## Last time: duality in linear programs

Given $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $G \in \mathbb{R}^{r \times n}$, $h \in \mathbb{R}^r$:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & Gx \le h \end{aligned}$$

Primal LP

$$\begin{aligned} \max_{u,v} \quad & -b^T u - h^T v \\ \text{subject to} \quad & -A^T u - G^T v = c \\ & v \ge 0 \end{aligned}$$

Dual LP

Explanation: for any $u$ and $v \ge 0$, and $x$ primal feasible,

$$u^T(Ax - b) + v^T(Gx - h) \le 0, \quad \text{i.e.,}$$
$$(-A^T u - G^T v)^T x \ge -b^T u - h^T v$$

So if $c = -A^T u - G^T v$, we get a bound on primal optimal value

2

Explanation # 2: for any $u$ and $v \ge 0$, and $x$ primal feasible

$$c^T x \ge c^T x + u^T(Ax - b) + v^T(Gx - h) := L(x, u, v)$$

So if $C$ denotes primal feasible set, $f^\star$ primal optimal value, then for any $u$ and $v \ge 0$,

$$f^\star \ \ge \ \min_{x \in C} L(x, u, v) \ \ge \ \min_x L(x, u, v) \ := \ g(u, v)$$

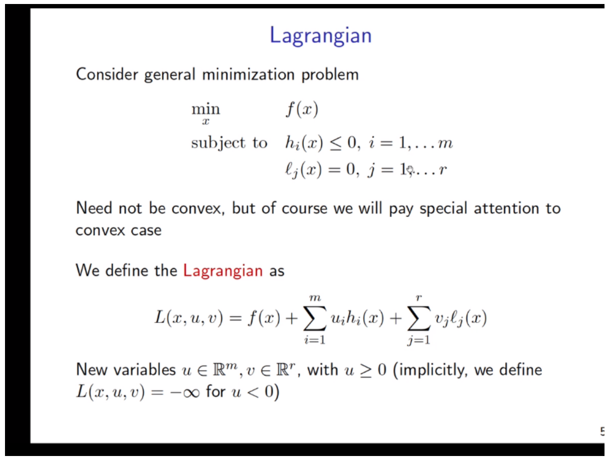
In other words, $g(u, v)$ is a lower bound on $f^\star$ for any $u$ and $v \ge 0$. Note that

$$g(u, v) = \begin{cases} -b^T u - h^T v & \text{if } c = -A^T u - G^T v \\ -\infty & \text{otherwise} \end{cases}$$

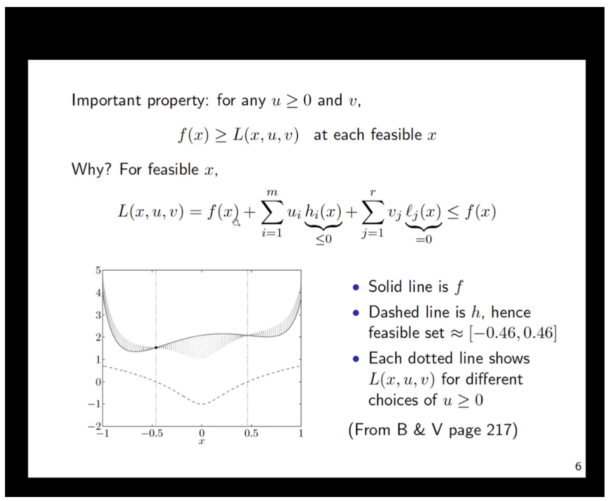
This second explanation reproduces the same dual, but is actually completely general and applies to arbitrary optimization problems (even nonconvex ones)

3

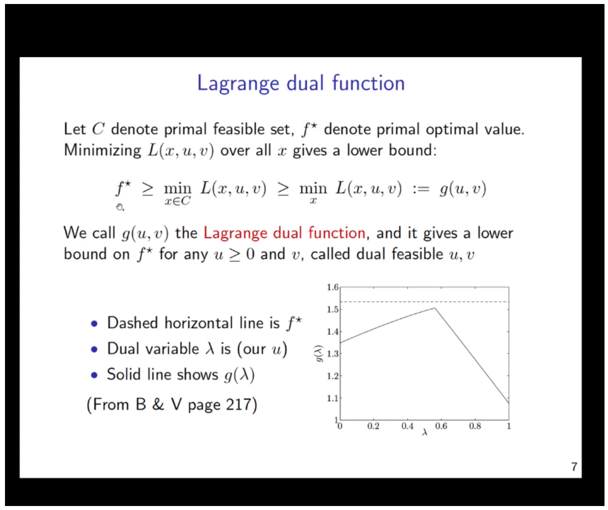
- the key insight is that removing the constraint on x is a lower bound on the constrained version of the problem
  - this is useful in most cases, because the constrained form of the problem is often not computable - this is why we are resorting to langrangians in the first place!
  - that is, if you could solve $min_{x \in C}\ L(x,\ u,\ v)$ then you could just as easily solve $min_{x \in C}\ f(x)$ to begin with
- the unconstrained Lagrangian minimum often has a nice closed form solution
- The lagrangian is a general approach - does not need linear program form
- **g(u, v) ≡ min of L(x, u, v) over all x evaluated at its argument!**
  - find expression for x in terms of u and v at this point
  - substitute to get an expression for g()




- u_i $\geqslant$ 0 for the inequality constraints



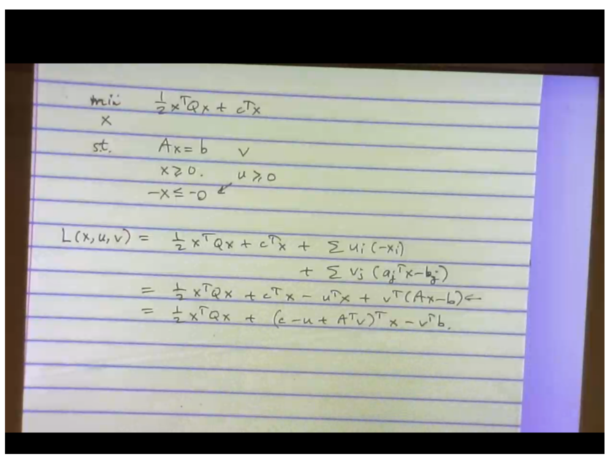

- in this simple example, when we are outside the feasible set [-0.46, 0.46], the Lagrangian instead upper bounds the function. Is this true in general?
  - no. In general you could violate a constraint and still provide a lower bound. e.g if h(x) = 0, obeying the constraint, but l(x) < 0, violating the constraint
  - you are outside the feasible region, but you will get a lower bound on the criterion function
- The claim is only that this inequality holds if x is feasible



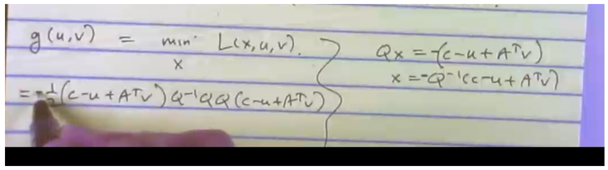

- Q: in this problem, why was the best lower bound not = the optimum criterion value? it is strictly less than it? Ans: the primal problem was not convex!

We have seen the Lagrange dual function g(u, v) for linear programs.

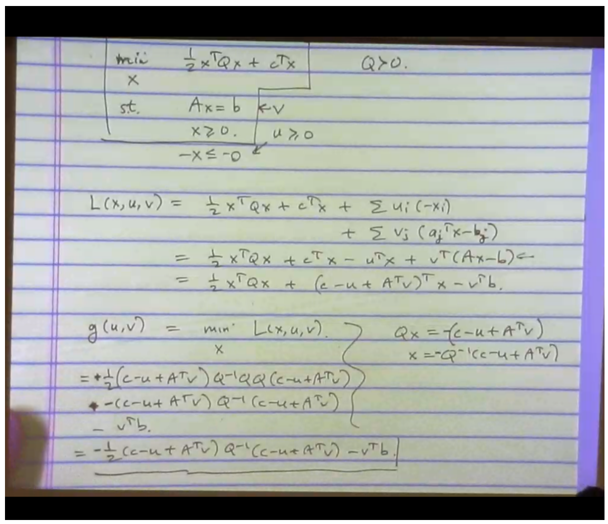
Expanding now to see what the dual is for quadratic programs.




Note that the earlier reasoning on why L(x, u, v) is a lower bound at feasibility - based onthe values of the constraints




- Set the gradient of L(x, u, v) to zero to get the min, since it is a quadratic (convex) function in x
- Note that this calculation assumes that Q is invertible i.e. all its eigenvalues are greater than 0 i.e. it is positive definite




- it is not immediately obiously why this is the lower bound on the primal quadratic! Even though we can reason through it easily

So we know that

$$g(u,\ v)\ =\ min_x\ L(x,\ u,\ v)\ =\ -\frac{1}{2}\left(c-u+A^Tv\right)Q^{-1}\left(c-u+A^Tv\right)\ -\ v^Tb$$

- To get the tightest lower bound possible, we would find $max\ g(u,\ v)$ over all u and v, which is a QP

**For LPs the dual is an LP, for QPs the dual is a QP!**

This is usually true - the structure of the primal is the structure of the dual

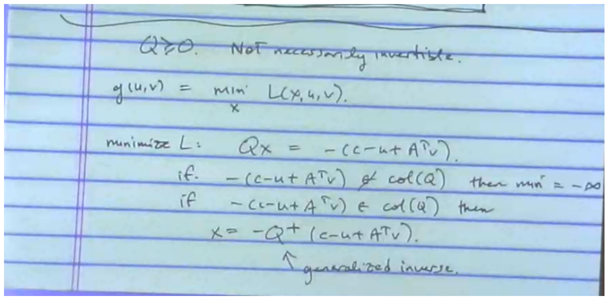
Testing

$$[x_1\ x_2\ x_3].\begin{matrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{matrix}.X\ =\ x_i(Q_{11}x_1+Q_{21}x_2\ +\ Q_{31}x_3)\ \ +x_2(Q_{12}x_1+Q_{22}x_2+Q_{32}x_3)\ \ +\ x_3(Q_{1}$$

$$\partial f/\partial x_1\ =\ 2Q_{11}x_1\ +\ Q_{21}x_2\ +\ Q_{31}x_3\ +\ x_2Q_{12}\ +\ x_3Q_{13}\ =\ Q_1x\ +Q_{\ 1}x$$

**Question: what if you relax the assumption that Q is positive defintite?** make it postiive semi definitey i.e. not necessarily invertible?

- we will see more constrains popping up in the dual




Have to consider two cases now when solving the system $Qx = -(c - u - A^T v)$

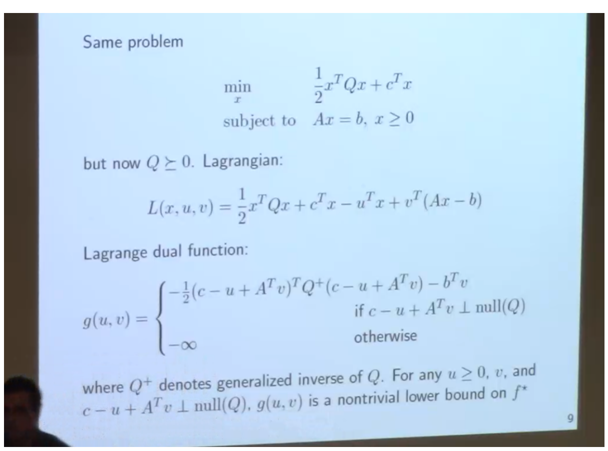
- case 1: $-(c - u + A^T v) \notin col(Q)$: this means there is no x for which this equality is
  true, which is possible when Q is singular. then there exists some x neq 0 for which $Qx = 0$ (i.e. 0 is an eigenvalue and x is the eigenvector) but will make $(c - u + A^T v)^T x$ arbitrarily large negative
  - https://math.stackexchange.com/questions/503585/show-that-a-matrix-a-is-singular-if-and-only-if-0-is-an-eigenvalue/503644

So for case 1, $g(u,\ v) = min_x\ L(x,\ u,\ v) = -inf$

- case 2: $-(c - u + A^T v) \in col(Q)$: in the case where Q is singular and the rhs is in the column space of Q, there are infinitely many solutions x to the system. The generalized or pseudo inverse $Q^+$gives us the min L2 norm solution to the linear system $Qx = -(c - u - A^T v) \implies x^* = Q^+(c - u + A^T v)$

So for case 2,

$g(u,\ v) = min_x\ L(x,\ u.\ v) \implies plug\ in\ the\ value\ of\ x^*\ back\ in\ the\ equation\ for\ L$

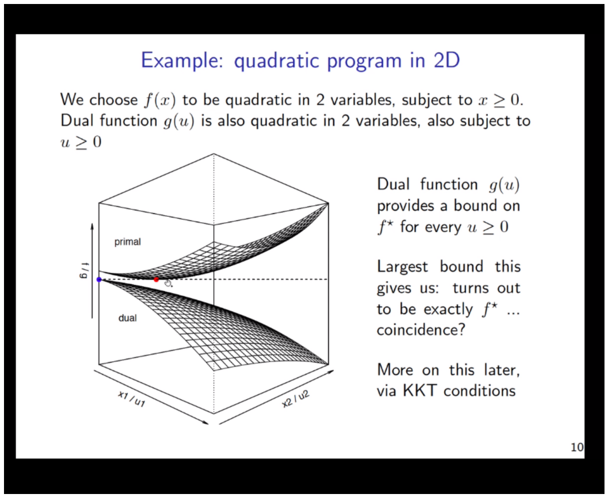


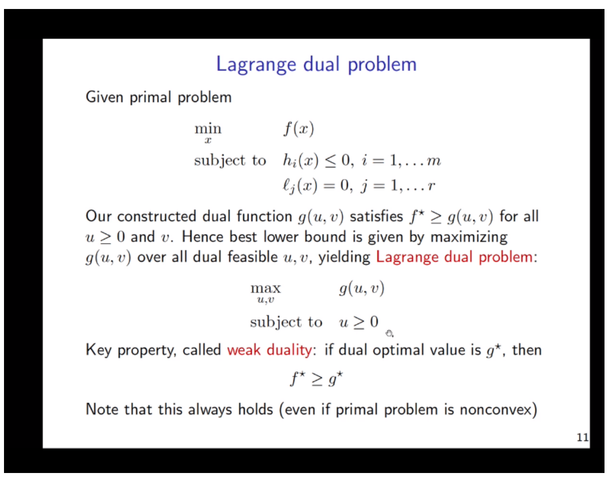
Note that the condition $c - u + A^T v \perp null(Q) \equiv c - u + A^T v \in col(Q)$

- for a symmetric matrix, row space = col space. And the null space is $\perp$ row space

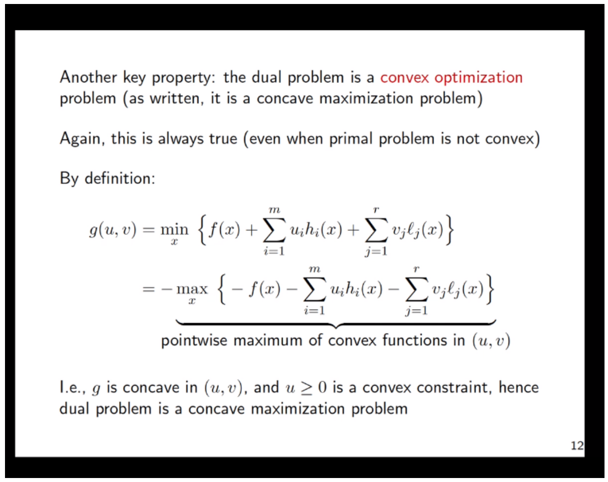
Also note that it forms a linear constraint on the quadratic form of $g()$ - the constraint is

$(c - u + A^T c)^T y = 0 \; for \; y \in null(Q)$

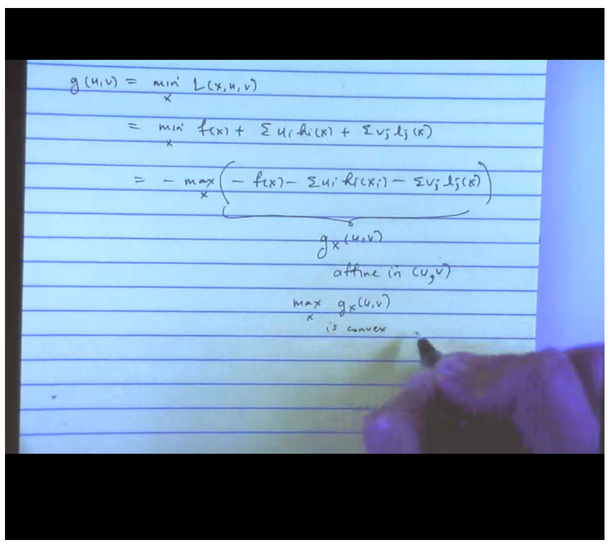


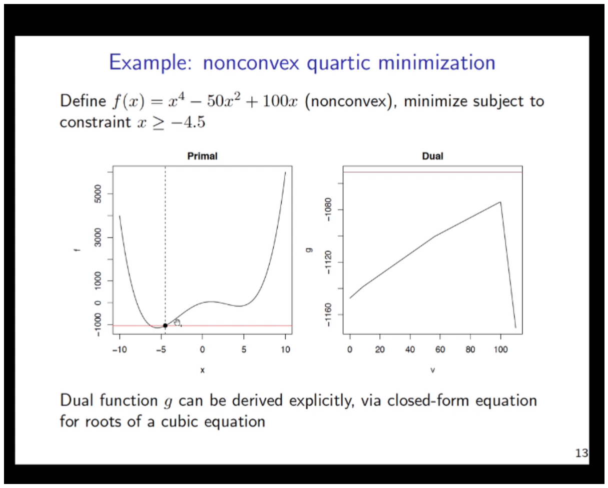




Note that **weak duality** ie the max of the dual $\geqslant$ min of the primal always holds! Even for non convex problems




The claim is that **g(u, v) = min_x L(x, u, v) is always concave in (u, v)**, no matter what form the f takes!
So the concave max problem = convex min problem (=min -g, since -g is convex if g is concave)

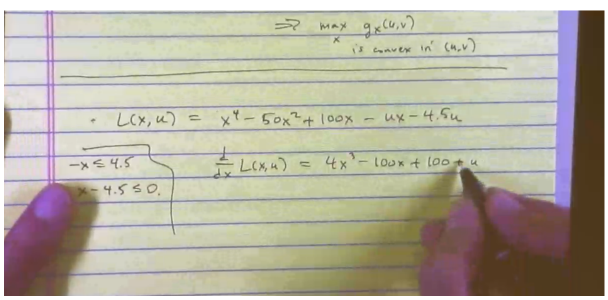


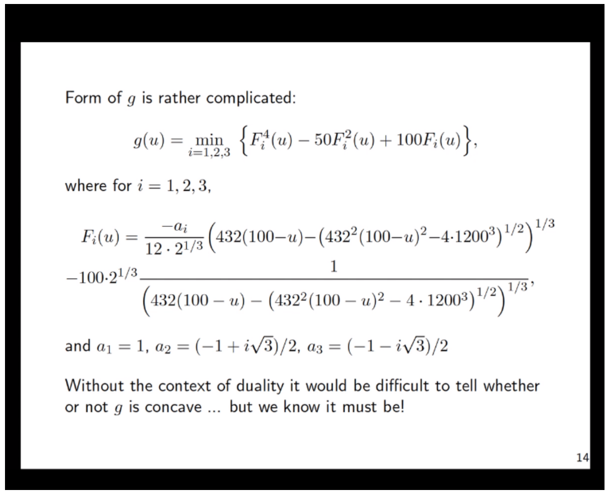



As a reminder:
we want to min f(x) st x in C
we know that f(x) st x in C ⩾ L(x, u, v) = f(x) + u(0) + v(⩽0) st x in C ⩾ g(u, v) = min_x f(x) + u(0) + v(⩽0) over all x
to get to the form g(u, v), we need to get the value of x that minimizes L(x, u, v) in terms of (u, v)

- do this by setting $\nabla L(x)$ to 0 and seeing what form the x takes (See B4 in Lecture 6 for why: 0 is a subgradient of any function, convex or not, at minimum)

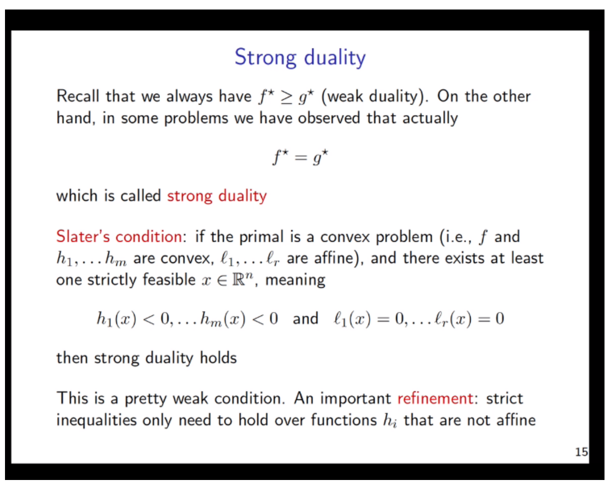
so if find max g(u, v), we find the best lower bound on f(x)






- this form comes from the closed form solution for the roots of a cubic: Cardano's formula
- you can plug in the roots values and see which one yields the smallest

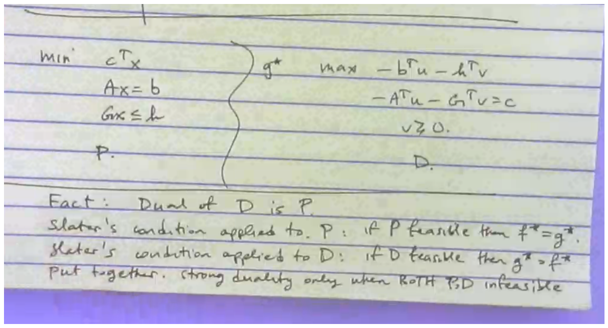
Lots of nonconvex problems do have closed form dual forms!




**Slater's condition:** If the primal is convex, and there is at least one x that fulfills the inequality constraints strictly (i.e is in the relative interior of the constraint set), then $min\ f(x)\ =\ max\ g(u,\ v)$!

- if all the non-affine constraints have a feasible point that satisfy them strictly, then this is true
- if we only have affine constraints, this means there are no conditions for Slater's condition to hold if the problem is convex!

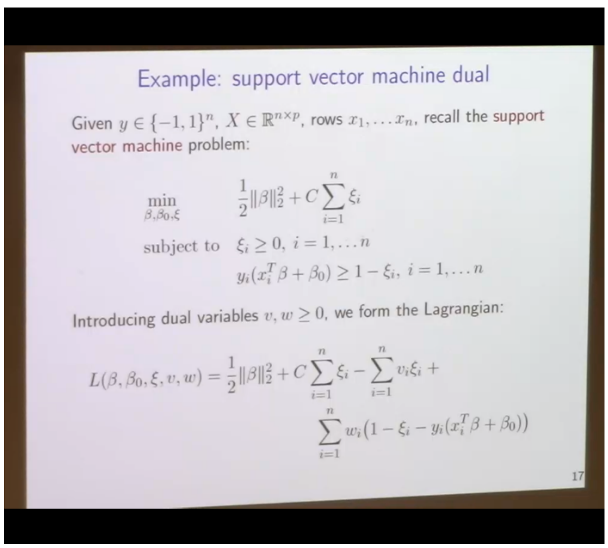
**So what does this mean for LPs?**

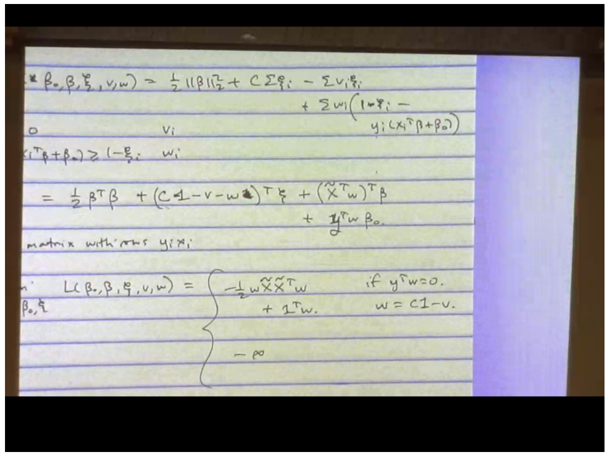



It is a challenge to construct such an LP for which Strong duality does not hold, with infeasible P and D!

- many LPs with infeasible P have a feasible D

**Example: SVM duals**





$L(B_0,\ B,\ \xi,\ v,\ w)\ =\ \dfrac{1}{2}B^T B\ +\ (c.1\ -\ v-w)^T \xi\ +\ \left(\widetilde{X}^T w\right)^T B\ +\ y^T w B_0\ +\ w.1$

*at min,* $\nabla L(B_0)\ =\ 0\ =\ y^T w$

*and* $\nabla L(\xi)\ =\ 0\ =\ (c.1\ -\ v-w)$

*and* $\nabla L(B)\ =\ 0\ =\ B\ +\ \left(\widetilde{X}^T w\right)\ \Longrightarrow\ B\ =\ -\left(\widetilde{X}^T w\right)$

*so then,* $g(v,\ w)\ =\ min\ L_{B_0,\ B,\ \xi}(B_0,\ B,\ \xi,\ v,\ w)\ =\ \dfrac{1}{2}B^T B\ +\ 0\ +\ \left(\widetilde{X}^T w\right)^T B\ \ +\ 0\ +\ w.1$

*substituing* $B\ =\ -\left(\widetilde{X}^T w\right),$

$=\ \dfrac{1}{2}B^T B\ +\ \left(\widetilde{X}^T w\right)^T B\ +\ w.1$

$=\ \left(\dfrac{1}{2} - \left(\widetilde{X}^T w\right)^T .\ -\left(\widetilde{X}^T w\right)\right)\ +\ \left(\left(\widetilde{X}^T w\right)^T .\ -\left(\widetilde{X}^T w\right)\right)\ +\ w.1$

$=\ \dfrac{1}{2}w\widetilde{X}\widetilde{X}^T w\ -\ w\widetilde{X}\widetilde{X}^T w\ +\ w.1$

$=\ -\dfrac{1}{2}w\widetilde{X}\widetilde{X}^T w\ +\ w.1$

*So the dual problem is*

*max* $g(v,\ w)$

$=\ max\ \left\{-\dfrac{1}{2}w\widetilde{X}\widetilde{X}^T w\ +\ w.1\right\}$

*subject to*

$0\ =\ y^T w$

$0\ =\ (c.1\ -\ v-w)\ \Longrightarrow\ w\ =\ c.1\ -\ v\ \Longrightarrow\ 0 \leqslant\ w\ \leqslant\ c.1$

*but we know* $w \geqslant 0$ *and* $v \geqslant\ 0,$ *because they are weights on the inequality* $\leqslant\ 0$ *conditions* *so we can eliminate v and instead say*

$\widetilde{X}\widetilde{X}^T$ *is the kernel matrix*!

Through Slater's condition, we know that max g() = min f() - i.e. the criterion values match.

In this case, we can even derive the *solution* from the dual! B = $\widetilde{X}\widetilde{X}^T$

Minimizing over $\beta, \beta_0, \xi$ gives Lagrange dual function:

$$g(v,w) = \begin{cases} -\frac{1}{2}w^T\tilde{X}\tilde{X}^T w + 1^T w & \text{if } w = C1 - v,\ w^T y = 0 \\ -\infty & \text{otherwise} \end{cases}$$

where $\tilde{X} = \mathrm{diag}(y)X$. Thus SVM dual problem, eliminating slack variable $v$, becomes

$$\begin{array}{ll} \max\limits_{w} & -\dfrac{1}{2}w^T\tilde{X}\tilde{X}^T w + 1^T w \\ \text{subject to} & 0 \le w \le C1,\ w^T y = 0 \end{array}$$

Check: Slater's condition is satisfied, and we have strong duality. Further, from study of SVMs, might recall that at optimality

$$\beta = \tilde{X}^T w$$

This is not a coincidence, as we'll later via the KKT conditions



## Duality gap

Given primal feasible $x$ and dual feasible $u, v$, the quantity

$$f(x) - g(u,v)$$

is called the duality gap between $x$ and $u, v$. Note that

$$f(x) - f^\star \le f(x) - g(u,v)$$

so if the duality gap is zero, then $x$ is primal optimal (and similarly, $u, v$ are dual optimal)

From an algorithmic viewpoint, provides a stopping criterion: if $f(x) - g(u,v) \le \epsilon$, then we are guaranteed that $f(x) - f^\star \le \epsilon$

Very useful, especially in conjunction with iterative methods ... more dual uses in coming lectures