Lecture #20 **Duality and KKT conditions** EE 260

- lower bound on f^*
- Duality : Lagrangian, dual function, dual problem
- Slater's conditions, strong duality
- KKT conditions.

$$\underset{x}{\text{minimize}} \quad f(x)$$

$$\text{s. to} \quad \underline{x} \in C$$

Suppose we want to find a lower bound

$$B \leq f(\underline{x}^*) \implies f^* \geq B$$

how can we do that?

Example:

1. $\underset{x,\, y}{\text{minimize}} \quad x + y$

s.to. $x + y \geq 2$

$x, y \geq 0$

$\Rightarrow \quad B = 2$

2. $\underset{x,\, y}{\text{minimize}} \quad x + 3y$

s. to $x + y \geq 2$

$x, y \geq 0$

$$\begin{array}{rr} & x + y \geq 2 \\ + & 2y \geq 0 \\ \hline & x + 3y \geq 2 \end{array}$$

$\Rightarrow \quad B = 2$

$$\begin{array}{ll} \underset{x,\,y}{\text{minimize}} & px + qy \\ & x + y \geq 2 \\ & x, y \geq 0 \end{array}$$

Lower bound :

$$\begin{array}{l} a + b = p \\ a + c = q \\ a, b, c \geq 0 \end{array}$$

$$\underbrace{a(x+y)}_{\geq 2} + \underbrace{bx}_{\geq 0} + \underbrace{cy}_{\geq 0}$$

$$B = 2a$$

So the best lower bound is obtained by

$$\boxed{\begin{array}{ll} \underset{a,\,b,\,c}{\text{maximize}} & 2a \\ & a + b = p \\ & a + c = q \\ & a, b, c \geq 0 \end{array}}$$

This is called the dual problem (dual LP)

- Number of variables in the dual problem = no. of constraints in the primal problem.

Lagrangian :

Standard form problem (not necessarily convex)

$$\underset{\underline{x}}{\text{minimize}} \quad f(\underline{x})$$

$$\text{subject to} \quad h_i(\underline{x}) \leq 0 \quad i = 1, \ldots, m$$

$$l_j(\underline{x}) = 0 \quad j = 1, \ldots, r$$

- Call $f(\underline{x}^*) = p^*$

Define Lagrangian as

$$L(\underline{x}, \underline{u}, \underline{v}) = f(\underline{x}) + \sum_{i=1}^{m} u_i \underbrace{h_i(\underline{x})}_{\leq 0} + \sum_{j=1}^{r} v_j \underbrace{l_j(\underline{x})}_{= 0}$$

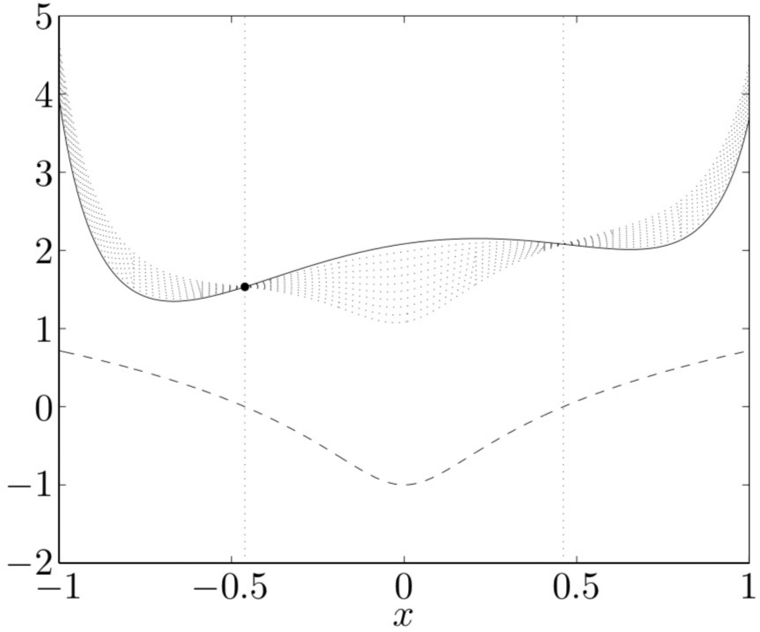
at feasible x

- New variables : $\underline{u} \in \mathbb{R}^m$ and $\underline{v} \in \mathbb{R}^r$

$\underline{u} \geq 0$ (else $L(\underline{x}, \underline{u}, v) = -\infty$)

• We have, for each feasible x:

$$f(x) \geq L(\underline{x}, \underline{u}, v)$$



- Solid line is f
- Dashed line is h, hence feasible set $\approx [-0.46, 0.46]$
- Each dotted line shows $L(x, u, v)$ for different choices of $u \geq 0$

(From B & V page 217)

Lagrange dual function:

Minimizing $L(\underline{x}, \underline{u}, v)$ over all $\underline{x}$ gives a lower bound on the primal optimal value f^*

$$f^* \geq \underset{\underline{x} \in C}{\text{minimize}} \; L(\underline{x}, \underline{u}, \underline{v}) \geq \underbrace{\underset{\underline{x}}{\text{minimize}} \; L(\underline{x}, \underline{u}, v)}_{= g(\underline{u}, \underline{v})}$$

Lagrange dual function

$$g(\underline{u}, \underline{v}) = \inf_{\underline{x} \in \operatorname{dom}(f)} L(\underline{x}, \underline{u}, \underline{v})$$

$$= \min_{\underline{x} \in \operatorname{dom}(f)} \left\{ \left(f(\underline{x}) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{i=1}^{r} v_i l_i(x) \right) \right\}$$

- $g(\underline{u}, v)$ is concave (even when the primal is not convex)

$$g(\underline{u}, \underline{v}) = -\max_{x} \underbrace{\left\{ -f(\underline{x}) - \sum_{i=1}^{m} u_i h_i(\underline{x}) - \sum_{i=1}^{r} v_i l_i(\underline{x}) \right\}}_{\text{pointwise max. of convex functions in } (\underline{u}, \underline{v})}$$

- Lower bound property: $\underline{u} \geq 0$, then $g(\underline{u}, \underline{v}) \leq p^*$

if $\tilde{x}$ is feasible and $\underline{u} \geq 0$, then

$$g(\underline{u}, \underline{v}) \leq f(\tilde{x})$$

$$f(\tilde{\underline{x}}) \geq L(\tilde{x}, \underline{u}, \underline{v}) \geq \inf_{\underline{x} \in \operatorname{dom} f} L(\underline{x}, \underline{u}, v) = g(\underline{u}, \underline{v})$$

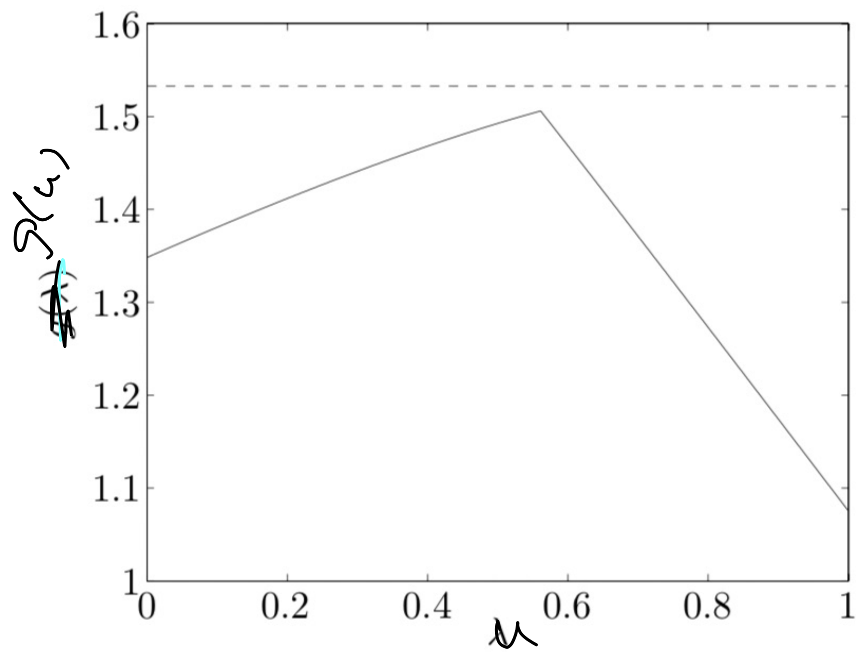
minimize over all feasible $\tilde{x}$ gives $p^* \geq g(\underline{u}, \underline{v})$.

u and v are dual feasible

dual variables.

- Dashed horizontal line is $f^\star$
- Dual variable λ is (our u)
- Solid line shows $g(\lambda)$

(From B & V page 217)



Lagrange dual problem:

$$p^* \geq g(\underline{u}, \underline{v}) \quad \text{for any } \underline{u} \geq 0 \text{ and } \underline{v}$$

Hence the best lower bound is obtained by solving the Lagrange dual problem

Convex optimization problem

$$\begin{cases} \underset{\underline{u},\, \underline{v}}{\text{maximize}} & g(\underline{u}, \underline{v}) \\ \text{s. to} & u \geq 0 \end{cases}$$

$g: \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}$

Weak duality: $p^* \geq g^*$

g^* is dual optimal value

Example: Quadratic program

minimize $\quad \frac{1}{2} \underline{x}^T Q \underline{x} + \underline{c}^T \underline{x}$

s.to. $\quad A\underline{x} = b \ , \quad \underline{x} \geq 0$

with $Q \succ 0$

- $L(\underline{x}, \underline{u}, \underline{v}) = \frac{1}{2} \underline{x}^T Q \underline{x} + \underline{c}^T \underline{x} - \underline{u}^T \underline{x} + \underline{v}^T (A\underline{x} - b)$

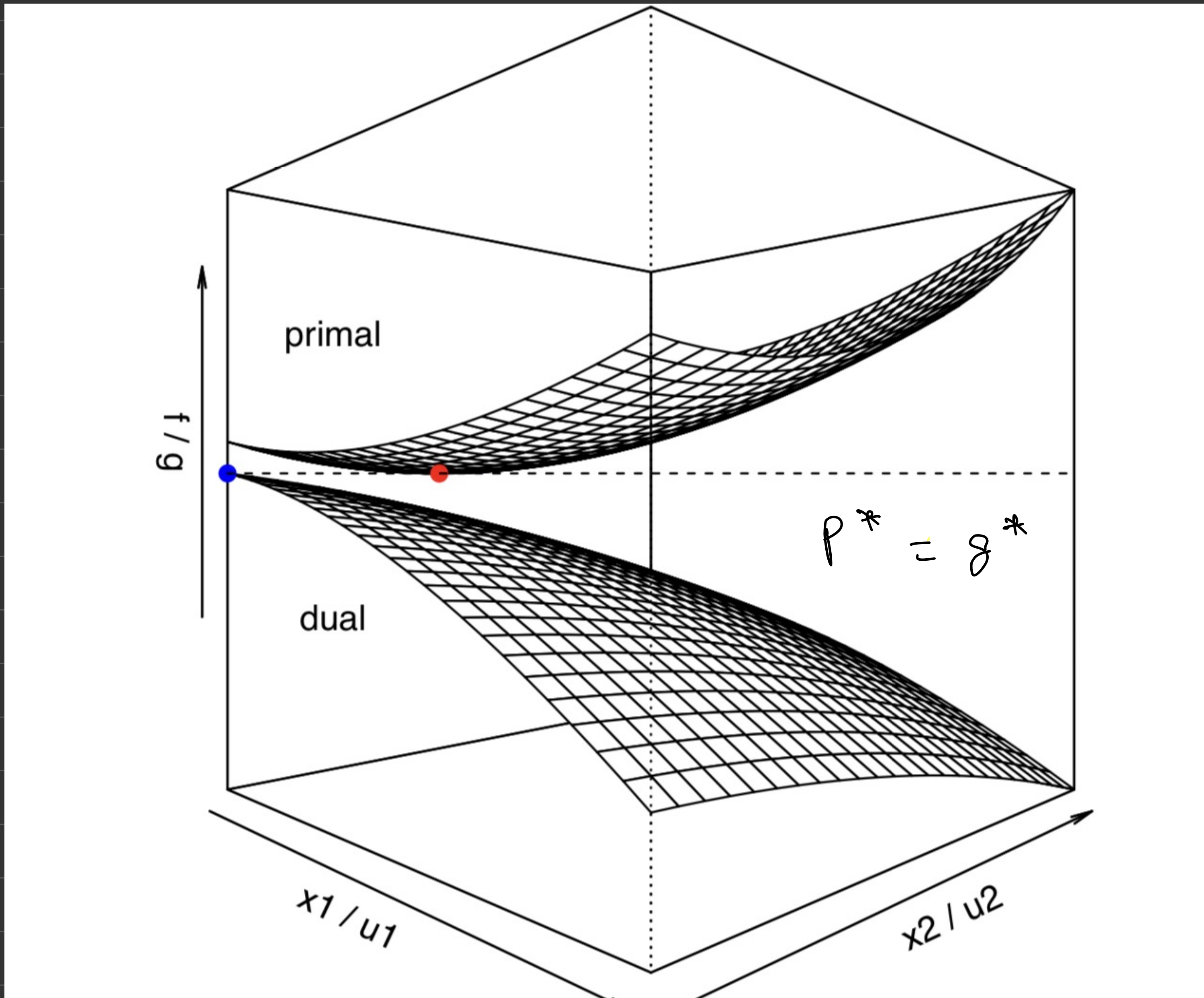
- $g(\underline{u}, \underline{v}) = \inf_{x} L(\underline{x}, \underline{u}, \underline{v})$

$= -\frac{1}{2} (\underline{c} - \underline{u} + A^T \underline{v})^T Q^{-1} (\underline{c} - \underline{u} + A^T \underline{v}) - b^T \underline{v}$

$Q\underline{x} + \underline{c} - \underline{u} + A^T \underline{v} = 0$

$\underline{x} = -Q^{-1} (\underline{c} - \underline{u} + A^T \underline{v})$

Quadratic in 2 variables



Strong duality:

$$p^* = g^*$$

Slater's conditions: if primal is a convex problem with at least one strictly feasible $\underline{x} \in \mathbb{R}^n$

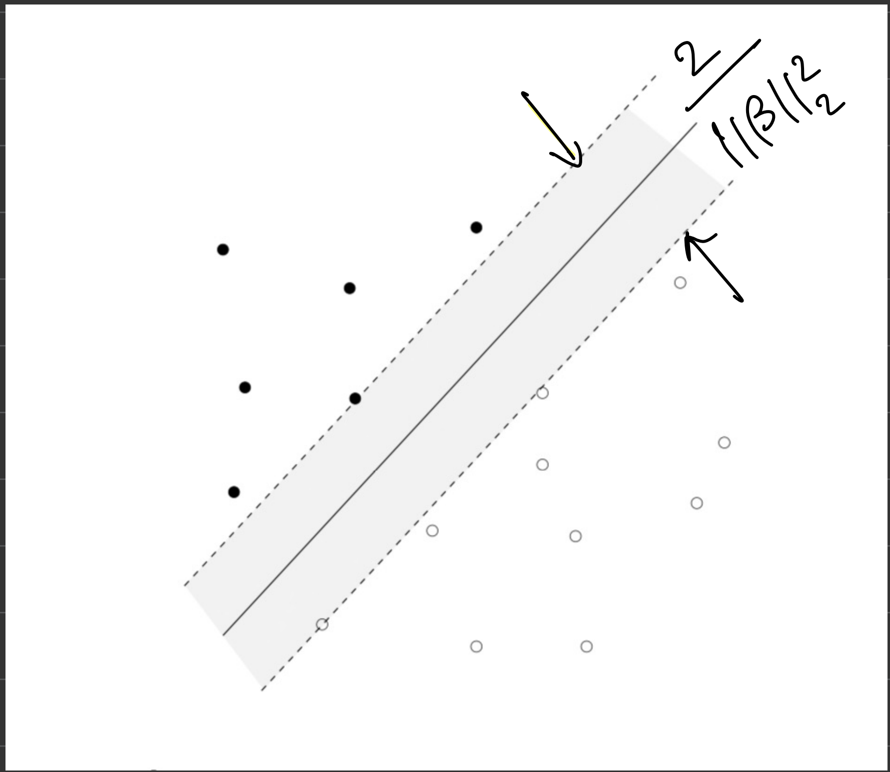
i.e.,

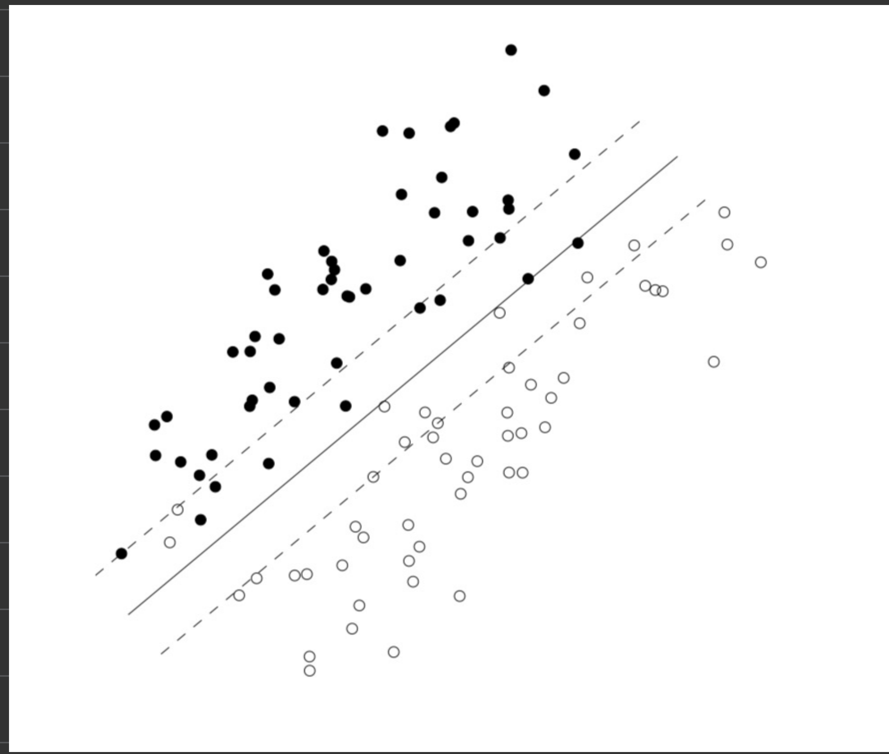
$$h_1(\underline{x}) < 0, \ldots, h_m(x) < 0$$

$$l_1(\underline{x}) = 0, \ldots, l_r(x) = 0$$

then strong duality holds.

Example : Support vector machine Classifier





Given $y \in \{1, -1\}^n$, $X : n \times p$ with rows $\{\underline{x}_i\}$

$$\underset{\beta, \beta_0, \underline{\xi}}{\text{minimize}} \quad \frac{1}{2} \|\beta\|_2^2 + C \sum_{i=1}^{n} \xi_i$$

s.t. $\xi_i \geq 0$

$y_i (x_i^T \beta + \beta_0) \geq 1 - \xi_i \quad i = 1, \ldots, n$

Lagrangian:

$$L(\underline{\beta}, \beta_0, \underline{\xi}, \underline{v}, \underline{w}) = \frac{1}{2} \|\underline{\beta}\|^2 + C \sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} v_i \xi_i$$

$$+ \sum_{i=1}^{n} w_i \left(1 - \xi_i - y_i(\underline{x}_i^T \underline{\beta} + \beta_0)\right)$$

Dual function:

$$g(\underline{v}, \underline{w}) = \begin{cases} -\frac{1}{2} \underline{w}^T \tilde{X} \tilde{X}^T \underline{w} + \underline{1}^T \underline{w} & \text{if } \underline{w} = C\underline{1} - \underline{v} \\ & \underline{w}^T \underline{y} = 0 \\ -\infty & \text{o.w.} \end{cases}$$

$$\tilde{X} = \operatorname{diag}(\underline{y})\, X$$

SVM dual: (eliminating slack variable $\underline{v}$):

$$\underset{\underline{\omega}}{\text{maximize}} \quad -\frac{1}{2}\,\underline{\omega}^T \tilde{X}\tilde{X}^T \underline{\omega} + \underline{1}^T \underline{\omega}$$

$$0 \le \underline{\omega} \le c\,\underline{1}, \quad \underline{\omega}^T \underline{y} = 0$$

Slater's condition is satisfied, we have strong duality

$$\underline{\beta} = \tilde{X}^T \underline{\omega}$$

Duality gap:

Given primal feasible $\underline{x}$ and dual feasible $\underline{u}, \underline{v}$:

duality gap: $f(\underline{x}) - g(\underline{u}, v)$

We have

$$f(x) - f(\underline{x}^*) \leq f(\underline{x}) - g(\underline{u}, \underline{v})$$

if $f(x) - g(x, v) = 0$, then x is primal optimal (and $\underline{u}$ & $\underline{v}$ are dual optimal)

Karush – Kuhn – Tucker conditions:

(KKT)

$$\underset{\underline{x}}{\text{minimize}} \quad f(\underline{x})$$

$$\text{s. to} \quad h_i(\underline{x}) \leq 0 \quad , \quad i = 1, \ldots m$$

$$l_j(\underline{x}) = 0 \qquad j = 1, \ldots r$$

KKT conditions:

① Stationarity: $\underline{0} \in \partial_x \left(f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j l_j(x) \right)$

② Complementary slackness: $u_i \, h_i(x) = 0 \qquad i = 1, \ldots, m$

③ primal feasibility: $h_i(x) \leq 0 \; ; \; l_j(x) = 0 \quad \forall i, j$

4. Dual feasibility: $u_i \geq 0 \quad \forall i$

For a problem with strong duality (i.e., Slater's condition holds):

x^* and $\underline{u}^*, \underline{v}^*$ are primal and dual solutions

$\iff$ $\underline{x}^*$ and $\underline{u}^*, \underline{v}^*$ satisfy the KKT conditions.

(1) if $\underline{x}^*$ & $\underline{u}^*, \underline{v}^*$ are primal and dual solutions with zero duality gap:

$$f(\underline{x}^*) = g(\underline{u}^*, \underline{v}^*)$$

$$= \min_{\underline{x}} \left\{ f(\underline{x}) + \sum_{i=1}^{m} u_i^* h_i(x) + \sum_{j=1}^{r} v_j^* l_j(\underline{x}) \right\}$$

$$\leq f(\underline{x}^*) + \sum_{i=1}^{m} u_i^* h_i(\underline{x}^*) + \sum_{j=1}^{r} v_j^* l_j(\underline{x}^*)$$

$$\leq f(\underline{x}^*)$$

Two inequalities hold with equality:

$\Rightarrow$ $\underline{x}^*$ minimizes $L(\underline{x}, \underline{u}^*, \underline{V}^*)$ [Stationary condition]

$\Rightarrow \quad u_i^* h_i(x^*) = 0$ or

$u_i^* > 0 \Rightarrow h_i(x^*) = 0$; $h_i(x^*) < 0 \Rightarrow u_i^* = 0$

$\underbrace{}_{\text{active constraint}}$ [complementary slackness]

② If there exists $\underline{x}^*$ and $\underline{u}^*, \underline{v}^*$ that satisfy the KKT condn.

$$g(\underline{u}^*, \underline{v}^*) = f(\underline{x}^*) + \sum_{i=0}^{m} u_i h_i(\underline{x}^*) + \sum_{j=1}^{r} v_j l_j(\underline{x}^*)$$

(Stationarity)

$$= f(\underline{x}^*)$$

(complementary slackness)

Thus duality gap is zero.

Example: $\underset{\underline{x}}{\text{minimize}} \quad \frac{1}{2}\underline{x}^T Q \underline{x} + \underline{c}^T \underline{x}$

$\text{s.t.} \quad A\underline{x} = b$

KKT: $L(\underline{x}, \underline{v}) = \frac{1}{2}\underline{x}^T Q \underline{x} + \underline{c}^T\underline{x} + \underline{v}^T(A x - b)$

$Ax^* = b \qquad Q\underline{x} + \underline{c} + A^T \underline{v}^* = 0$

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

$$\left.\begin{array}{l} \underset{\underline{x}}{\text{minimize}} \quad f(x) \approx \frac{1}{2}\underline{x}_t^T Q \underline{x}_t + \underline{c}^T \underline{x}_t \\ \qquad A\underline{x} = b \end{array}\right\}$$

Example: SVM

Given $y \in \{1, -1\}^n$, $X : n \times p$ with rows $\{\underline{x}_i\}$

$$\underset{\beta, \beta_0, \underline{\xi}}{\text{minimize}} \quad \frac{1}{2} \|\beta\|_2^2 + C \sum_{i=1}^{n} \xi_i$$

s.t. $\xi_i \geq 0$

$y_i (x_i^T \beta + \beta_0) \geq 1 - \xi_i \quad i = 1, \ldots n$

Dual variables: $\underline{v}$, $\underline{w}$

$$0 = \sum_{i=1}^{n} w_i y_i \ ; \quad \beta = \sum_{i=1}^{n} w_i y_i x_i \ , \quad \underline{w} = C\underline{1} - \underline{v}$$

Complementary slackness:

$$v_i \xi_i = 0 \ ; \quad w_i \left(1 - \xi_i - y_i (x_i^T \beta + \beta_0)\right) = 0$$

$i = 1, \ldots, n$

At optimality we have,

$$\underline{\beta} = \sum_{i=1}^{n} w_i y_i \underline{x_i} = \tilde{X}^T \underline{w}$$

and

$$w_i \neq 0 \quad \text{only if} \quad y_i(\underline{x_i}^T \beta + \beta_0) = 1 - \xi_i$$

Such points are called support points