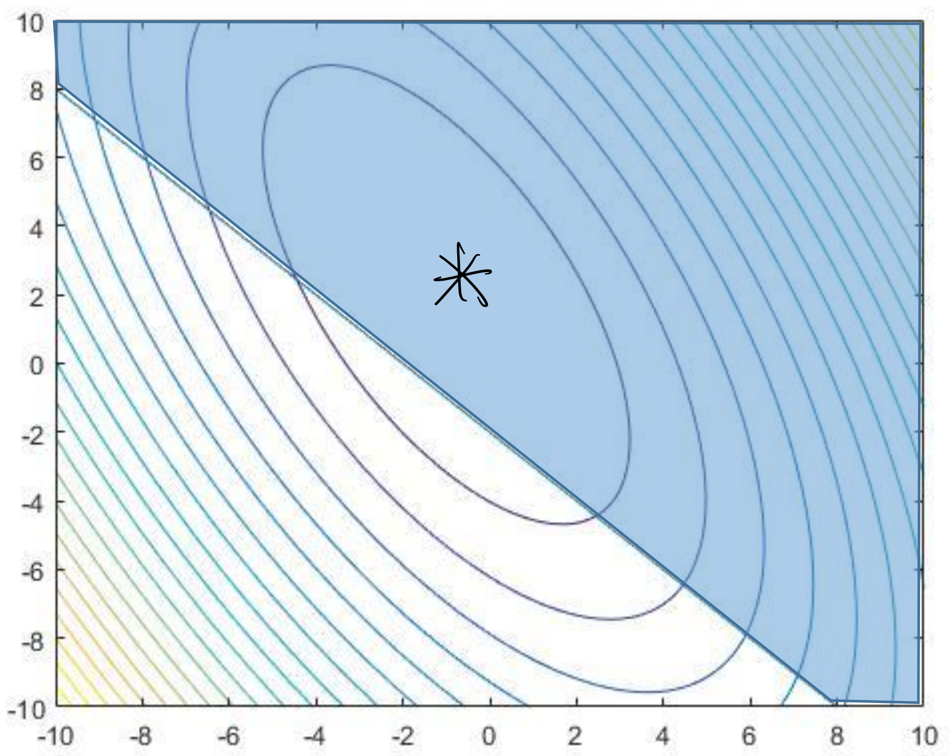



```
A = [5 2;2 2];
b = [2 -4];
cvx_begin
    variable x(2)
    minimize quad_form(x,A)+b*x
    subject to
        x1+x2+2 <= 0;
cvx_end
```

$x = [-1; -1]$.

Linear inequality constraint

We add a linear inequality constraint: $x_1 + x_2 + 2 \geq 0$.




```
A = [5 2;2 2];
b = [2 -4];
cvx_begin
    variable x(2)
    minimize quad_form(x,A)+b*x
    subject to
        x1+x2+2 >= 0;
cvx_end
```

$x = [-1; 2]$.

Kernalized Linear Regression

When can kernal trick be used:

① Model parameter only depends on inner products of features.

② At test time, the resule also only depends on inner product.

$$w = (\underline{X^T X} + \lambda I)^{-1} \underline{X^T y} \qquad \times \text{ Kernal trick}$$

Rewriting

$$(X^T X + \lambda I)\, w = X^T y$$

$$w = \frac{1}{\lambda}(X^T y - X^T X w) = X^T \cdot \underbrace{\frac{1}{\lambda}(y - Xw)}_{a}$$

$$w = X^T a$$

$$\lambda a = Y - X \cdot X^T a$$

$$a = (X \cdot X^T + \lambda I)^{-1} y$$

$$X \cdot X^T \rightarrow \begin{bmatrix} x_1^T x_1 & \cdots & \\ x_1^T x_2 & \ddots & \\ \vdots & & \ddots \end{bmatrix} \rightarrow \begin{bmatrix} k(x_1, x_1) & \cdots & \\ \vdots & \ddots & \\ \vdots & & \ddots \end{bmatrix}$$

At test time:

$$w^T z = X^T a \cdot z = a^T X^T z$$

$$= \sum_{i=1}^{N} a_i \underline{x_i^T z}$$

$$x_i^T z \rightarrow k(x_i, z)$$

Kernalized Perceptron

$$w^{(l+1)} = w^{(l)} + x_n y_n \quad \text{when misclassified}$$

$$w = a_1 x_1 y_1 + a_2 x_2 y_2 + a_3 x_3 y_3 \; \cdots$$

$$\{x_1, y_1\} \; \cdots \; \{x_2 \; y_2\} \; \cdots$$

$$a_1 \qquad\qquad a_2$$

$$w = \sum_{i=1}^{N} a_i x_i y_i$$

At test time:

$$\text{sign}(w^T z) = \text{sign} \sum_{i=1}^{N} a_i \underbrace{x_i^T z}_{K(x_i, z)} y_i$$