Lagrange dualization, examples

Péter Hajnal

2024 Spring

The Original Problem

• Let's consider the following optimization problem specified with explicit conditions:

$$\boxed{\begin{array}{lll} \text{Minimize} & c(x) & \\ \text{Subject to} & f_i(x) \leq 0, & i = 1, \ldots, k, \\ & g_j(x) = 0, & j = 1, \ldots, \ell, \end{array}} \qquad \text{(P)}$$

where $x \in \mathbb{R}^n$, $c : \text{dom } (c)(\subset \mathbb{R}^n) \to \mathbb{R}$, $x = (x_1, \ldots, x_n)^\mathsf{T}$, and f_i and g_j are real-valued functions of n variables.

• Let's introduce a concise and simple notation. Let

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_k \end{pmatrix} : \cap_{i=1}^k \text{dom } f_i \subset \mathbb{R}^n \to \mathbb{R}^k, \text{ and } g = \begin{pmatrix} g_1 \\ \vdots \\ g_\ell \end{pmatrix} : \cap_{j=1}^\ell \text{dom } g_j \subset \mathbb{R}^n \to \mathbb{R}^\ell$$

Notation Technique

• Expressing our problem in this notation, it takes the following form:

Minimize	$c(x)$-t
subject to	$f(x) \preceq 0,$
	$g(x) = 0.$

• It's always good to keep in mind what the concise notation represents. For example, in the above, the 0s represent zero vectors in $\mathbb{R}^k$ and $\mathbb{R}^\ell$.

Dual Variables

• Next, we'll introduce new variables for the conditions. For each inequality, we'll introduce a λ_i and for each equality, we'll introduce a μ_i. These are called Lagrange multipliers or alternatively dual variables. As before, we'll utilize the vector notation:

$$\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_\ell \end{pmatrix}.$$

Lagrange Function

• We introduce the concept of the Lagrange function associated with the optimization problem.

Definition

$$L(x;\lambda,\mu) = c(x) + \sum_{i=1}^{k} \lambda_i f_i(x) + \sum_{j=1}^{\ell} \mu_j g_j(x) = c(x) + \lambda^\mathsf{T} f(x) + \mu^\mathsf{T} g(x).$$

• The domain of the Lagrange function coincides with the domain of the original optimization problem denoted by (P), which we labeled as $\mathcal{D}$.

Lagrange Function: Remark

Remark

If x is a feasible solution (i.e., $x \in \mathcal{L}$), and $0 \preceq \lambda$, then we have $c(x) \geq L(x; \lambda, \mu)$.

- Indeed: Since $x \in \mathcal{L}$, then for every j, $g_j(x) = 0$ and hence $\sum \mu_j g_j(x) = 0$.

For every i, $\lambda_i \geq 0$, and $f_i(x) \leq 0$, from which $\sum \lambda_i f_i(x) \leq 0$.

Adding $c(x) = c(x)$ to this and summing up the above, we get precisely the following:

$$L(x, \lambda, \mu) \leq c(x).$$

- Thus, for every non-negative coordinate λ and any μ, we obtain a lower bound for $c(x)$ by evaluating L.

Dual Objective Function

• It's advantageous if our lower bound doesn't depend on x. The following definition makes our lower bound depend solely on the dual variables.

Definition: Lagrange/Objective Function

$$\widetilde{c}(\lambda, \mu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \mu).$$

• Note that this also represents an optimization problem, but it has no constraints. More precisely, „the original constraints are incorporated into the objective function".

It immediately follows from the previous remark that for $x \in \mathcal{L}$ and $\lambda \succeq 0$,

$$c(x) \geq \widetilde{c}(\lambda, \mu).$$

This is because $c(x) \geq L(\lambda, \mu, x) \geq \widetilde{c}(\lambda, \mu)$.

Dual Problem

- Let's define the dual of problem (P).

Definition: Dual Optimization Problem

$$\boxed{\begin{array}{ll} \text{Maximize} & \widetilde{c}(\lambda, \mu) \\ \text{Subject to} & \lambda \succeq 0. \end{array}} \qquad \text{(D)}$$

- We denote the dual problem as (D), and its optimal value as d^*. (The original problem (P) is the primal problem; its optimal value is p^*).

Weak Duality Theorem

Weak Duality Theorem

$$p^* \geq d^*.$$

• This is obvious from the earlier discussions.

• The objective function of the dual problem is „guaranteedly nice”, expressed as a minimization problem,

Minimize	$-\widetilde{c}(\lambda,\mu)$-t
subject to	$\lambda \succeq 0.$

it will be convex:

Theorem

$\widetilde{c}(\lambda,\mu)$ is concave, thus $-\widetilde{c}(\lambda,\mu)$ is convex.

Duality: Terminology

- Many times, weak duality theorem holds with equality.
- In such cases, we say that strong duality holds.
- However, this is not necessary.
- When $p^* - d^* > 0$, we say there is a (positive) duality gap.

Break Time

Example I: Dualization of LP in Simplex Form

Example

Minimize	$c^{\mathsf{T}}x$-t
subject to	$Ax = b$
	$x \succeq 0$

- In this case, the Lagrange function is:

$$L(\lambda, \mu, x) = c^{\mathsf{T}}x - \lambda^{\mathsf{T}}x + \mu^{\mathsf{T}}(Ax - b) = (c^{\mathsf{T}} - \lambda^{\mathsf{T}} + (A^{\mathsf{T}}\mu)^{\mathsf{T}})x - \mu^{\mathsf{T}}b$$
$$= (c - \lambda + A^{\mathsf{T}}\mu)^{\mathsf{T}}x - b^{\mathsf{T}}\mu.$$

- To determine the value of the dual objective function at (λ, μ), we need to find the global minimum of a linear function.

Example I: Dualization of LP in Simplex Form: Digression

• Our task is to minimize the function $a^{\mathsf{T}}x + \alpha$.

• Minimizing linear functions with one variable is easy to visualize. The graph is a line. If this graph is a horizontal line (our function is a constant α), then α is the minimum. Otherwise, our function can take any small value.

• The situation is similar for multiple variables. If the coefficient vector is the 0 vector, then our linear function is constant. If one coordinate of a (one coefficient of x_i) is not 0, then the linear function can take any small value.

Remark

$$\inf_{x\in\mathbb{R}^n} a^{\mathsf{T}}x + \alpha = \begin{cases} \alpha, & a = 0 \\ -\infty, & a \neq 0 \end{cases}$$

Dualization Example I: LP in Simplex Form (continued)

- After the digression, the dualization becomes clear:

$$\widetilde{c}(\lambda,\mu) = \inf_{x\in\mathbb{R}^n}(c-\lambda+A^{\mathsf{T}}\mu)^{\mathsf{T}}x-b^{\mathsf{T}}\mu = \begin{cases}-b^{\mathsf{T}}\mu, & \text{if } c-\lambda+A^{\mathsf{T}}\mu=0,\\ -\infty, & \text{otherwise.}\end{cases}$$

- The dual problem:

Maximize	$\widetilde{c}(\lambda,\mu)$-t
subject to	$\lambda\succeq 0$

- Or equivalently:

Maximize	$-b^{\mathsf{T}}\mu$-t
subject to	$c-\lambda+A^{\mathsf{T}}\mu=0$
	$\lambda\succeq 0$

Dualization Example I: LP in Simplex Form (continued)

- Or equivalently:

Minimize	$b^{\mathsf{T}}\mu$-t
subject to	$c + A^{\mathsf{T}}\mu \succeq 0$

Thus, the dual of the LP problem in simplex form is also an LP problem. This LP problem is known as the polyhedral form of the LP problem.

Dualization Example II: LP in Polyhedral Form

Example

Minimize	$c^T x$-t
subject to	$Ax \preccurlyeq b.$

- We have:

$$L(x,\lambda) = c^{\mathsf{T}} x + \lambda^{\mathsf{T}} (Ax - b) = \left(c + A^{\mathsf{T}} \lambda\right)^{\mathsf{T}} x - b^{\mathsf{T}} \lambda.$$

- Thus,

$$\widetilde{c}(\lambda) = \begin{cases} -b^{\mathsf{T}}\lambda, & \text{if } c + A^T\lambda = 0, \\ -\infty, & \text{otherwise.} \end{cases}.$$

Dualization Example II: LP in Polyhedral Form (continued)

- The dual problem is:

Maximize	$-b^{\mathsf{T}}\lambda$-t
subject to	$c + A^{\mathsf{T}}\lambda = 0$
	$\lambda \succcurlyeq 0.$

Equivalently:

Minimize	$b^{\mathsf{T}}\lambda$-t
subject to	$c + A^{\mathsf{T}}\lambda = 0$
	$\lambda \succcurlyeq 0.$

- In the previous two examples, we dualized perhaps the two most common normal forms of LP problems. Both formalize the same problem domain. Due to the different forms, the dualization followed different paths. It turns out that the two forms are dual to each other.

Dualization Example II: LP: Strong Duality

• It may be known from operations research that while weak duality theorem always holds for LP problems, strong duality often holds as well. The only possibility for a positive duality gap is when $p^* = \infty$ and $d^* = -\infty$ simultaneously. That is, if one problem has finite optimum, then so does the other, and the two optimal values coincide.

Strong Duality Theorem for LP

Consider any LP problem. Exactly one of the following two possibilities holds:

(1)
$$p^* = \infty > -\infty = d^*,$$

(2)
$$p^* = d^*.$$

Dualization Example III: Flow Problem

Example

$\mathcal{H} = (\overrightarrow{G}, s, t, c)$ is a network, where $\overrightarrow{G}$ is a directed graph, s and t are two distinguished vertices (source and sink), and c is the capacity function. $c : E(\overrightarrow{G}) \to \mathbb{R} \equiv c \in \mathbb{R}^{E(\overrightarrow{G})}$.

Flow assigns a quantity to each edge such that it lies between 0 and the capacity of the corresponding edge (capacity constraints). Furthermore, it ensures the conservation of flow at every vertex except the source and sink.

We seek the flow f with maximum value.

- The flow function $f : E(\overrightarrow{G}) \to \mathbb{R}$ can be described as $f \in \mathbb{R}^{E(\overrightarrow{G})}$, i.e., $x = (f(e_1), \ldots, f(e_m))^{\mathsf{T}} \in \mathbb{R}^E$ as a vector.

- Capacities can also be handled as vectors. Algebraically, the capacity constraints are: $0 \preceq x \preceq c$.

Dualization Example III: Flow Problem (continued)

- The conservation law can also be written in algebraic form:

$$\sum_{e:vKe} x_e - \sum_{e:vBe} x_e = 0 \quad \text{for all } v \in V\setminus\{s,t\}.$$

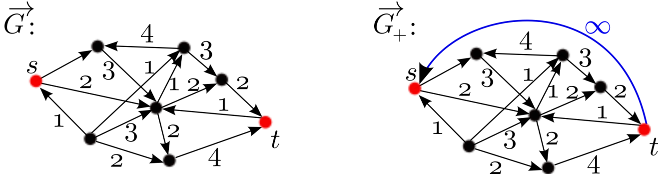
- The objective function/the value of the flow ($x = x(\text{flow})$)

$$c(x) = \text{val}(f) = \sum_{e:sKe} x_e - \sum_{e:sBe} x_e.$$

- Before us is the LP form of the flow problem. We introduce a little *twist* into the obvious formalization.

Dualization Example III: Flow Problem (continued)

- Consider the following modification of the network: We add an edge of infinite capacity (an edge without capacity constraint), leading from t to s in $\overrightarrow{G}$.



- In this $\overrightarrow{G}_+$ graph, a flow should remain on the old edges, and the e_+ edge should have the value of the flow. Thus, conservation is satisfied at every vertex (our network becomes a so-called circulation). Let $x_+ = \binom{x}{v}$ be the extended variable vector with the variable corresponding to the new edge, i.e., the new coordinate $v = \mathrm{val}(f)$, the value of the flow.

Dualization Example III: Flow Problem (continued)

- Let A be the incidence matrix of $\overrightarrow{G}$, and A_+ be the incidence matrix of $\overrightarrow{G}_+$.
- The flow problem is the following:

Maximize	v-t
subject to	$0 \preceq x \preceq c$,
	$A_+ x_+ = 0$.

- The linear equation system in matrix form corresponds to $|V|$ equations, each representing the conservation law written for all vertices except the source and sink. To dualize, we switch to the standard form:

Minimize	$-v$-t
subject to	$-x \preceq 0$,
	$x - c \preceq 0$,
	$A_+ x_+ = 0$.

Dualization Example III: Flow Problem (continued)

- The Lagrange function is given by:

$$L(x_+;\lambda_1,\lambda_2,\mu) = -v + \lambda_1^\mathsf{T}(-x) + \lambda_2^\mathsf{T}(x-c) + \underbrace{\underbrace{\mu^\mathsf{T} A_+ x_+}_{(A_+^\mathsf{T}\mu)^\mathsf{T} x_+}}_{(A^\mathsf{T}\mu)^\mathsf{T} x + (\mu_s - \mu_t)v} =$$

$$=(-1+\mu_s-\mu_t)v + (\lambda_2 - \lambda_1 + A^\mathsf{T}\mu)^\mathsf{T} x - \lambda_2^\mathsf{T} c$$

- From here, the dual objective function is:

$$\widetilde{c} = \begin{cases} -\lambda_2^\mathsf{T} c, & \text{if } (-1+\mu_s-\mu_t)=0 \text{ and } \lambda_2-\lambda_1+A^\mathsf{T}\mu = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Dualization Example III: Flow Problem (continued)

- The (D) dual problem is:

Maximize	$-\lambda_2^{\mathsf{T}} c$-t
subject to	$\mu_s - \mu_t = 1$
	$\lambda_2 = \lambda_1 - A^{\mathsf{T}} \mu$
	$\lambda_1, \lambda_2 \succeq 0$

- Below, we reconsider the dual problem using elementary methods.

Dualization Example III: Flow Problem: 1st + 2nd Observation

1st Observation

The goal is to minimize the components of λ_2, i.e., to make the coordinates of the (non-negative) variable vector as close to zero as possible. To achieve this, it suffices to choose μ and λ_1 components wisely.

- If μ is given, then choosing λ_1 is straightforward: If $(A^{\mathsf{T}}\mu)_e \geq 0$, then $(\lambda_1)_e = (A^{\mathsf{T}}\mu)_e$ is the optimal choice (in this case, $(\lambda_2)_e = 0$). If $(A^{\mathsf{T}}\mu)_e < 0$, then $(\lambda_1)_e = 0$ leads to the *best* $(\lambda_2)_e$.

2nd Observation

There exists an integral optimal solution.

- This is not trivial. We will prove it later in the semester.

Dualization Example III: Flow Problem: 3rd Observation

3rd Observation

In the constraints of the dual problem, the μ vector appears only as differences of two μ coordinates: $e = \overrightarrow{uv}$ edge has $(A^{\mathsf{T}}\mu)_e = \mu_v - \mu_u$. Moreover, it is advantageous if this difference — when negative — is as close to 0 as possible.

- If $\mu \in \mathbb{R}^V$ is a feasible solution, then for any constant c,

$$\mu + \begin{pmatrix} c \\ c \\ \vdots \\ c \end{pmatrix} = \mu + c \cdot 1^{\mathsf{T}}$$ is also a feasible solution and equivalent to the original μ.

- Due to such shifts, we can assume that the μ vector satisfies $\mu_s = 1$ and $\mu_t = 0$ (normalization).

Dualization Example III: Flow Problem (continued)

- From this, it can be computed which c_e edge capacities will have non-zero coefficients in the objective function. Exactly those edges leading from the set $S = \{v \in V : \mu(v) = 1\}$ to the set $T = \{v \in V : \mu(v) = 0\}$ (on such an edge $(A^{\mathsf{T}}\mu)_e = -1$, when the optimal choice assigns $(\lambda_1)_e = 0$ and $(\lambda_2)_e = 1$).
- That is, the capacity of the s-t cut defined by μ will be the value of the objective function.
- After the considerations, the dual problem turns out to be the problem of finding the minimal capacity cut:

Minimize	$C(\mathcal{V})$-t
subject to	$\mathcal{V}$ is an s-t cut.

- The weak duality theorem states that every cut capacity is an upper bound for every flow value. We also know that the optimal values of the two optimization problems are equal.

Break

Dualization Example IV: Least Squares Problem

Example

Minimize	$x^{\mathsf{T}}x$-t
subject to	$Ax = b$,

where $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{\ell \times n}$, $b \in \mathbb{R}^{\ell}$.

- Then
$$L(x;\mu) = x^{\mathsf{T}}x + \mu^{\mathsf{T}}(Ax - b).$$

- Expressing $\widetilde{c}$:

$$\widetilde{c}(\mu) = \inf_{x\in\mathbb{R}^n} L(\mu, x) = \inf_{x\in\mathbb{R}^n} (x^{\mathsf{T}}x + \underbrace{\mu^{\mathsf{T}}(Ax)}_{(A^{\mathsf{T}}\mu)^{\mathsf{T}}x}) - \underbrace{\mu^{\mathsf{T}}b}_{\text{independent of } x}$$

Dualization Example IV: Least Squares Problem (continued)

- The infimum over x is taken for a function depending on x as

$$\widetilde{L} = x^{\mathsf{T}}x + (A^{\mathsf{T}}\mu)^{\mathsf{T}}x.$$

- $\widetilde{L} : \mathbb{R}^n \to \mathbb{R}$ is a quadratic polynomial function, differentiable, so calculus tools can be applied for finding extrema.

- Gradient of L:

$$\nabla L = \text{grad } L = 2x + A^{\mathsf{T}}\mu.$$

- We know that at an extremum, the gradient is zero. $\nabla_x L = 0$ if and only if $x = -\frac{1}{2}A^{\mathsf{T}}\mu$.

Dualization Example IV: Least Squares Problem (continued)

• The zero gradient does not necessarily mean a minimum, but in our case, we have a convex function, so there will definitely be a minimum here. Therefore, substituting x with $-\frac{1}{2}A^{\mathsf{T}}\mu$, we get that

$$\widetilde{c}(\mu) = \left(-\frac{1}{2}A^{\mathsf{T}}\mu\right)^{\mathsf{T}} \cdot \left(-\frac{1}{2}A^{\mathsf{T}}\mu\right) + (A^{\mathsf{T}}\mu)^{\mathsf{T}} \cdot \left(-\frac{1}{2}A^{\mathsf{T}}\mu\right) - b^{\mathsf{T}}\mu =$$
$$= -\frac{1}{4}\mu^{\mathsf{T}}AA^{\mathsf{T}}\mu - b^{\mathsf{T}}\mu.$$

• The dual problem is then

Maximize	$-\frac{1}{4}\mu^{\mathsf{T}}AA^{\mathsf{T}}\mu - \mu^{\mathsf{T}}b$-t

• So the dual (D) problem is an unconstrained optimization question.

Dualization Example V: Maximum Cut

Example

Consider a simple graph G. The task is to find such a cut $\mathcal{V}$ (a partition of the vertices into two classes) where the $|E(\mathcal{V})|$ is maximized (maximize the number of edges crossing).

• First, let's formalize/arithmeticize the problem.

• We can describe a cut by encoding with an additional plus or minus 1 component for each vertex to indicate which side of the cut it falls on:

$$\mathcal{V} \equiv x \in \{-1, 1\}^V \subset \mathbb{R}^V.$$

Dualization Example V: Maximum Cut (continued)

- Let $A = A_G$ be the adjacency matrix of G.

- For the quadratic form $x^{\mathsf{T}}Ax$, each edge $e = uv$ contributes $2x_u x_v$. The value of $x_u x_v$ is $+1$ if edge e belongs to one side of the cut, and -1 if edge e belongs to the cut set (crosses).

- It's easy to calculate that

$$x^{\mathsf{T}}Ax = 2\,|E(G)| - 4\,|E(\mathcal{V})|\,.$$

- So the original problem's formalization is

Minimize	$x^{\mathsf{T}}Ax$,	$x \in \mathbb{R}^V$-t
subject to	$x_v^2 = 1$,	for all $v \in V$.

- This formalization of the problem is $\mathcal{NP}$-hard. Of course, we can determine its dual.

Dualization Example V: Maximum Cut (continued)

- The Lagrange function of the dual problem is

$$\begin{aligned} L(\mu, x) &= x^T A x + \sum_{v \in V} \mu_v \left(x_v^2 - 1\right) \\ &= x^T A x + \sum_v \mu_v x_v^2 - \sum_v \mu_v \\ &= x^T \left(A + \text{diag } \mu\right) x - 1^T \mu, \end{aligned}$$

where for a vector $a \in \mathbb{R}^n$ $\text{diag }(a) = \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & a_n \end{pmatrix}$ is an $n \times n$ diagonal matrix.

- From this, the dual objective function is

$$\widetilde{c}(\mu) = -\underline{1}^T \mu + \inf_{x \in \mathbb{R}^n} x^T \left(A + \text{diag } \mu\right) x.$$

Dualization Example V: Maximum Cut (continued)

• The infimum needs to be determined for every vector in $\mathbb{R}^n$, because the expression is defined for every such x, there are no restrictions.

• Now the question is, what is the (unconstrained) minimum of a homogeneous quadratic function in $\mathbb{R}^n$?

• To answer this, let's take a little detour.

Notation

Let $M \in \mathcal{S}^n \subset \mathbb{R}^{n\times n}$ be a symmetric matrix.

If M is positive semidefinite, then we write

$$M \succeq 0, \text{ or } M \in \mathcal{S}^n_+.$$

If M is positive definite, then we write

$$M \succ 0, \text{ or } M \in \mathcal{S}^n_{++}.$$

Dualization Example V: Maximum Cut: Digression

• What is the minimum of the real function wx^2? The answer is simple based on our elementary school studies:

$$\inf_{x\in\mathbb{R}} wx^2 = \begin{cases} 0, & \text{if } w \geq 0, \\ -\infty, & \text{if } w < 0. \end{cases}$$

• Let $W \in \mathcal{S}^n \subset \mathbb{R}^{n\times n}$, that is, an $n \times n$ symmetric matrix. Then

$$\inf_{x\in\mathbb{R}^n} x^T W x = \begin{cases} 0, & \text{if } W \succeq 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

• The case $W \succeq 0$ corresponds to the definition of positive semidefiniteness, and because of $0^T W 0 = 0$, it holds true.

• In the second case, since W is not positive semidefinite, substituting an appropriate vector x results in a negative value in the expression to be minimized. However, through scaling, the quadratic form can take arbitrarily small values as well.

Dualization Example V: Maximum Cut (continued)

- After the digression, we can now determine the dual objective function:

$$\widetilde{c}(\mu) = \begin{cases} -1^T\mu, & \text{if } A + \text{diag}\ \mu \succeq 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

- Thus, the dual problem is as follows:

Maximize	$-1^T\mu$-t
subject to	$A + \text{diag}\ \mu \succeq 0$.

- The dual problem is a semidefinite programming problem, manageable. Strong duality is not expected. However, every possible dual solution provides a lower bound on the value of p^*. We hope that a clever dual solution can provide a good approximation to p^*.

Dualization Example V: Maximum Cut (continued)

Consequence

Let G be any arbitrary simple graph, and λ_{min} be the smallest eigenvalue of graph G (the adjacency matrix A). Then

$$\frac{1}{2}\left|E(G)\right| \overset{(1)}{\leq} \max_{\mathcal{V}\text{ cut}}\left|E(\mathcal{V})\right| \overset{(2)}{\leq} \frac{1}{2}\left|E(G)\right| - \frac{\lambda_{min}}{4}\left|V(G)\right|.$$

Dualization Example V: Maximum Cut (continued)

- The inequality (1) follows from

$$\max_{\mathcal{V}\ \text{cut}} |E(\mathcal{V})| \geq \mathbb{E}(|E(\underline{\mathcal{V}})|),$$

where $\underline{\mathcal{V}}$ is a random cut in graph G (vector chosen uniformly at random from $\{-1,1\}^n$). We need to check how many times each edge contributes to the expected value. Let ξ be the following random variable:

$$\xi_e = \begin{cases} 1, & \text{if } e \in E(\underline{\mathcal{V}}), \\ 0, & \text{if } e \notin E(\underline{\mathcal{V}}), \end{cases}$$

which can be easily calculated to have $P(\xi_e = 1) = P(\xi_e = 0) = \frac{1}{2}$ for each e edge. Then

$$\mathbb{E}(E(\underline{\mathcal{V}})) = \mathbb{E}\left(\sum_{e\in E(G)} \xi_e\right) = \sum_{e\in E(G)} \mathbb{E}\left(\xi_e\right) = \sum_{e\in E(G)} \frac{1}{2} = \frac{1}{2}\left|E(G)\right|.$$

Dualization Example V: Maximum Cut (continued)

- To prove inequality (2), let's denote the optimal solution of the dual problem as d^*, and let μ be any feasible dual solution. Then it's obvious that $d^* \geq \widetilde{c}(\mu)$.
- The condition $A + \text{diag}\ \mu \succeq 0$ is equivalent to saying that all eigenvalues of the matrix $A + \text{diag}\ \mu$ are non-negative.
- Now we just need to choose a good μ vector. Let

$$\mu = -\lambda_{min}1 = \begin{pmatrix} -\lambda_{min} \\ -\lambda_{min} \\ \vdots \\ -\lambda_{min} \end{pmatrix}.$$

Dualization Example V: Maximum Cut (continued)

• The first claim is that the μ chosen in this way is a feasible solution: This means that the condition $A + \text{diag}\ \mu \succeq 0$ must be satisfied. If $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n = \lambda_{min}$ are the eigenvalues of matrix A, then the eigenvalues of matrix $A + \text{diag}\ \mu$ are $\lambda_1 - \lambda_{min} \geq \ldots \geq \lambda_n - \lambda_{min} \geq 0$. This can be easily proved by considering that the eigenvectors of A are also eigenvectors of $A + \text{diag}\ \mu$. Hence all eigenvalues of $A + \text{diag}\ \mu$ are non-negative.

• Thus, the μ chosen in this way satisfies the positive semidefiniteness condition. We have already seen that the optimal solution of the primal problem is $p^* = 2\,|E(G)| - 4\max|E(\mathcal{V})|$. Then

$$2\,|E(G)| - 4\max|E(\mathcal{V})| = p^* \geq d^* \geq c(\mu) = |V|\,\lambda_{min}.$$

Rearranging this gives us equality (2).

Break Time

Dualization Example VI: Minimizing Norm under Linear Constraints

Example

Minimize	$\|x\|$,	$x \in \mathbb{R}^n$,-t
subject to	$Ax = b$,	

where $\|.\| : \mathbb{R}^n \to \mathbb{R}$ is an arbitrary norm.

- For the L_2 norm, we have already seen how easy it is to dualize.
- For an arbitrary norm, the Lagrange function is

$$L(\mu, x) = \|x\| + \mu^T (Ax - b) = \|x\| + \left(A^T \mu\right)^T x - b^T \mu.$$

- Now we seek an infimum of this, and here too we need a little digression, as in the previous example.

Digression

• Our task is essentially to determine the infimum of an expression in the form $\|x\| + v^T x$.

Definition

Let $\|.\|$ be any norm. The

$$\|v\|_* = \sup \left\{ v^T x : \ \|x\| = 1 \right\}$$

norm is called the dual norm.

• From the definition of the dual norm and the properties of norms, it follows that $|v^T x| \leq \|v\|_*$, if $\|x\| = 1$. (Why?)

• Scaling leads to the following Lemma.

Digression (continued)

Lemma

$$\left|v^T x\right| \leq \|v\|_* \, \|x\| \, .$$

• Using the dual norm, the sought infimum can be expressed as follows:

$$\inf_{x \in \mathbb{R}^n} \|x\| + v^T x = \begin{cases} 0, & \text{if } \|v\|_* \leq 1, \\ -\infty, & \text{if } \|v\|_* > 1. \end{cases}$$

• In the first case, it is equivalent to $\|v\|_* \leq 1$ implying $\|x\| + v^T x \geq 0$ for all x vectors. This can be easily deduced from the above statement.

• In the second case, if we reach a negative value, then through scaling we can achieve arbitrarily large negative values, leading to the $-\infty$ infimum.

Dualization Example VI: Minimizing Norm under Linear Constraints (continued)

The dual objective function is as follows:

$$\widetilde{c}(\mu) = \begin{cases} -b^T\mu, & \text{if } \left\|A^T\mu\right\|_* \leq 1, \\ -\infty, & \text{if } \left\|A^T\mu\right\|_* > 1. \end{cases}$$

Thus, the dual problem is as follows:

Maximize	$-b^T\mu$-t
subject to	$\left\|A^T\mu\right\|_* \leq 1.$

Dualization Example VII: Optimization under Linear Constraints

Example

Minimize	$c(x)$-t
subject to	$Ax \preceq b,$
	$Cx = d.$

- The problem is much more general than an LP problem, where we also work with linear constraints. Here, the objective function can be any function.

- The Lagrange function of the problem is

$$\begin{aligned} L(\lambda, \mu, x) &= c(x) + \lambda^T(Ax - b) + \mu^T(Cx - d) \\ &= c(x) + \left(A^T\lambda + C^\mathsf{T}\mu\right)^T x - (b^T\lambda + d^T\mu). \end{aligned}$$

Digression

• Now we are seeking the infimum of an expression in the form $f(x) + v^T x$ over the domain dom (f).

Definition

The convex conjugate, or alternatively the Fenchel conjugate, of the function f is defined as

$$f^*(u) = \sup_{x \in \text{dom }(f)} u^T x - f(x).$$

• In the expression to be minimized, the role of u in the objective function will be played by $-v$:

$$\inf_x f(x) + v^\mathsf{T} x = -\sup_x -f(x) - v^\mathsf{T} x = -\sup_x -v^\mathsf{T} x - f(x) = -f^*(-v).$$

Dualization Example VII: Optimization under Linear Constraints (continued)

- Following this, the dual problem becomes

Maximize	$-c^*(-A^{\mathsf{T}}\lambda - C^{\mathsf{T}}\mu) - (b^{\mathsf{T}}\lambda + d^{\mathsf{T}}\mu)$-t
subject to	$\lambda \succeq 0.$

Dualization Example VIII: Maximizing Entropy

Example

Minimize	$\sum_{i=1}^{n} x_i \log x_i$-t
subject to	$Ax \preceq b$,
	$1^{\mathsf{T}} x = 1$.

• The objective function is negative entropy, resolving the apparent contradiction between the maximization sign and the minimization optimization task.

• Due to the appearance of the logarithm function, the components of the feasible solutions are positive. $1^{\mathsf{T}} x = 1$ indicates that x encodes a probability distribution. The linear inequalities $Ax \preceq b$ can be statistical observations about the distribution. For example, its expected value, variance, moments, estimation of the tail of the distribution, etc.

Dualization Example VIII: Maximizing Entropy (continued)

- This example is a special case of the previous one. Let's write down the

$$c(x) = \sum_{i=1}^{n} x_i \log x_i.$$

- It is easy to calculate that in the one variable case

$$(x \log x)^* = e^{y-1}.$$

- It can be easily shown that this implies

$$c^*(x) = \sum_{i=1}^{n} e^{y_i - 1}.$$

- Now we write down $\widetilde{c}\,(\lambda, \mu)$ function:

$$\widetilde{c}\,(\lambda, \mu) = -b^T\lambda - \mu - \sum_{i=1}^{n} e^{-a_i^T\lambda - \mu - 1} = -b^T\lambda - \mu - e^{-\mu-1}\sum_{i=1}^{n} e^{-a_i^T\lambda},$$

where a_i^{T} is the i-th row of the A^{T} matrix, i.e., the i-th column of A.

Dualization Example VIII: Maximizing Entropy (continued)

- The dual problem is:

Maximize	$\widetilde{c}\,(\lambda, \mu)$-t
subject to	$\lambda \succcurlyeq 0$.

- This can be easily simplified: If λ is fixed, then $\widetilde{c}$ is a single-variable real function, and thus a good μ value can be determined for each λ:

$$\mu = \log \sum_{i=1}^{n} e^{-a_i^{\mathsf{T}}\lambda} - 1.$$

- Substituting, the dual becomes equivalent to the following:

Maximize	$-b^{\mathsf{T}}\lambda - \log\left(\sum_{i=1}^{n} e^{-a_i^{\mathsf{T}}\lambda}\right)$-t
subject to	$\lambda \succcurlyeq 0$.

Break

Dualization Example IX

Example

Minimize	x_1x_2-t
subject to	$x_1 \geq 0$
	$x_2 \geq 0$
	$x_1^2 + x_2^2 \leq 1$

- The optimization problem is trivial: The product of non-negative numbers is non-negative, in our case $(0, 0)$ is a feasible solution.
- Thus, $p^* = 0$.
- Nevertheless, let's practice the learned dualization formalism.

Dualization Example IX (continued)

- The dual variables are λ_1, λ_2, and λ_3.
- The Lagrange function is:

$$x_1x_2 - \lambda_1 x_1 - \lambda_2 x_2 + \lambda_3(x_1^2 + x_2^2 - 1).$$

- The dual objective function is:

$$\widetilde{c}(\lambda) = \inf_{x\in\mathbb{R}^2} \left(x_1x_2 - \lambda_1 x_1 - \lambda_2 x_2 + \lambda_3(x_1^2 + x_2^2 - 1)\right) =$$

$$= \inf_{x\in\mathbb{R}^2} \left((x_1, x_2) \begin{pmatrix} \lambda_3 & 1/2 \\ 1/2 & \lambda_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - (\lambda_1, \lambda_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \lambda_3 \right).$$

Dualization Example IX (continued)

• The matrix of the quadratic part is positive definite if $\lambda_3 > 1/2$, positive semidefinite if $\lambda_3 = 1/2$, and indefinite if $\lambda_3 < 1/2$.

• It can be easily seen that in the indefinite case (having both positive and negative eigenvalues), $\widetilde{c}$ can take arbitrarily small (arbitrarily large absolute value negative) values.

• It can be easily seen that in the positive semidefinite case ($\lambda_3 = 1/2$), if $\lambda_1 - \lambda_2 \neq 0$, the $\widetilde{c}(\lambda_1, \lambda_2, \lambda_3)$ objective function can be arbitrarily small.

• For positive semidefinite matrices and $\lambda_1 = \lambda_2$, the value of $\widetilde{c}(\lambda_1, \lambda_2, \lambda_3)$ is $-\lambda_3$.

• In the case of positive definite matrix, it can be easily seen that a finite minimum exists. With our analytical knowledge, the minimum value can be easily determined as:

$$-\frac{1}{4}(\lambda_1, \lambda_2) \begin{pmatrix} \lambda_3 & 1/2 \\ 1/2 & \lambda_3 \end{pmatrix}^{-1} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} - \lambda_3.$$

Dualization Example IX (continued)

- The dual problem is:

Maximize	$\widetilde{c}(\lambda_1, \lambda_2, \lambda_3)$-t
subject to	$\lambda_1 \geq 0$
	$\lambda_2 \geq 0$
	$\lambda_3 \geq \frac{1}{2}$

- Elementary consideration suggests that the optimal points are: $(\lambda, \lambda, 1/2)$ and $d^* = -1/2$.

- Weak duality inequality is of course satisfied, but as a strict inequality.

Dualization Example $\widetilde{IX}$

Example

Minimize	x_1x_2-t
subject to	$x_1 \geq 0$
	$x_2 \geq 0$
	$x_1^2 + x_2^2 \leq 1$
	$x_1x_2 \geq 0$

- We repeated the previous example with an additional, obviously (mathematically) redundant constraint.
- Of course, p^* remains 0.

Dualization Example $\widetilde{\text{IX}}$ (continued)

- However, the dualization will change.
- A dual variable λ_4 appears.
- The Lagrange function is:

$$x_1x_2 - \lambda_1 x_1 - \lambda_2 x_2 + \lambda_3(x_1^2 + x_2^2 - 1) - \lambda_4 x_1 x_2.$$

- The dual objective function is:

$$\widetilde{c}(\lambda) = \inf_{x\in\mathbb{R}^2} \left(x_1x_2 - \lambda_1 x_1 - \lambda_2 x_2 + \lambda_3(x_1^2 + x_2^2 - 1) - \lambda_4 x_1 x_2\right) =$$

$$= \inf_{x\in\mathbb{R}^2} \left((x_1, x_2) \begin{pmatrix} \lambda_3 & \frac{1-\lambda_4}{2} \\ \frac{1-\lambda_4}{2} & \lambda_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - (\lambda_1, \lambda_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \lambda_3 \right).$$

Dualization Example $\widetilde{\text{IX}}$ (continued)

- The analysis/solution of the dual problem can be easily done following the previous reasoning.

- The calculation results in: the optimal point is $(0,0,0,1)$ and $d^* = 0$, strong duality holds.

- The optimal point and the dual optimal value were controlled by the fourth (corresponding to the redundant) dual variable. Adding the seemingly redundant constraint can be justified afterwards.

Dualization Example X

Example

Minimize	$c(u)$-t
subject to	$u \leq 0$

where $c(u) = -\left(\frac{u+1}{2}\right)^2$, $dom\,(c) = [-1, 1]$, meaning the graph of c is a parabolic arc.

- The domain of the objective function is unnatural. With this strange, unnatural domain, we built constraints into the objective function. This is not fair. This is cheating.

- Our goal is not to present an application, but to demonstrate the geometric view of duality gap.

Dualization Example X (continued)

- First, let's dualize the original problem. We use the notation $v = c(u)$.
- The Lagrange function is:

$$L(u, \lambda) = -\left(\frac{u+1}{2}\right)^2 + \lambda u = \lambda u + v.$$

- The dual objective function is:

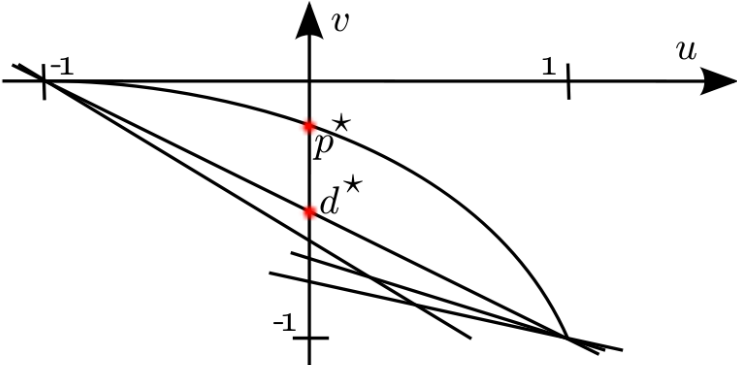
$$\widetilde{c}(\lambda) = \inf_{u \in [-1,1], v = \frac{1}{4}(u+1)^2} \lambda u + v.$$

- The dual is:

Maximize	$\widetilde{c}(\lambda) = \inf_{u \in [-1,1], v = \frac{1}{4}(u+1)^2} \lambda u + v$-t
subject to	$\lambda \geq 0$

Dualization Example X: The Figure

- With the following figure, we can visualize the solutions to the primal and dual problems:



The function $v = -\frac{1}{4}(u+1)^2$ is plotted in the u-v coordinate plane. Additionally, there are $v + \lambda u = \alpha$-type functions visible, which bound the objective function from below.

Dualization Example X: The Figure (continued)

• The solution to the primal is evident: On the feasible values ($[-1, 0]$), the function is monotonically decreasing, thus attaining a minimum at 0. So, $x^* = 0$ and $p^* = -1/4$.

• The Lagrange function is of the form $v + \lambda u$. If it takes the value α_0 for some λ_0, it bounds the objective function from below for $u \in [-1, 1]$. $v + \lambda_0 u \geq \alpha$ intersects the parabolic arc in the half-space.

• Thus, the lines of the form $\lambda u + v = \alpha$ are going *underneath* the graph. A specific line (λ_0) corresponds to the intersection with the v-axis.

• The dual optimum is the highest intersection with the v-axis.

• In the figure, the optimal value d^* is clearly visible, as well as the strict inequality $d^* < p^*$. There is no strong duality.

Dualization Example XI

Example

Let $n = 2$, $c(x, y) : \mathbb{R} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$, $(x, y) \mapsto e^{-x}$. Our optimization problem is the following:

Minimize	$c(x, y)$-t
subject to	$\frac{x^2}{y} \leq 0$

- Then, the domain of the optimization problem is as follows: $\mathcal{D} = \mathbb{R} \times \mathbb{R}_{>0}$.
- To satisfy the constraint, due to the non-negativity of x^2 and $y > 0$, x must be 0.
- The feasible solution set for the problem is

$$\{(x, y) : x = 0, y > 0\}.$$

Dualization Example XI (continued)

- If we restrict the objective function to $\mathcal{L}$, then $c|_{\mathcal{L}} = e^{-0} = 1$ becomes a constant function.

- It follows that the optimal value of the primal problem is 1, thus $p^* = 1$.

- Let's write the Lagrange function for the problem:

$$L(x, y; \lambda) = c(x, y) + \lambda \frac{x^2}{y} = e^{-x} + \lambda \frac{x^2}{y}.$$

- Then, the dual objective function is as follows:

$$\widetilde{c}(\lambda) = \inf_{(x,y)\in\mathcal{D}} L(x, y; \lambda) = \inf_{(x,y)\in\mathcal{D}} \left(e^{-x} + \lambda \frac{x^2}{y} \right) = \begin{cases} 0 & \text{if } \lambda \geq 0 \\ -\infty & \text{otherwise} \end{cases}$$

Dualization Example XI (continued)

- So we can write down the dual optimization problem:

Maximize	$\widetilde{c}(\lambda)$-t
subject to	$\lambda \geq 0.$

- Its optimal value is $d^* = 0$.
- Note that the weak duality theorem holds in this case, since $p^* \geq d^*$.
- In our case, the inequality is strict, creating a so-called *duality gap*, because $p^* - d^* = 1 > 0$.

This is the End!

Thank you for your attention!