Computational Method of Optimization

Duality- Lecture VIII

Duality

- Agenda
- Lagrange dual problem
- weak and strong duality
- geometric interpretation
- optimality conditions
- perturbation and sensitivity analysis
- examples
- generalized inequalities

Source: Convex Optimization — Boyd & Vandenberghe

Lagrangian

standard form problem (not necessarily convex)

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{array}$$

variable $x \in \mathbf{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$

Lagrangian: $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, with $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

We start with a problem like this so minimize the objectives, some inequality constraints and quality constraints.
We form a Lagrangian and the Lagrangian is simply this. You take the objective and to the objective, you add a linear combination of the constraint functions and the equality functions.

Then we look at the so-called dual function or the Lagrangian dual function, it's got other names, but let's look at it. So the Lagrangian dual function is actually very, very simple. It just says minimize this Lagrangian over all xs. That's what it says. So you just minimize the Lagrangian.

Lagrange dual function

Lagrange dual function: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\
&= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)
\end{aligned}
$$

g is concave, can be $-\infty$ for some λ, ν

lower bound property: if $\lambda \succeq 0$, then $g(\lambda, \nu) \leq p^\star$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq 0$, then

$$
f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)
$$

minimizing over all feasible $\tilde{x}$ gives $p^\star \geq g(\lambda, \nu)$

This is the constrained approach where you simply say by $f_i(x)$ has to be less than zero, $h_i(x)$ has to be zero then we're going to let f_i float above zero, we'll charge you for it the λ_i are positive. That would be the meaning here. But if f_i goes less than zero, we'll pay you for it. You'll be subsidized and this is sort of the optimal cost under these sort of market prices.

If you look at this function as a function of λ and ν , for each x. It's affine. Yeah. It's linear plus, that's a constant. Okay. So, it's affine. And the family of affine functions is concave. So, this function g, the optimal – as a function of the prices, even if the original problem – even if these are affine, that's not convex. This dual function is concave.

Lagrange dual function

Lagrange dual function: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

$$
\begin{aligned}
g(\lambda,\nu) &= \inf_{x\in\mathcal{D}} L(x,\lambda,\nu) \\
&= \inf_{x\in\mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)
\end{aligned}
$$

g is concave, can be $-\infty$ for some λ, ν

lower bound property: if $\lambda \succeq 0$, then $g(\lambda,\nu) \le p^\star$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq 0$, then

$$f_0(\tilde{x}) \ge L(\tilde{x},\lambda,\nu) \ge \inf_{x\in\mathcal{D}} L(x,\lambda,\nu) = g(\lambda,\nu)$$

minimizing over all feasible $\tilde{x}$ gives $p^\star \ge g(\lambda,\nu)$

It says, basically, if you can evaluate g as a dual function, you're going to get a lower bound on the optimal value of the original problem. Imagine x is feasible. For any feasible x, $f_i(x) \le 0$, but if $\lambda_i \ge$ this term is ≤ 0 so therefore, this whole thing ≤ 0. If it is feasible, $h_i(x)$ is zero so it doesn't even matter what the side of new i, this is zero and, therefore, Lagrangian (L) $\le f_0(x)$ or any feasible x.

Now, for infeasible x's that's false, but for feasible x's it's true that $L(x, \lambda, \nu) \le f_0(x)$ because that's zero and that's ≤ 0. Now, if $\tilde{x}$ is well there, then $L < f_0(\tilde{x})$. This is true for any feasible $\tilde{x}$ and there's no conclusion possible other than this .

Least-norm solution of linear equations

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b \end{array}$$

dual function

- Lagrangian is $L(x, \nu) = x^T x + \nu^T (Ax - b)$
- to minimize L over x, set gradient equal to zero:

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \Longrightarrow \quad x = -(1/2) A^T \nu$$

- plug in in L to obtain g:

$$g(\nu) = L((-1/2)A^T \nu, \nu) = -\frac{1}{4} \nu^T A A^T \nu - b^T \nu$$

a concave function of ν

lower bound property: $p^\star \geq -(1/4)\nu^T A A^T \nu - b^T \nu$ for all ν

Let's do least norm solution of linear equations, so here's. We know how to solve this problem analytically. The Lagrangian function is this; $x^T x$, that's the objective, we add new transpose Ax-b so this as Ax-b = 0. You have to decide what h is. h is either Ax-b or something that's b-Ax so all that would happen there is the sign on ν, I've written it this way, Ax-b = 0 so. We're going to minimize this over x, that's completely trivial because that's a convex quadratic, that's affine in x and you just take the gradient, you get 2x +A^T ν = 0 and this is the optimal, that's the x that minimizes the Lagrangian here. Now, we take that x and we plug it back into here to get g so x that minimizes the Lagrangian, you get this and let's take a look at it.

Least-norm solution of linear equations

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b \end{array}$$

dual function

- Lagrangian is $L(x, \nu) = x^T x + \nu^T (Ax - b)$
- to minimize L over x, set gradient equal to zero:

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \Longrightarrow \quad x = -(1/2) A^T \nu$$

- plug in in L to obtain g:

$$g(\nu) = L((-1/2) A^T \nu, \nu) = -\frac{1}{4} \nu^T A A^T \nu - b^T \nu$$

  a concave function of ν

lower bound property: $p^\star \geq -(1/4) \nu^T A A^T \nu - b^T \nu$ for all ν

Evidently, this function here is concave because that is a positive semi-definite quadratic form. All we care about is a positive semi-definite quadratic form, but it happens to be positive definite. No, it doesn't matter in this case because I'm making no assumption about A whatsoever in this case; everything is true no matter what I assume about A. So this whole function is concave quadratic so there it is, which we knew had to happen because the dual function is always concave.

Standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array}$$

-x ≤ 0

dual function

- Lagrangian is

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\ &= -b^T \nu + (c + A^T \nu - \lambda)^T x \end{aligned}$$

- L is affine in x, hence

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

g is linear on affine domain $\{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\}$, hence concave

lower bound property: $p^\star \geq -b^T \nu$ if $A^T \nu + c \succeq 0$

That strings together three or four totally obvious things,
Let's do a quality constraint norm minimization. So here you minimize the norm of X subject to Ax = b.

Let's look at a standard form LP. So I want to minimize $c^T x$ subject to Ax = b and x ≽ 0.
Just standard LP. The Lagrangian is $c^T x$+, add a Lagrange multiplier v^T of Ax-b and -x is ≤ 0 so -λ^Tx. So that's Lagrangian.
What kind of function is the lagrangian in x?
L is affine in x. It's affine because there's this constant term – $v^T b$ here.

g is -∞ , and note that it is indeed a lower Bound. What's that? When the function is zero. And that's it. So in fact, if you minimize this over x, you will get -∞ . One exception. If c+ $A^T v$- λ is zero then this whole thing goes away and you get that.

However, if you find any vector v by any means, if you found a vector nu such as $A^T v$+c is a non-negative vector, then you evaluate –$b^T v$, that's the lower bound on this LP.

Equality constrained norm minimization

$$\begin{array}{ll} \text{minimize} & \|x\| \\ \text{subject to} & Ax = b \end{array}$$

dual function

$$g(\nu) = \inf_x (\|x\| - \nu^T Ax + b^T \nu) = \begin{cases} b^T\nu & \|A^T\nu\|_* \le 1 \\ -\infty & \text{otherwise} \end{cases}$$

where $\|v\|_* = \sup_{\|u\| \le 1} u^T v$ is dual norm of $\|\cdot\|$

proof: follows from $\inf_x(\|x\| - y^T x) = 0$ if $\|y\|_* \le 1$, $-\infty$ otherwise

- if $\|y\|_* \le 1$, then $\|x\| - y^T x \ge 0$ for all x, with equality if $x = 0$
- if $\|y\|_* > 1$, choose $x = tu$ where $\|u\| \le 1$, $u^T y = \|y\|_* > 1$:

$$\|x\| - y^T x = t(\|u\| - \|y\|_*) \to -\infty \quad \text{as } t \to \infty$$

lower bound property: $p^\star \ge b^T \nu$ if $\|A^T \nu\|_* \le 1$

Well, the Lagrangian is to minimize norm x - ν^TAx, this is a constant so it's totally irrelevant. Now, this goes back to the idea of a dual norm and let's look at that and if I want to minimize this thing – the question is what is this thing, what do you get here, right?
And the answer is straight from dual norms.

$$\inf_x \left(\|x\| - y^T x \right)$$

$$\inf_x \left(\|x\| - y^T x \right) = \begin{cases} -\infty & \|y\|_* > 1 \\ 0 & \|y\|_* \le 1 \end{cases}$$

In fact, we can do it for the two norm first. So, for the two norm, if norm y >1 in two norm then I can align x with it in that direction of -∞. Now, if norm y ≤ 1 and, therefore, this whole thing is ≥ 0. So I could never ever make this thing negative. On the other hand, by choosing x =0, I can make it zero, so that's clearly the optimal. So, this generalizes now to a general norm. It's either equal to -∞ if the dual norm of y is >1 or at 0 otherwise.

Equality constrained norm minimization

$$\begin{array}{ll} \text{minimize} & \|x\| \\ \text{subject to} & Ax = b \end{array}$$

dual function

$$g(\nu) = \inf_x (\|x\| - \nu^T Ax + b^T \nu) = \begin{cases} b^T\nu & \|A^T\nu\|_* \le 1 \\ -\infty & \text{otherwise} \end{cases}$$

where $\|v\|_* = \sup_{\|u\|\le 1} u^T v$ is dual norm of $\|\cdot\|$

proof: follows from $\inf_x(\|x\| - y^T x) = 0$ if $\|y\|_* \le 1$, $-\infty$ otherwise

- if $\|y\|_* \le 1$, then $\|x\| - y^T x \ge 0$ for all x, with equality if $x = 0$
- if $\|y\|_* > 1$, choose $x = tu$ where $\|u\| \le 1$, $u^T y = \|y\|_* > 1$:

$$\|x\| - y^T x = t(\|u\| - \|y\|_*) \to -\infty \quad \text{as } t \to \infty$$

lower bound property: $p^\star \ge b^T\nu$ if $\|A^T\nu\|_* \le 1$

And then you can go through the argument here and you've got something totally non-obvious. It basically says if you can come up with a vector ν, for which $A^T\nu \le 1$ in dual norm, then $b^T\nu$ is a lower bound on the optimal value of this problem

Two-way partitioning

$$\begin{array}{ll} \text{minimize} & x^TWx \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

- a nonconvex problem; feasible set contains 2^n discrete points
- interpretation: partition $\{1, \ldots, n\}$ in two sets; W_{ij} is cost of assigning i, j to the same set; $-W_{ij}$ is cost of assigning to different sets

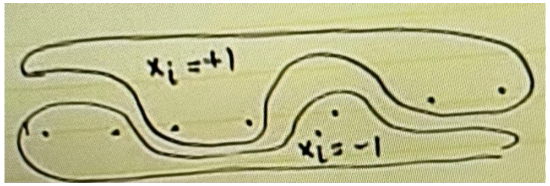
dual function

$$\begin{aligned} g(\nu) = \inf_x (x^TWx + \sum_i \nu_i(x_i^2 - 1)) &= \inf_x x^T(W + \mathbf{diag}(\nu))x - \mathbf{1}^T\nu \\ &= \begin{cases} -\mathbf{1}^T\nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

lower bound property: $p^\star \geq -\mathbf{1}^T\nu$ if $W + \mathbf{diag}(\nu) \succeq 0$

example: $\nu = -\lambda_{\min}(W)\mathbf{1}$ gives bound $p^\star \geq n\lambda_{\min}(W)$

It's two-way partitioning. I want to minimize a quadratic form subject to x_i^2=1 and this means x_i is +1 or -1. And let me first just say about this problem. We're gonna see it a lot and just so you get a rough idea of what it means.



So x_i is +1 and x_i is -1 so we can really think of this as the above is you have a set of points, like, M points and what you want to do is you're going to partition them into two groups. That's one group and then the other group will be this. And we encode that by saying here's where x_i is +1 and here's where x_i is -1 . It's a partition. It's just that it's a numeric data structure to encode a partition.

Two-way partitioning

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

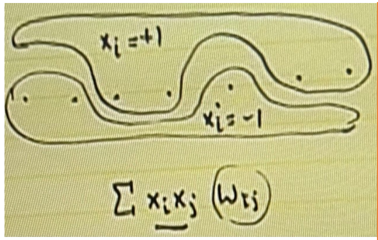
- a nonconvex problem; feasible set contains 2^n discrete points
- interpretation: partition $\{1, \ldots, n\}$ in two sets; W_{ij} is cost of assigning i, j to the same set; $-W_{ij}$ is cost of assigning to different sets

dual function

$$\begin{aligned} g(\nu) = \inf_x \left(x^T W x + \sum_i \nu_i (x_i^2 - 1)\right) &= \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu \\ &= \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

lower bound property: $p^\star \geq -\mathbf{1}^T \nu$ if $W + \mathbf{diag}(\nu) \succeq 0$

example: $\nu = -\lambda_{\min}(W)\mathbf{1}$ gives bound $p^\star \geq n\lambda_{\min}(W)$



Let's look at the objective is $\Sigma x_i x_j w_{ii}$ and so you sum over all pairs. If x_i and x_j are in the same partition, what is x_i and x_j?

Now, if x_i and x_j are in opposite partitions, this is negative w_{ii}. If w_{ii} is very high, it means that if x_i and x_j have the same side, you're assessed a big charge in the cost. I mean, if x_i is high and they're in opposite things, you're going to decrement the cost a lot. So w_{ii} is basically how much i aggravates j, but it's symmetric so it's the average of how much i aggravates j and j aggravates i.

Now, if w_{ii} is small, it means they don't care much. So in fact, you have a group of people, you have social network and you want to partition it.

There would be obvious ones. If the sign pattern in w_{ii} were such that like everybody liked everybody except one then it would be very simple. In general, finding a solution to this problem is extremely hard. You can't do it.

Two-way partitioning

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

- a nonconvex problem; feasible set contains 2^n discrete points
- interpretation: partition $\{1, \ldots, n\}$ in two sets; W_{ij} is cost of assigning i, j to the same set; $-W_{ij}$ is cost of assigning to different sets

dual function

$$\begin{aligned} g(\nu) = \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) &= \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu \\ &= \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

lower bound property: $p^\star \geq -\mathbf{1}^T \nu$ if $W + \mathbf{diag}(\nu) \succeq 0$

example: $\nu = -\lambda_{\min}(W)\mathbf{1}$ gives bound $p^\star \geq n\lambda_{\min}(W)$

If the quadratic form is positive semi-definite then the only values it takes on are non-negative so it couldn't be -∞ then so that's exactly the condition. The minimum of a quadratic form is -∞ if that matrix is not positive semi-definite so if it has one negative eigenvalue, the minimum is -∞. Otherwise, if it's positive semi-definite the minimum is zero because it can't be any lower than zero and it can be zero by plugging in zero. Let me tell you what the Lagrange dual function is. It is a lower bound on an optimization problem, gives a lower bound.

Lagrange dual and conjugate function

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax \preceq b, \quad Cx = d \end{array}$$

dual function

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathbf{dom}\, f_0} \left(f_0(x) + \underbrace{(A^T\lambda + C^T\nu)^T x} - b^T\lambda - d^T\nu \right) \\ &= \boxed{-f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu} \end{aligned}$$

- recall definition of conjugate $f^*(y) = \sup_{x \in \mathbf{dom}\, f}(y^T x - f(x))$
- simplifies derivation of dual if conjugate of f_0 is known

example: entropy maximization

$$f_0(x) = \sum_{i=1}^n x_i \log x_i, \qquad f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

The idea of the conjugate function and Lagrange duality. So if you have a function with just linear inequality and equality constraints, a problem, and you work out what the dual function is, it's a minimum of f_0 plus multiply by x and then that's, of course, a constant. And what this means is the following. If I focus on this and then go and look up what the conjugate function is, which was the conjugate over $y^T x$- f(x), that's the dual, if you plug in also all the right minuses, you get this. It's equal to that. Now, what that means is the Lagrange dual function of this thing is exactly equal to this.

Lagrange dual and conjugate function

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & Ax \preceq b, \quad Cx = d
\end{array}
$$

dual function

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_{x \in \mathbf{dom}\, f_0} \left(f_0(x) + (A^T\lambda + C^T\nu)^T x - b^T\lambda - d^T\nu \right) \\
&= \boxed{-f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu}
\end{aligned}
$$

- recall definition of conjugate $f^*(y) = \sup_{x \in \mathbf{dom}\, f}(y^T x - f(x))$
- simplifies derivation of dual if conjugate of f_0 is known

example: entropy maximization

$$
f_0(x) = \sum_{i=1}^n x_i \log x_i, \qquad f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}
$$

here's an example. The maximum entropy problem is maximize $-\Sigma$ x_i log x_i subject to some inequalities and equalities. It says find the Maximum entropy distribution that has these expected values – these are just known expected values. These can be moments, it could be probabilities, and these are inequalities on expected values. What's the maximum entropy distribution, for example, on these points that, for example, has the following variance and has the probability in the left tail less than that. That's the maximum entropy problem.

when $f_i(x)$ is the negative entropy here, that's minimized negative entropy, you will get the sum of exponentials. So, the dual function for a maximum entropy problem is going to involve a sum of exponentials. It's a connection between exponential families and maximum entropy.

The dual problem

Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

$$\begin{cases} -b^T\boldsymbol{\nu} & A^T\boldsymbol{\nu} + c - \lambda \\ -\infty & \text{otherwise} \end{cases}$$

- finds best lower bound on $p^\star$, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \mathbf{dom}\, g$ explicit

example: standard form LP and its dual

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c \succeq 0 \end{array}$$

We get to the dual problem and to write down the dual problem. This is always a convex optimization problem no matter what the primal problem was. This is called Lagrange dual, and it's just shortened to the dual problem. And it says simply maximize the dual function subject to this. The subject to the λ's being positive. Now, g is -∞ for some values of λ and $\boldsymbol{\nu}$. To find a point where it's -∞ , you know, this thing could be -∞ , that can happen. So, in fact, often what happens is you pull the implicit constraints in g out and make them explicit.

The dual problem

Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda,\nu) \\ \text{subject to} & \lambda \succeq 0 \end{array} \qquad \begin{cases} -b^T\boldsymbol{\nu} & A^T\boldsymbol{\nu} + c - \lambda \\ -\infty & \text{otherwise} \end{cases}$$

- finds best lower bound on $p^\star$, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda,\nu) \in \mathbf{dom}\, g$
- often simplified by making implicit constraint $(\lambda,\nu) \in \mathbf{dom}\, g$ explicit

example: standard form LP and its dual

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T\nu + c \succeq 0 \end{array}$$

The dual function for this LP looks like this. It is kind of going up on a line, but off that line, the thing falls off to -∞ and we're just simply going to maximize that subject to λ positive; however, it's easier to simply take the implicit constraint out and you end up with something that looks like this.

The Lagrange dual is also an LP. If you have a feasible ν here then $-b^T\boldsymbol{\nu}$ is a lower bound on the optimal value of this problem. This thing says, you have a family of lower bounds, please get for me the best lower bound." That's what the meaning of this problem is. So, for example in engineering design. if x's feasible, it satisfies the constraints, that's feasible here would be a sub optimal design.

The dual problem

Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array} \qquad \begin{cases} -b^T\boldsymbol{\nu} & A^T\boldsymbol{\nu} + c - \lambda \\ -\infty & \text{otherwise} \end{cases}$$

- finds best lower bound on $p^\star$, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \mathbf{dom}\, g$ explicit

example: standard form LP and its dual

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c \succeq 0 \end{array}$$

The big picture is you have an optimization problem and you form another one called the Lagrange dual. That Lagrange dual problem, essentially, is saying what is the best lower bound on the optimal value of the first one I can get using the Lagrange dual function. That is what's important.

Weak and strong duality

weak duality: $d^\star \leq p^\star$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

  for example, solving the SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array}$$

gives a lower bound for the two-way partitioning problem: Refer to slide 13

strong duality: $d^\star = p^\star$

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**

So now we get to the idea of weak and strong duality. The weak duality d* ≤ p*. So, in this context, the original problem is called the primal problem and the Lagrange dual is then called the dual problem. we've already assigned the symbol p* to mean the optimal value here. We're going to let d* be the optimal value of the dual problem. You always have d* ≤ p*. Any dual feasible point is a lower bound p* so the best one is also a lower bound. This is called weak duality.

You can solve this SDP (semi-definite program) and you will get a lower bound on the two-way partitioning problem. So, you do some heuristic. It's the $x^T wx$ **(Refer to slide 13)**. You can solve this SDP. You would have a partition, and such sub optimal.

Weak and strong duality

weak duality: $d^\star \leq p^\star$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

  for example, solving the SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array}$$

gives a lower bound for the two-way partitioning problem: Refer to slide 13

strong duality: $d^\star = p^\star$

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**

Strong duality is going to be that that lower bound is tight. That says, there's a lower bound that goes all the way up to the optimal value. That's strong duality and we'll see what its equivalent to, but that is not trivial. And by the way, it often doesn't happen. So in two-way partitioning problems, if it were true there, you'd have P = NP because this problem we can solve in polynomial time and so in fact, if p* = d* and there's even approximation. When a problem is convex, you usually have strong duality.

So a constraint qualification theorem goes like this. It says if the primal problem is convex and then you insert your constraint qualification here, then p* = d*. That's a constraint qualification.

Slater's constraint qualification

strong duality holds for a convex problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

if it is strictly feasible, *i.e.*,

$$\exists x \in \mathbf{int}\, \mathcal{D}: \qquad f_i(x) < 0, \quad i = 1, \ldots, m, \qquad Ax = b$$

- also guarantees that the dual optimum is attained (if $p^\star > -\infty$)
- can be sharpened: *e.g.*, can replace $\mathbf{int}\, \mathcal{D}$ with $\mathbf{relint}\, \mathcal{D}$ (interior relative to affine hull); linear inequalities do not need to hold with strict inequality, . . .
- there exist many other types of constraint qualifications

So let's call it Slater's Constraint Qualification and it says this, if you have a convex problem like this, it says if there is a strictly feasible point, if there exists one, then p⋆ = d⋆. Strictly feasible means not just that you met the inequality constraints, but you do so with positive margin for each one. That's the condition.
Now, I should add that it's completely clear, that for most problems that covers everything in engineering, where you don't have a strictly feasible point, but for most problems that come up in engineering, anything in machine learning, this makes perfect sense.

For example, if the third inequality was a limit on power, it doesn't make any sense to say just think about it? If Slater's condition failed to hold, it means their existing circuit dissipates 100 milli-watts, but there's no circuit that dissipates 99.999999 because if there were, Slater's condition would hold.

Inequality form LP

primal problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

dual function

$$g(\lambda) = \inf_x \left((c + A^T\lambda)^T x - b^T\lambda\right) = \begin{cases} -b^T\lambda & A^T\lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

dual problem

$$\begin{array}{ll} \text{maximize} & -b^T\lambda \\ \text{subject to} & A^T\lambda + c = 0, \quad \lambda \succeq 0 \end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ except when primal and dual are infeasible

Let's go to the inequality form linear program. Here you want to minimize c^Tx subject to $Ax \preccurlyeq b$. $g(\lambda)=c^Tx + \lambda^T(A^Tx - b)$ because we put the b on the left-hand side to make this $f \le 0$. We infamize (infimum) this, but we know how to infamize an affine function. You get $-\infty$ unless the linear part disappears so I get this and so this is the dual problem. Notice this is not the dual problem. This is a problem that is trivially equivalent to the dual problem.

Now, Slater's condition says that if the feasible set, the feasible set is a polyhedron and by the way, one possibility is the feasible set could be empty, which in fact, is a polyhedron. What Slater's condition says geometrically if that polyhedron has non-empty interior, it means, that there's an interior point, if it has a non-empty interior then you have strong duality, so you have $p^\star = d^\star$.

Quadratic program

primal problem (assume $P \in \mathbf{S}^n_{++}$)

$$\begin{array}{ll} \text{minimize} & x^T P x \\ \text{subject to} & Ax \preceq b \end{array}$$

dual function

$$g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b) \right) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

dual problem

$$\begin{array}{ll} \text{maximize} & -(1/4) \lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ always

Let's look at a quadratic program. Let's minimize x^TPx subject to Ax less than b. That's minimizing quadratic form over a polyhedron, the dual function is this x^TPx and we assume P is positive definite. So here the dual function is you infamize over x, $x^TPx+ \lambda^T(A^Tx – b$ here like that and now I minimize over x. Now, P is positive definite so I know how to minimize this. It's P inverse times whatever something. It's easy to minimize a strictly convex quadratic function so I minimize it. I plug that X back in here and I get this thing, which is I get $(1/4)\lambda^T AP^{-1} A^T \lambda - b^T \lambda$ and the dual problem looks like this

Quadratic program

primal problem (assume $P \in \mathbf{S}^n_{++}$)

$$\begin{array}{ll} \text{minimize} & x^T P x \\ \text{subject to} & Ax \preceq b \end{array}$$

dual function

$$g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b) \right) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

dual problem

$$\begin{array}{ll} \text{maximize} & -(1/4) \lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ always

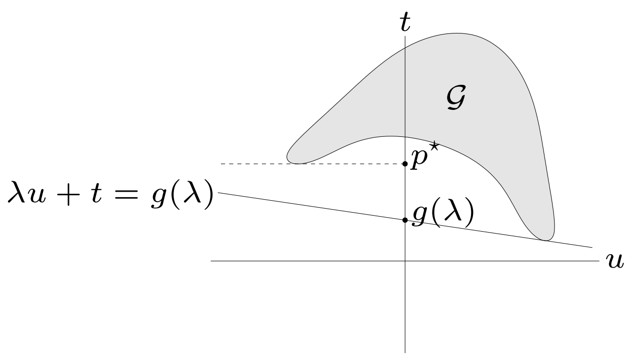
Should we believe that that's the optimal x, how do you prove it? Check out this λ ≽ 0 , and then you evaluate that number and that number is equal to the value up here of the point. That x is feasible and by the way, you would call that lambda a certificate proving it. Now, Slater's condition says the following: If this polyhedron has non-empty interior, then these are equal then their always exists a certificate proving optimality of the optimal x.

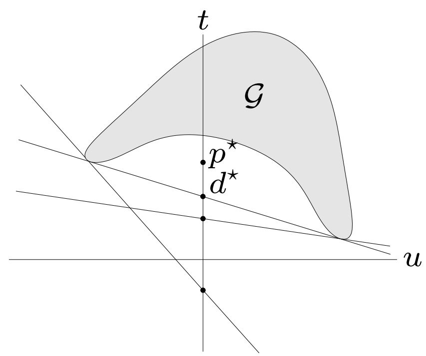
Geometric interpretation

for simplicity, consider problem with one constraint $f_1(x) \leq 0$

interpretation of dual function:

$$g(\lambda) = \inf_{(u,t)\in\mathcal{G}} (t + \lambda u), \qquad \text{where} \quad \mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$$





- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to $\mathcal{G}$
- hyperplane intersects t-axis at $t = g(\lambda)$

So for example, if you fix λ and then
to evaluate the dual. You fix a slope here and
you march down this way until
You would be right there. And then when
you work out what g(λ) is, it's this intersection
here. So this is g(λ) and now the dual problem
says, Optimize over all λ, so if λ is zero. You go
down there and g of zero is this number, which
is a lower bound on p*

Now, I move up the slope g and it keeps rising
until it just hits here, this point, at which point
here its right there. Now as I keep increasing λ,
the optimal is rotating around – it's not a fixed
point, it's rolling the context, but because it's
got sharp curves. It's rolling along here and
as I increase λ, g gets worse and worse. In fact,
if λ is huge, it looks like this, and g is very
negative. It's still a lower bound. So d^* is
that point.

Complementary slackness

assume strong duality holds, $x^\star$ is primal optimal, $(\lambda^\star, \nu^\star)$ is dual optimal

$$\begin{aligned} f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\ &\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\ &\leq f_0(x^\star) \end{aligned}$$

hence, the two inequalities hold with equality

- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
- $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, \ldots, m$ (known as complementary slackness):

$$\lambda_i^\star > 0 \Longrightarrow f_i(x^\star) = 0, \qquad f_i(x^\star) < 0 \Longrightarrow \lambda_i^\star = 0$$

We'll talk about complimentary slackness. we're covering right now involving duality – full optimality conditions. Assume strong duality holds. Now we have a primal optimal point x* and (λ*, ν*) be a dual optimal point . This means that f_0(x*) = g(λ*, ν*)

This means that the value achieved by x*, which is feasible, is equal to g(λ*, ν*). Now that's a lower bound on the optimal value. So if you've got a point that's feasible and has an objective value equal to the lower bound, you're done. It's optimal. So (λ*, ν*) as a certificate proving x* is optimal. By the way, the same is true the other way around.

Complementary slackness

assume strong duality holds, $x^\star$ is primal optimal, $(\lambda^\star, \nu^\star)$ is dual optimal

$$\begin{aligned} f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\ &\le f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\ &\le f_0(x^\star) \end{aligned}$$

hence, the two inequalities hold with equality

- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
- $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, \ldots, m$ (known as complementary slackness):

$$\lambda_i^\star > 0 \Longrightarrow f_i(x^\star) = 0, \qquad f_i(x^\star) < 0 \Longrightarrow \lambda_i^\star = 0$$

$x^\star$ is a primal certificate proving $(\lambda^\star, \nu^\star)$ are optimal. $g(\lambda^\star, \nu^\star)$ is by definition – it's the infimum over all x–here of this Lagrangian. That has to be less than or equal to the value of the Lagrangian if you plug in $x^\star$.

You plug that in and you get this. But $x^\star$ is feasible. But therefore, $f_i(x^\star)$ is ≤ 0. The $\lambda^\star_i \ge 0$. They have to be. That's dual feasibility. So this whole term $\lambda^\star_i f_i(x^\star) \le 0$. This term $\nu_i^\star h_i(x^\star)$ is equal to zero because these h_i's are zero. Therefore, this thing here is less than or equal to that $f_0(x^\star)$ because it's this thing plus zero, plus something less than or equal to zero.

if $\lambda^\star_i$ is positive, then the constraint is tight. If a constraint is slack, that means $f_i(x^\star)$ is ≤ 0, then you must have $\lambda^\star_i = 0$.

So this is called complimentary slackness, and the vector of these f's and the vector of λ's have complimentary sparsity patterns

Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i):

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$
2. dual constraints: $\lambda \succeq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

$\nabla_x L = 0$

from pages 26-27: if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

It's called the KKT conditions. So this is a generalization of an unconstrained problem, the necessary and sufficient conditions for optimality, or a convex function, is if the gradient is zero. these are for differentiable f_i and h_i.

You have to have primal feasibility. So x by definition, it could not be optimal if it weren't feasible. So feasible means that it satisfies the inequality constraints and the equality constraints. That's feasible. Dual constraint – dual feasibility says that the $\lambda_i \geq 0$. Complementary slackness says $\lambda_i f_i(x)$ is zero, and then the final one is gradient of the lagrangian with respect to x vanishes. That's this expression here. $\nabla_x L=0$. That's a gradient. So these four conditions together are called the KKT conditions.

KKT conditions for convex problem

if $\tilde{x}$, $\tilde{\lambda}$, $\tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal:

- from complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$

if **Slater's condition** is satisfied:

x is optimal if and only if there exist λ, ν that satisfy KKT conditions

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem

Now suppose you have a convex problem and some points, $\tilde{x}, \tilde{\lambda}$ $\tilde{\nu}$ satisfy KKT, in fact, they're optimal. x, is primal optimal and λ, ν are dual optimal. that works is this – from complimentary slackness – if complimentary slackness holds – if all of these conditions hold then, in fact, $\lambda_i f_i$ is zero, because for each i, λ_i is zero or f_i is zero. These are – all – the other things, the ν_i h_i – those are all zero, therefore, L($\tilde{x}, \tilde{\lambda}$ $\tilde{\nu}$), is equal to f_0 ($\tilde{x}$).

Now the fourth condition, that's this gradient condition here. If this holds, then

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

Because the h Is are affine. You can multiply them by any number and still have an affine function. The f Is are convex. I can multiply them by any non-negative numbers, still get a convex function. I can add up a bunch of convex functions to get a convex function differentiable

Introducing new variables and equality constraints

$$\text{minimize} \quad f_0(Ax+b)$$

- dual function is constant: $g = \inf_x L(x) = \inf_x f_0(Ax+b) = p^\star$
- we have strong duality, but dual is quite useless

reformulated problem and its dual

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & Ax+b-y=0 \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T\nu - f_0^*(\nu) \\ \text{subject to} & A^T\nu = 0 \end{array}$$

dual function follows from

$$\begin{aligned} g(\nu) &= \inf_{x,y}(f_0(y) - \nu^T y + \nu^T Ax + b^T\nu) \\ &= \begin{cases} -f_0^*(\nu) + b^T\nu & A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

A non-constrained problem – we haven't had a discussion of the dual of an unconstrained problem, so let's have it now.

So there are no constraints, so what's the Lagrangian? for each constraint, we should add a term, $\lambda_i f_i$, to this objective, and for each equality constraint we should add $v_i h_i$. But there are no constraints, so you're looking at Lagrangian. And there are no dual variables. there's no constraints. The next step is to calculate g, the dual function. you minimize the Lagrangian. So you minimize this. The minimum of that is p^*.

Minimize $f_0(y)$, subject to y= Ax+b. The dual of this, this is sort of a Lagrange dual, so you can say the Lagrange dual exists in this – the Lagrange dual maximize $b^T v - f_0^*(v)$ subject to $A^T v=0$

Norm approximation problem

norm approximation problem: minimize $\|Ax - b\|$

$$\begin{array}{ll} \text{minimize} & \|y\| \\ \text{subject to} & y = Ax - b \end{array}$$

can look up conjugate of $\|\cdot\|$, or derive dual directly

$$\begin{aligned} g(\nu) &= \inf_{x,y}(\|y\| + \nu^T y - \nu^T Ax + b^T\nu) \\ &= \begin{cases} b^T\nu + \inf_y(\|y\| + \nu^T y) & A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases} \\ &= \begin{cases} b^T\nu & A^T\nu = 0, \quad \|\nu\|_* \le 1 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

(See Slide 6)

dual of norm approximation problem

$$\begin{array}{ll} \text{maximize} & b^T\nu \\ \text{subject to} & A^T\nu = 0, \quad \|\nu\|_* \le 1 \end{array}$$

Unconstrained norm approximation problem where ‖ · ‖ is any norm. Here too the Lagrange dual function is constant, equal to the optimal value of ‖Ax − b‖ and therefore, not useful.

You work out what the conjugate function of the norm is, and the conjugate function of the norm is the indicator function of the dual norm unit ball. We've actually done that. But it ends up with the following problem.
The dual of this is sort of a Lagrange dual, so you can say the Lagrange dual exists in this – this the Lagrange dual of that.

you maximize $b^T\nu$, $A^T\nu=0$, and ‖ ν ‖ is less than one n dual mode

Implicit constraints

LP with box constraints: primal and dual problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -\mathbf{1} \preceq x \preceq \mathbf{1} \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T\nu - \mathbf{1}^T\lambda_1 - \mathbf{1}^T\lambda_2 \\ \text{subject to} & c + A^T\nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0 \end{array}$$

reformulation with box constraints made implicit

$$\begin{array}{ll} \text{minimize} & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \preceq x \preceq \mathbf{1} \\ \infty & \text{otherwise} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

dual function

$$\begin{aligned} g(\nu) &= \inf_{-\mathbf{1} \preceq x \preceq \mathbf{1}} (c^T x + \nu^T(Ax - b)) \\ &= -b^T\nu - \|A^T\nu + c\|_1 \end{aligned}$$

dual problem: maximize $-b^T\nu - \|A^T\nu + c\|_1$

There's another trick, which is this – and we've already seen this a couple of times. There's a difference explicit and implicit constraints. It doesn't really matter generally. It's kind of just a trick in labeling in general. When you form a dual, it is not a trick. So you will get different duals. So if you include, you know, when you have a constraint, you can either explicitly declare it in your list of constraints, or you can sort of secretly attach it to the function, the objective object, as part of the domain.

So here's a problem which says minimize a linear function subject to equality constraints and a box constraint on x. That's an LP here. You can write it out as an LP. This is an LP. That's the dual LP. And you have Lagrange multipliers for the two inequalities, that λ_1 is for this one, λ_2 is for this one.

Implicit constraints

LP with box constraints: primal and dual problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -\mathbf{1} \preceq x \preceq \mathbf{1} \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T\nu - \mathbf{1}^T\lambda_1 - \mathbf{1}^T\lambda_2 \\ \text{subject to} & c + A^T\nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0 \end{array}$$

reformulation with box constraints made implicit

$$\begin{array}{ll} \text{minimize} & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \preceq x \preceq \mathbf{1} \\ \infty & \text{otherwise} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

dual function

$$\begin{aligned} g(\nu) &= \inf_{-\mathbf{1} \preceq x \preceq \mathbf{1}} (c^T x + \nu^T(Ax - b)) \\ &= -b^T\nu - \|A^T\nu + c\|_1 \end{aligned}$$

dual problem: maximize $-b^T\nu - \|A^T\nu + c\|_1$

That's another L P. We're going to rewrite this problem this way, clearly they're equivalent. By the way, they're not the same problem because you would get different exceptions thrown at you when you pass in x, which violates the – if you give an x over here, the objective will be evaluated.

Now let's form the Lagrange dual and it'll make a big difference. So the dual function is f_0, as usual, plus v^T (Ax – b).

Problems with generalized inequalities

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad i = 1, \ldots, p
\end{array}
$$

$\preceq_{K_i}$ is generalized inequality on $\mathbf{R}^{k_i}$

definitions are parallel to scalar case:

- Lagrange multiplier for $f_i(x) \preceq_{K_i} 0$ is vector $\lambda_i \in \mathbf{R}^{k_i}$
- Lagrangian $L : \mathbf{R}^n \times \mathbf{R}^{k_1} \times \cdots \times \mathbf{R}^{k_m} \times \mathbf{R}^p \to \mathbf{R}$, is defined as

$$L(x, \lambda_1, \cdots, \lambda_m, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i^T f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- dual function $g : \mathbf{R}^{k_1} \times \cdots \times \mathbf{R}^{k_m} \times \mathbf{R}^p \to \mathbf{R}$, is defined as

$$g(\lambda_1, \ldots, \lambda_m, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda_1, \cdots, \lambda_m, \nu)$$

Last topic in duality is problems with generalized inequalities. So here we have $f_i(x)$ is a vector. Everything works the same, except how it works in our case is we form $\lambda_i f_i$ in the scaler case, but now f_i is a vector. So that λ_i to be a vector too. λ_i for scalers greater than or equal to zero. So for vectors, it means for λ_i to be positive, It had to be λ_i has to be positive in the dual cone. So here's what it is. This is the lagrangian. So f_i has to be less than or equal to zero in the cone K_i. If you look at sort of dual cones and everything, you can mark each thing as actually sort of being in the primal or dual space.

Lower bound property

lower bound property: if $\lambda_i \succeq_{K_i^*} 0$, then $g(\lambda_1, \ldots, \lambda_m, \nu) \leq p^\star$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq_{K_i^*} 0$, then

$$\begin{array}{rcl} f_0(\tilde{x}) & \geq & f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i^T f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \\ & \geq & \inf_{x \in \mathcal{D}} L(x, \lambda_1, \ldots, \lambda_m, \nu) \\ & = & g(\lambda_1, \ldots, \lambda_m, \nu) \end{array}$$

minimizing over all feasible $\tilde{x}$ gives $p^\star \geq g(\lambda_1, \ldots, \lambda_m, \nu)$

dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda_1, \ldots, \lambda_m, \nu) \\ \text{subject to} & \lambda_i \succeq_{K_i^*} 0, \quad i = 1, \ldots, m \end{array}$$

- weak duality: $p^\star \geq d^\star$ always
- strong duality: $p^\star = d^\star$ for convex problem with constraint qualification (for example, Slater's: primal problem is strictly feasible)

You have the lower bound property. If these λ_i's are larger than in the dual cone – larger than zero, then g is a lower bound on p^*. And, everything it's the same. These are zero, and then this is a product of elements in the cone – that's in the negative cone, and that's in the dual cone. And inner products of things in the dual cone and the cone are non-negative. Inner products of things in the dual cone and the negative dual cone are non-positive. And so this thing's less than or equal to zero. Everything works. Everything works, including Slater.

Semidefinite program

primal SDP ($F_i, G \in \mathbf{S}^k$)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n \preceq G \end{array}$$

- Lagrange multiplier is matrix $Z \in \mathbf{S}^k$
- Lagrangian $L(x, Z) = c^T x + \mathbf{tr}\left(Z(x_1 F_1 + \cdots + x_n F_n - G)\right)$
- dual function

$$g(Z) = \inf_x L(x, Z) = \begin{cases} -\mathbf{tr}(GZ) & \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\ -\infty & \text{otherwise} \end{cases}$$

dual SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{tr}(GZ) \\ \text{subject to} & Z \succeq 0, \quad \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \end{array}$$

$p^\star = d^\star$ if primal SDP is strictly feasible ($\exists x$ with $x_1 F_1 + \cdots + x_n F_n \prec G$)

It's a semi-definite program. So you want to minimize $c^T x$ subject to an LMI that's a linear matrix is less than or equal to, in matrix sense, another one. And the Lagrange multiplier here is going to be a matrix inequality. So it's going to be matrix Z symmetric. And it's going to be non-negative in the dual cone.

Now the dual of the positive semi-definite cone is the positive semi-definite cone, so Z's is going to end up being a positive semi-definite matrix. The inner product between two symmetric matrices is trace –it's trace $A^T B$. But if it's symmetric, it's just trace. So it's the trace of -GZ. So that's the Lagrangian. It's affine in x. It's minus infinity unless x_i component of the linear thing is equal to zero, in which case, this all drops away and what's left is the trace ZG. And so this is the dual STP.