COMP9501: Machine Learning

Computer Science Department
University of Hong Kong

Duality

Tools prepared for SVM

Jia PAN

Lecture 5

Duality in linear programs

- Suppose we want to find lower bound on the optimal value in our convex problem:

$$B \leq \min_{x \in C} f(x)$$

- For instance, consider the simple LP:

$$\min_{x,y} \; x + y$$

$$\text{subject to } x + y \geq 2$$

$$x, y \geq 0$$

- Luckily, the lower bound here is easy: take B = 2

Duality in linear programs

Try again

$$\min_{x,y} \; x + 3y$$
$$\text{subject to } x + y \geq 2$$
$$x, y \geq 0$$

$$x + y \geq 2$$
$$+ \; 2y \geq 0$$
$$= x + 3y \geq 2$$

Lower bound B = 2

More general

$$\min_{x,y} \; px + qy$$
$$\text{subject to } x + y \geq 2$$
$$x, y \geq 0$$

$$a + b = p$$
$$a + c = q$$
$$a, b, c \geq 0$$

Lower bound B = 2a, for any a, b, c satisfying above

Duality in linear programs

- What is the best we can do? Maximize our lower bound over all possible a, b, c

$$\min_{x,y} \; px + qy$$
$$\text{subject to } x + y \geq 2$$
$$x, y \geq 0$$

called primal LP

$$\max_{a,b,c} \; 2a$$
$$\text{subject to } a + b = p$$
$$a + c = q$$
$$a, b, c \geq 0$$

called dual LP

- Note: the number of dual variables is the number of primal constraints

Duality in linear programs

- Try another one

$$\min_{x,y} \; px + qy$$

$$\text{subject to } x \geq 0$$

$$y \leq 1$$

$$3x + y = 2$$

called primal LP

$$\max_{a,b,c} \; 2c - b$$

$$\text{subject to } a + 3c = p$$

$$-b + c = q$$

$$a, b \geq 0$$

called dual LP

- Note: for the ≤ constraints, need to convert them first into ≥ constraints
- Note: for = constraints, the corresponding dual variables are not constrained
- Note: for max program, the dual program is min, and since we need to find the minimum upper bound, ≤ constraint unchanged , but ≥ converts to ≤

Duality in linear programs

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

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

called primal LP

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

called dual LP

- Explanation #1: for any u and $v \ge 0$, and x primal feasible,
$$-u^T(Ax - b) - v^T(Gx - h) \ge 0$$
- Or $$(-A^T u - G^T v)^T x \ge -b^T u - h^T v$$
- So if let $c = -A^T u - G^T v$, we get a lower bound on primal optimal value

Another "deeper" perspective on LP duality

$$\min_x \; c^T x$$
$$\text{subject to } Ax = b$$
$$Gx \le h$$

called primal LP

$$\max_{u,v} \; -b^T u - h^T v$$
$$\text{subject to } -A^T u - G^T v = c$$
$$v \ge 0$$

called dual LP

- Explanation #2: for any u and $v \geq 0$, and x primal feasible:
- $c^T x \geq c^T x + u^T(Ax - b) + v^T(Gx - h) := L(x, u, v)$
- So if C denotes the unknown primal feasible set, f^* primal optimal value, then for any u and $v \geq 0$,
- $f^* = \min_{x \in C} c^T x \geq \min_{x \in C} L(x, u, v) \geq \min_{x \in R^n} L(x, u, v) := g(u, v)$

Another "deeper" perspective on LP duality

- In other words,

$$g(u, v) = \min_{x \in R^n} c^T x + u^T(Ax - b) + v^T(Gx - h)$$

is a lower bound on f^* for any u and $v \geq 0$

- Note that:

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

- Thus, if we maximize $g(u, v)$ over u and $v \geq 0$ and get the tightest bound, the result gives exactly the dual LP as before
- This last perspective is actually completely general, and applies to arbitrary optimization problems (even nonlinear and nonconvex ones)

Lagrangian

- Consider general minimization problem

$$\min_{x \in \mathcal{R}^n} \; f(x)$$

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

$$l_j(x) = 0, j = 1, \cdots, r$$

- Need not to be convex, but for sure we will pay special attention to convex cases
- We define the Lagrangian as

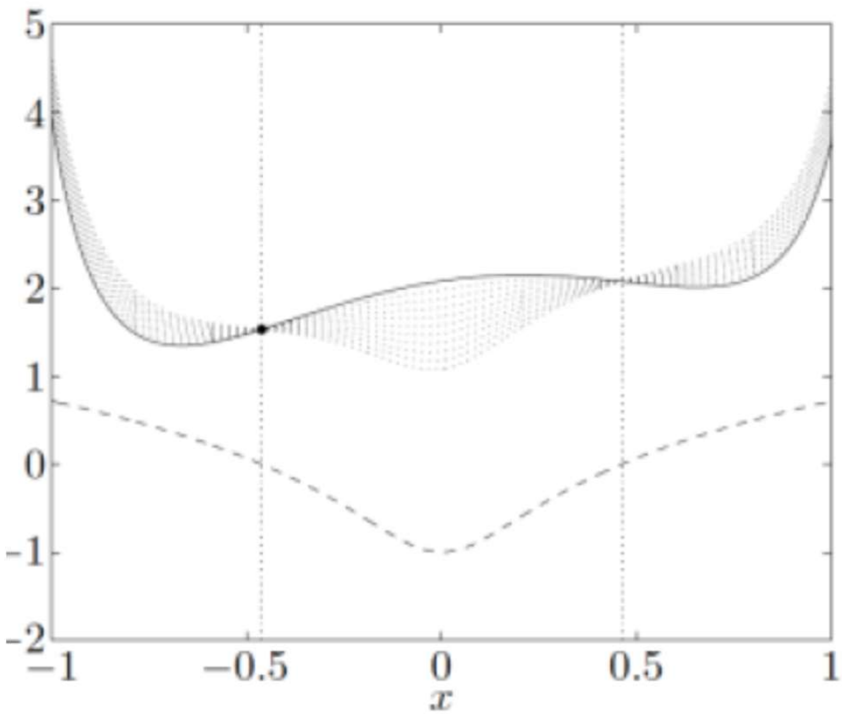
$$L(x, u, v) = f(x) + \textstyle\sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j l_j(x)$$

- New variables $u \in R^m, v \in R^r$ with $u \geq 0$ (implicitly we define $L(x, u, v) = -\infty$ for $u < 0$

Lagrangian

- For any $u \geq 0$ and v, at each feasible x, there is

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



- 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 and v

Lagrange dual function

- Same relaxation as before
- Let C denote primal feasible set, f^* denote primal optimal value
- Minimizing $L(x, u, v)$ over all $x \in R^n$ gives a lower bound:

$$f^* \geq \min_{x \in C} L(x, u, v) \geq \min_{x \in \mathcal{R}^n} L(x, u, v) := g(u, v)$$

- We call $g(u, v)$ the Lagrange dual function, and it gives a lower bound on f^* for any $u \geq 0$ and v (called the dual feasible u, v)

Example: Quadratic program

- Consider quadratic program (QP)

$$\min_{x \in \mathcal{R}^n} \ \frac{1}{2} x^T Q x + c^T x$$
$$\text{subject to } Ax = b, x \geq 0$$

  where $Q \succ 0$ is a positive definite matrix

- Lagrangian:

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

- Lagrange dual function:

$$\begin{aligned} g(u, v) &= \min_{x \in \mathcal{R}^n} L(x, u, v) \\ &= -\frac{1}{2}(c - u + A^T v)^T Q^{-1} (c - u + A^T v) - b^T v \\ &\leq f^* \end{aligned}$$

Example: Quadratic program

- Consider quadratic program (QP)

$$\min_{x \in \mathcal{R}^n} \ \frac{1}{2} x^T Q x + c^T x$$

$$\text{subject to } Ax = b, x \geq 0$$

where $Q \succeq 0$ is a positive semidefinite matrix

- Lagrangian is the same

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

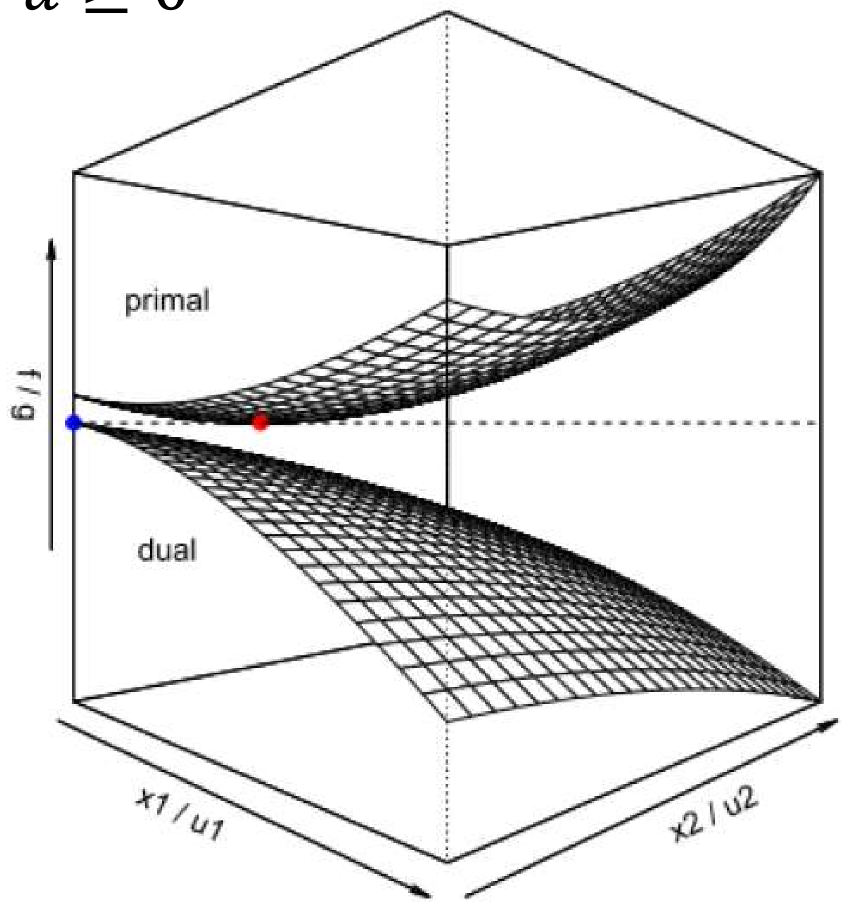
- Lagrange dual function:

$$g(u, v) = \begin{cases} -\frac{1}{2}(c - u + A^T v)^T Q^+ (c - u + A^T v) - b^T v & \text{if } c - u + A^T v \perp \text{null}(Q) \\ -\infty & \text{otherwise} \end{cases}$$

Q^+ denotes generalized inverse of Q

QP in 2D

- Choose $f(x)$ to be quadratic in 2 variables, s.t $x \geq 0$
- Dual function $g(u)$ is also quadratic in 2 variables, s.t. $u \geq 0$



- Dual function $g(u)$ provides a lower bound on f^* for every $u \geq 0$
- The computed largest bound turns out to be exactly f^*
- Is this coincidence?
- More on this later

General Lagrange dual problem

- Given primal problem

$$\min_{x \in \mathcal{R}^n} f(x)$$

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

$$l_j(x) = 0, j = 1, \cdots, r$$

- Our constructed dual function $g(u, v)$ satisfies $f^* \geq g(u, v)$ for all dual feasible (u, v)
- Hence the best lower bound is given by maximizing $g(u, v)$ over all dual feasible (u, v), yielding the Lagrange dual problem:

$$\max_{u \in \mathcal{R}^m, v \in \mathcal{R}^r} g(u, v)$$

$$\text{subject to } u \geq 0$$

Duality property

- Weak duality: the dual optimal value g^* and the primal optimal value f^* satisfy: $f^* \geq g^*$ (holds for general f)
- Dual problem is a convex optimization problem (i.e. a concave maximization problem)

Proof: By definition:

$$g(u,v) = \min_{x \in \mathcal{R}^n} \left\{ f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j l_j(x) \right\}$$

$$= -\underbrace{\max_{x \in \mathcal{R}^n} \underbrace{\left\{ -f(x) - \sum_{i=1}^{m} u_i h_i(x) - \sum_{j=1}^{r} v_j l_j(x) \right\}}_{\text{linear function in } (u,v)}}_{\text{pointwise maximization of convex functions in } (u,v)}$$

Weak duality and minimax/maximin

- The original problem can be rewritten as

$$\min_{x \in \mathcal{R}^n} \max_{u \geq 0, v} L(x, u, v)$$

- The dual problem is actually

$$\max_{u \geq 0, v} \min_{x \in \mathcal{R}^n} L(x, u, v)$$

- Since minimax is always larger than maximin
- Hence,

$$\begin{aligned} f^* &= \min_{x \in \mathcal{R}^n} \max_{u \geq 0, v} L(x, u, v) \\ &\geq \max_{u \geq 0, v} \min_{x \in \mathcal{R}^n} L(x, u, v) \\ &= g^* \end{aligned}$$

Strong duality

- Weak duality always holds: $f^* \geq g^*$
- In QP, we have observed that actually $f^* = g^*$, which is called the strong duality.

Slater's condition: if the primal is a convex problem (i.e. f and $h_1, \cdots, h_m$ are convex, $l_1, \cdots, l_r$ are affine), and there exists at least one strictly feasible $x \in R^n$, meaning that

$$h_1(x) < 0, \cdots, h_m(x) < 0, \text{ and } l_1(x) = 0, \cdots, l_r(x) = 0,$$

then strong duality holds

- This is a pretty weak condition. And it can be further refined by requiring strict inequalities only over functions h_i that are not affine

Strong duality of LPs

- For linear programs:
  - Easy to check that the dual of the dual LP is the primal LP
  - Refined version of Slater's condition: strong duality holds for an LP if it is feasible
  - Apply the same logic to its dual LP: strong duality holds if it is feasible
  - Hence strong duality holds for LPs, except when both primal and dual are infeasible
- In other words, we pretty much always have strong duality for LPs

What we have seen so far

- Given a minimization problem

$$\min_{x \in \mathcal{R}^n} \; f(x)$$

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

$$l_j(x) = 0, j = 1, \cdots, r$$

- We defined the Lagrangian:

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

- And the Lagrange dual function

$$g(u, v) = \min_{x \in \mathcal{R}^n} L(x, u, v)$$

What we have seen so far

- The subsequent dual problem is:

$$\max_{u \in \mathcal{R}^m, v \in \mathcal{R}^r} \; g(u, v)$$

$$\text{subject to } u \geq 0$$

- Important properties:
  - Dual problem is always convex, i.e. g is always concave (even if primal problem is not convex)
  - The primal and dual optimal values, f^* and g^*, always satisfy weak duality $f^* \geq g^*$
  - Slater's condition: for convex primal, if there is an x such that

$$h_1(x) < 0, \cdots, h_m(x) < 0, \text{ and } l_1(x) = 0, \cdots, l_r(x) = 0,$$

then strong duality holds: $f^* = g^*$ (can be further refined)

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 $f(x) - f^* \leq f(x) - g(u, v)$
- 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 $f(x) - g(u, v) \leq \epsilon$, then we are guaranteed that $0 \leq f(x) - f^* \leq \epsilon$
- Very useful, especially in conjunction with iterative methods.

Subgradients

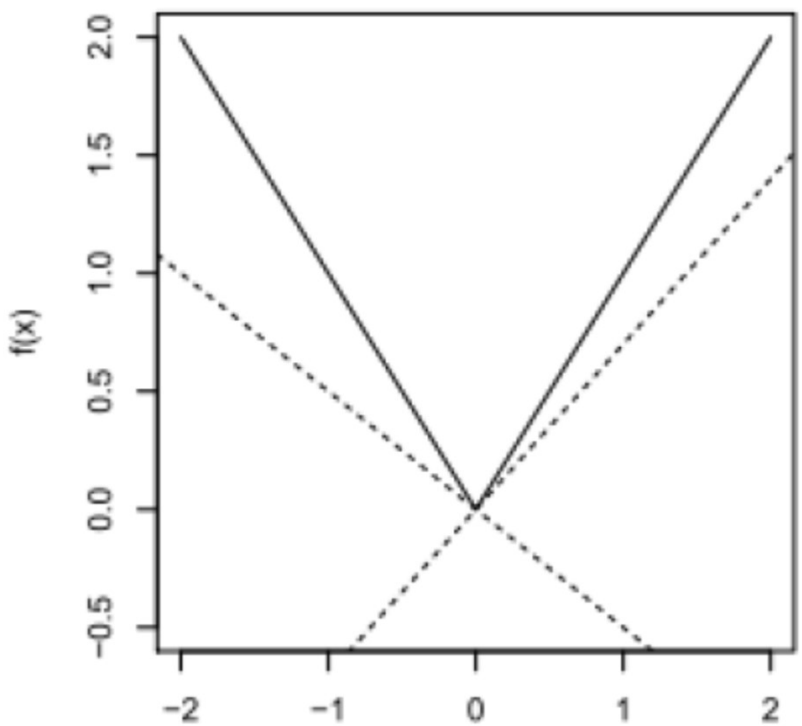
- Recall that for convex $f: R^n \to R$

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \text{ for all } x, y$$

- i..e linear approximation always underestimates f
- A subgradient of convex f at x is any $g \in R^n$ such that
- $f(y) \geq f(x) + g^T (y - x)$, for all y
- Subgradient
  – Always exists
  – If f differentiable at x, then $g = \nabla f(x)$ uniquely
  – Otherwise, subgradient can be among a set of vectors
  – Actually, the same definition works for nonconvex f (though subgradients need not exist for such f)

Subgradients: example

- Consider $f: R \to R, f(x) = |x|$



- For $x \neq 0$, unique subgradient $g = \mathrm{sgn}(x)$
- For $x = 0$, subgradient g is any element of [-1, 1]

Subdifferential

- Set of all subgradients of convex f is called the subdifferential:

$$\partial f(x) = \{g \in R^n \colon g \text{ is a subgradient of } f \text{ at } x\}$$

- $\partial f(x)$ is closed and convex (even for nonconvex f)
- Nonempty (can be empty for nonconvex f)
- If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$
- If $\partial f(x) = \{g\}$, then f is differentiable at x and $\nabla f(x) = g$

Optimality condition

For any f (convex or not),

$$f(x^*) = \min_{x \in R^n} f(x) \Leftrightarrow 0 \in \partial f(x^*)$$

- i.e., x^* is a minimizer if and only if 0 is a subgradient of f at x^*

Proof: g is a subgradient means that for all y:

$$f(y) \geq f(x^*) + 0^T(y - x^*)$$

Karush-Kuhn-Tucker (KKT) conditions

- Given a general problem

$$\min_{x \in \mathcal{R}^n} \; f(x)$$

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

$$l_j(x) = 0, j = 1, \cdots, r$$

- The KKT conditions are:
  - Stationarity: $0 \in \partial f(x) + \sum_{i=1}^{m} u_i \partial h_i(x) + \sum_{j=1}^{r} v_j \partial l_j(x)$
  - Complementary slackness: $u_i \cdot h_i(x) = 0, \forall i$
  - Primal feasibility: $h_i(x) \leq 0, l_j(x) = 0, \forall i, j$
  - Dual feasibility: $u_i \geq 0, \forall i$

Necessity – Part 1

- Let x^* and u^*, v^* be primal and dual solutions with zero duality gap (strong duality holds, e.g., under Slater's condition)
- Then $f(x^*) = g(u^*, v^*) = \min_{x \in \mathcal{R}^n} L(x, u^*, v^*)$

$$
\begin{aligned}
&= \min_{x \in \mathcal{R}^n} f(x) + \sum_{i=1}^{m} u_i^* h_i(x) + \sum_{j=1}^{r} v_j^* l_j(x) \\
&\leq f(x^*) + \sum_{i=1}^{m} u_i^* h_i(x^*) + \sum_{j=1}^{r} v_j^* l_j(x^*) \\
&\leq f(x^*)
\end{aligned}
$$

- In other words, all these inequalities are actually equalities

Necessity – Part 2

Two things to learn from this:

- The point x^* minimizes $L(x, u^*, v^*)$ over $x \in R^n$. Hence, the subdifferential of $L(x, u^*, v^*)$ must contain 0 at $x = x^*$ – this is exactly the stationary condition
- We must have $\sum_{i=1}^{m} u_i^* h_i(x^*) = 0$ so that the inequalities becomes equalities. And since each term here is ≤ 0, this implies $u_i^* h_i(x^*) = 0$ for all i. – this is exactly the complementary slackness
- Primal and dual feasibility obviously hold. Hence:

If x^* and u^*, v^* be primal and dual solutions with zero duality gap, then x^* and u^*, v^* satisfy the KKT conditions

- Note that this statement assumes nothing a priori about the convexity

Sufficiency

- If there exist x^* and u^*, v^* that satisfy the KKT conditions, then

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

- Where the first equality holds from stationarity, and the second holds from complementary slackness
- Therefore duality gap is zero (and x^* and u^*, v^* are primal and dual feasible), so x^* and u^*, v^* are primal and dual optimal

If x^* and u^*, v^* satisfy the KKT conditions, then x^* and u^*, v^* are primal and dual solutions

Putting everything together

- In summary, KKT conditions
  - Always sufficient
  - Necessary under strong duality
- Putting it together

For a problem with strong duality (e.g., assume Slater's condition: convex problem and there exists x strictly satisfying non-affine inequality constraints

x^* and u^*, v^* are primal and dual solutions

$\Leftrightarrow$ x^* and u^*, v^*satisfy the KKT conditions

- (Warning, concerning the stationarity condition: for a differentiable function f, we cannot use $\partial f(x) = \{\nabla f(x)\}$ until f is convex)

Example: quadratic with equality constraints

- Consider for $Q \succcurlyeq 0$

$$\min_{x \in \mathcal{R}^n} \; \frac{1}{2} x^T Q x + c^T x$$

$$\text{subject to } Ax = 0$$

- This is a core sub-problem used in Newton step for $\min_{x \in R^n} f(x)$ subject to $Ax = b$
- Convex problem, no inequality constraints, so by KKT conditions, x is a solution if and only if

$$\begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} -c \\ 0 \end{pmatrix}$$

- For some u. Linear system combines stationary and primal feasibility.