Duality (II)

Lecture 10, Convex Optimization

National Taiwan University

May 13, 2021

Table of contents

Certificate of suboptimality (1/3)

- If we can find a dual feasible (λ, ν), we establish a lower bound on the optimal value of the primal problem: $p^* \geq g(\lambda, \nu)$. It provides a proof or certificate that $p^* \geq g(\lambda, \nu)$.
- Dual feasible points allow us to bound how suboptimal a given feasible point is, without knowing the exact value of p^*. If x is primal feasible and (λ, ν) is dual feasible, then

$$f_0(x) - p^* \leq f_0(x) - g(\lambda, \nu).$$

- In particular, this establishes that x is ϵ-suboptimal, with $\epsilon = f_0(x) - g(\lambda, \nu)$.
    - It also establishes that (λ, ν) is ϵ-suboptimal for the dual problem.

Certificate of suboptimality (2/3)

- We refer to the gap between primal and dual objectives,

$$f_0(x) - g(\lambda, \nu),$$

as the duality gap associated with the primal feasible point x and dual feasible point (λ, ν).

- A primal dual feasible pair x, (λ, ν) localizes the optimal value of the primal (and dual) problems to an interval:

$$p^* \in [g(\lambda, \nu), f_0(x)], \quad d^* \in [g(\lambda, \nu), f_0(x)],$$

the width of which is the duality gap.

Certificate of suboptimality (3/3)

- If the duality gap of the primal dual feasible pair x, (λ, ν) is zero, i.e., $f_0(x) = g(\lambda, \nu)$, then x is primal optimal and (λ, ν) is dual optimal.
- We can think of (λ, ν) as a certificate that proves x is optimal.
  - Similarly, we can think of x as a certificate that proves (λ, ν) is dual optimal.
- These observations can be used in optimization algorithms to provide nonheuristic stopping criteria.

Stopping Criteria

- Suppose an algorithm produces a sequence of primal feasible $x^{(k)}$ and dual feasible $(\lambda^{(k)}, \nu^{(k)})$, for $k = 1, 2, \ldots$, and $\epsilon_{abs} > 0$ is a given required absolute accuracy.
- Then the stopping criterion (i.e., the condition for terminating the algorithm) $f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \leq \epsilon_{abs}$ guarantees that when the algorithm terminates, $x^{(k)}$ is ϵ_{abs}-suboptimal.
  - $\because f_0(x^{(k)}) - p^* \leq f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \leq \epsilon_{abs}$.
- Strong duality must hold if this method is to work for arbitrarily small tolerances ϵ_{abs}.

Complementary slackness (1/5)

- Suppose that the primal and dual optimal values are attained and equal (i.e., strong duality holds).
- Let x^* be a primal optimal and (λ^*, ν^*) be a dual optimal point. This means that

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

Complementary slackness (2/5)

- Let's examine the expressions

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

  - The first two equalities state that the optimal duality gap is zero, and the definition of the dual function, respectively.
  - The first inequality follows since the infimum of the Lagrangian over x is less than or equal to its value at $x = x^*$.
  - The last inequality follows from $\lambda_i^* \geq 0, f_i(x^*) \leq 0, i = 1, ..., m$, and $h_i(x^*) = 0, i = 1, ..., p$.

Complementary slackness (3/5)

- So, we conclude that in

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

the two inequalities in this chain hold with equality. We draw the following conclusions:

1. x^* minimizes $L(x, \lambda^*, \nu^*)$ over x. (The Lagrangian $L(x, \lambda^*, \nu^*)$ can have other minimizers; x^* is simply a minimizer.)
2. Another important observation is that

$$
\sum_{i=1}^m \lambda_i^* f_i(x^*) = 0.
$$

Complementary slackness (4/5)

- Now that
$$\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = 0$$
and each term in this sum is nonpositive ($\because \lambda_i \geq 0$, $f_i(x^*) \leq 0$), we conclude that
$$\lambda_i^* f_i(x^*) = 0, i = 1, ..., m.$$

- This condition is known as **complementary slackness**; it holds for any primal optimal x^* and any dual optimal (λ^*, ν^*) (when strong duality holds).

Complementary slackness (5/5)

- We can express the complementary slackness condition as

$$\lambda_i^* > 0 \Longrightarrow f_i(x^*) = 0,$$

or, equivalently,

$$f_i(x^*) < 0 \Longrightarrow \lambda_i^* = 0.$$

- Roughly speaking, this means the ith optimal Lagrange multiplier is zero unless the ith constraint is active at the optimum.

KKT optimality conditions

- We consider an optimization problem, not necessarily convex, with its objective function and constraint functions being differentiable.
- Assume that the functions $f_0, \ldots, f_m, h_1, \ldots, h_p$ are differentiable (and therefore have open domains).
- Let x^* and (λ^*, ν^*) be any primal and dual optimal points with zero duality gap.
- Since x^* minimizes $L(x, \lambda^*, \nu^*)$ over x, it follows that its gradient must vanish at x^*, i.e.,

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

KKT conditions for nonconvex problems

- For any optimization problem with differentiable objective and constraint functions for which strong duality obtains, any pair of primal and dual optimal points (x^*, λ^*, ν^*) must satisfy the following conditions

$$\begin{aligned} f_i(x^*) &\leq 0, i = 1, \dots, m \\ h_i(x^*) &= 0, i = 1, \dots, p \\ \lambda_i^* &\geq 0, i = 1, \dots, m \\ \lambda_i^* f_i(x^*) &= 0, i = 1, \dots, m \\ \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) &= 0. \end{aligned}$$

- These are called the **Karush-Kuhn-Tucker (KKT) conditions**.

KKT Conditions for convex Problems (1/3)

- When the primal problem is convex, the KKT conditions are also sufficient for the points to be primal and dual optimal.
- In other words, if f_i are convex and h_i are affine, and $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ are any points that satisfy the KKT conditions

$$
\begin{aligned}
f_i(\tilde{x}) &\leq 0, i = 1, \dots, m \\
h_i(\tilde{x}) &= 0, i = 1, \dots, p \\
\tilde{\lambda}_i &\geq 0, i = 1, \dots, m \\
\tilde{\lambda}_i f_i(\tilde{x}) &= 0, i = 1, \dots, m \\
\nabla f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{x}) &= 0,
\end{aligned}
$$

then $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ are primal and dual optimal, with zero duality gap.

KKT Conditions for convex Problems (2/3)

Reasons for KKT conditions for convex problems to guarantee primal and dual optimal:

- The first two conditions state that $\tilde{x}$ is primal feasible.
- Since $\tilde{\lambda}_i \geq 0$, $L(x, \tilde{\lambda}, \tilde{\nu})$ is convex in x; the last KKT condition states that its gradient with respect to x vanishes at $x = \tilde{x}$, so it follows that $\tilde{x}$ minimizes $L(x, \tilde{\lambda}, \tilde{\nu})$ over x.
- From this we conclude that

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

where in the last equality we use $h_i(\tilde{x}) = 0$ and $\tilde{\lambda}_i f_i(\tilde{x}) = 0$.

KKT Conditions for convex Problems (3/3)

- This shows that $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ have zero duality gap, and therefore are primal and dual optimal.
- In summary, for any convex optimization problem with differentiable objective and constraint functions, any points that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap.

KKT Optimality Conditions

- If a convex optimization problem with differentiable objective and constraint functions satisfies Slater's condition, then the KKT conditions provide necessary and sufficient conditions for optimality:
  - Slater's condition implies that the optimal duality gap is zero and the dual optimum is attained, so x is optimal if and only if there are (λ, ν) that, together with x, satisfy the KKT conditions.
- In a few special cases it is possible to solve the KKT conditions analytically.
- More generally, many algorithms for convex optimization are conceived as, or can be interpreted as, methods for solving the KKT conditions.

Example

Equality constrained convex quadratic minimization

- We consider the problem

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax = b, \end{array}$$

  where $P \in \mathbf{S}^n_+$.
- The KKT conditions for this problem are

$$Ax^* = b, \quad Px^* + q + A^T \nu^* = 0,$$

  which we can write as

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}.$$

- Solving this set of $m + n$ equations in the $m + n$ variables x^*, ν^* gives the optimal primal and dual variables for the problem.

Example – Water Filling (1/5)

- We consider the convex optimization problem

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{n} \log(\alpha_i + x_i) \\ \text{subject to} & x \succeq 0, \mathbf{1}^T x = 1, \end{array}$$

  where $\alpha_i > 0$, which arises in information theory, in allocating power to a set of n communication channels.
- The variable x_i represents the transmitter power allocated to the ith channel, and $\log(\alpha_i + x_i)$ gives the capacity or communication rate of the channel.
- So, the problem is to allocate a total power of one to the channels, in order to maximize the total communication rate.

Example – Water Filling (2/5)

- Introducing Lagrange multipliers $\lambda^* \in \mathbf{R}^n$ for the inequality constraints $x^* \succeq 0$, and a multiplier $\nu^* \in \mathbf{R}$ for the equality constraint $\mathbf{1}^T x = 1$, we obtain the KKT conditions

$$
\begin{aligned}
x^* &\succeq 0, \\
\mathbf{1}^T x^* &= 1, \\
\lambda^* &\succeq 0, \\
\lambda_i^* x_i^* &= 0, i = 1, ..., n, \\
-1/(\alpha_i + x_i^*) - \lambda_i^* + \nu^* &= 0, i = 1, ..., n.
\end{aligned}
$$

- We can directly solve these equations to find x^*, λ^*, and ν^*.

Example – Water Filling (3/5)

- We start by noting that λ_i^* acts as a slack variable in the last equation, so it can be eliminated, leaving

$$\begin{aligned} x^* &\succeq 0, \\ \mathbf{1}^T x^* &= 1, \\ x_i^*(\nu^* - 1/(\alpha_i + x_i^*)) &= 0, i = 1, \ldots, n, \\ \nu^* &\geq 1/(\alpha_i + x_i^*), i = 1, \ldots, n. \end{aligned}$$

- If $\nu^* < 1/\alpha_i$, this last condition can only hold if $x_i^* > 0$, which by the third condition implies that $\nu^* = 1/(\alpha_i + x_i^*)$. Solving for x_i^*, we conclude that $x_i^* = 1/\nu^* - \alpha_i$ if $\nu^* < 1/\alpha_i$.
- If $\nu^* \geq 1/\alpha_i$, then $x_i^* > 0$ is impossible, because it would imply

$$\nu^* \geq 1/\alpha_i > 1/(\alpha_i + x_i^*),$$

which violates the complementary slackness condition.
- Therefore,

$$x_i^* = 0 \text{ if } \nu^* \geq 1/\alpha_i.$$

Example – Water Filling (4/5)

- Thus we have

$$x_i^* = \begin{cases} 1/\nu^* - \alpha_i, & \nu^* < 1/\alpha_i \\ 0, & \nu^* \geq 1/\alpha_i, \end{cases}$$

  or, put more simply,

$$x_i^* = \max\{0, 1/\nu^* - \alpha_i\} = (1/\nu^* - \alpha_i)_+.$$

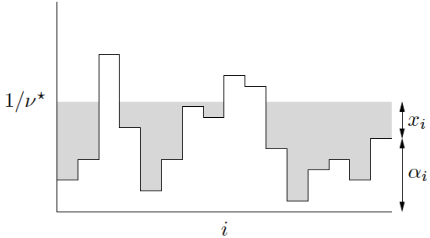
- Substituting this expression for x_i^* into the condition $\mathbf{1}^T x^* = 1$ we obtain

$$\sum_{i=1}^{n} \max\{0, 1/\nu^* - \alpha_i\} = 1.$$

  The left hand side is a piecewise-linear increasing function of $1/\nu^*$, with breakpoints at α_i, so the equation has a unique solution which is readily determined.

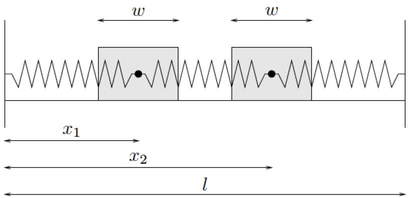
Example – Water Filling (5/5)

- This solution method is called water-filling for the following reason. We think of α_i as the ground level above patch i, and then flood the region with water to a depth $1/\nu$, as illustrated in the figure below.
- The total amount of water used is $\sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\}$.
- We then increase the flood level until we have used a total amount of water equal to one: $\sum_{i=1}^{n} \max\{0, 1/\nu^* - \alpha_i\} = 1$.
- The depth of water above patch i is then the optimal value x_i^* .



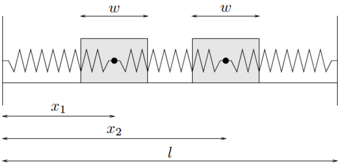
An optimization problem in mechanics (1/3)

- The KKT conditions can be given a nice interpretation in mechanics (which indeed, was one of Lagrange's primary motivations).
- We illustrate the idea with a simple example.



- The system consists of two blocks attached to each other, and to walls at the left and right, by three springs.

An optimization problem in mechanics (2/3)

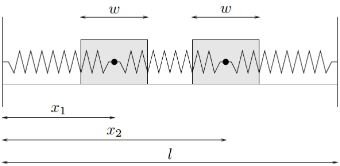


- The position of the blocks are given by $x \in \mathbf{R}^2$, where x_1 is the displacement of the (middle of the) left block, and x_2 is the displacement of the right block.
- The left wall is at position 0, and the right wall is at position l.
- The potential energy in the springs, as a function of the block positions, is given by

$$f_0(x_1, x_2) = \frac{1}{2}k_1 x_1^2 + \frac{1}{2}k_2(x_2 - x_1)^2 + \frac{1}{2}k_3(l - x_2)^2,$$

where $k_i > 0$ are the stiffness constants of the three springs.

An optimization problem in mechanics (3/3)



- The equilibrium position x^* is the position that minimizes the potential energy subject to the inequalities

$$w/2 - x_1 \leq 0, \quad w + x_1 - x_2 \leq 0, \quad w/2 - l + x_2 \leq 0.$$

- These constraints are called kinematic constraints, and express the fact that the blocks have width $w > 0$, and cannot penetrate each other or the walls.

KKT conditions for the mechanics problem (1/2)

- The equilibrium position is therefore given by the solution of the optimization problem

$$\begin{array}{ll} \text{minimize} & (1/2)\left(k_1 x_1^2 + k_2(x_2 - x_1)^2 + k_3(l - x_2)^2\right) \\ \text{subject to} & w/2 - x_1 \leq 0 \\ & w + x_1 - x_2 \leq 0 \\ & w/2 - l + x_2 \leq 0, \end{array}$$

  which is a QP.

- With $\lambda_1, \lambda_2, \lambda_3$ as Lagrange multipliers, what are the KKT conditions for this problem?

KKT conditions for the mechanics problem (2/2)

- The KKT conditions for this problem consist of
  1. the kinematic constraints

$$w/2 - x_1 \le 0, \quad w + x_1 - x_2 \le 0, \quad w/2 - l + x_2 \le 0,$$

  2. the nonnegativity constraints $\lambda_i \ge 0$,
  3. the complementary slackness conditions

$$\lambda_1(w/2 - x_1) = 0, \ \lambda_2(w - x_2 + x_1) = 0, \ \lambda_3(w/2 - l + x_2) = 0,$$

  4. and the zero gradient condition

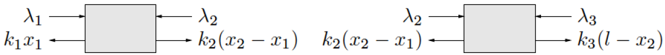
$$\begin{bmatrix} k_1 x_1 - k_2(x_2 - x_1) \\ k_2(x_2 - x_1) - k_3(l - x_2) \end{bmatrix} + \lambda_1 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \lambda_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0.$$

Mechanics interpretation of KKT conditions (1/4)

- The equation

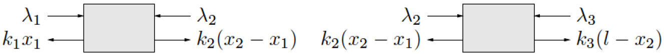
$$\begin{bmatrix} k_1 x_1 - k_2(x_2 - x_1) \\ k_2(x_2 - x_1) - k_3(l - x_2) \end{bmatrix} + \lambda_1 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \lambda_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0$$

can be interpreted as the force balance equations for the two blocks, provided we interpret the Lagrange multipliers as contact forces that act between the walls and blocks, as illustrated in the figure below.



Mechanics interpretation of KKT conditions (2/4)

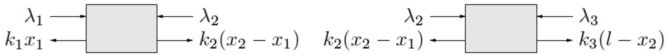
$$\begin{bmatrix} k_1 x_1 - k_2(x_2 - x_1) \\ k_2(x_2 - x_1) - k_3(l - x_2) \end{bmatrix} + \lambda_1 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \lambda_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0$$



- The first equation states that the sum of the forces on the first block is zero:
  1. $-k_1 x_1$: the force exerted on the left block by the left spring.
  2. $k_2(x_2 - x_1)$: the force exerted by the middle spring.
  3. λ_1: the force exerted by the left wall.
  4. $-\lambda_2$: the force exerted by the right block.

Mechanics interpretation of KKT conditions (3/4)

$$\begin{bmatrix} k_1 x_1 - k_2(x_2 - x_1) \\ k_2(x_2 - x_1) - k_3(l - x_2) \end{bmatrix} + \lambda_1 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \lambda_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0$$



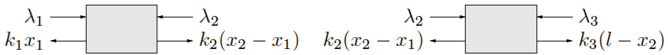
- The contact forces must point away from the contact surface (as expressed by the constraints $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$), and are nonzero only when there is contact (as expressed by the first two complementary slackness conditions $\lambda_1(w/2 - x_1) = 0,\ \lambda_2(w - x_2 + x_1) = 0$).
- In a similar way, the second equation is the force balance for the second block, and the last complementary slackness condition,

$$\lambda_3(w/2 - l + x_2) = 0,$$

states that λ_3 is zero unless the right block touches the wall.

Mechanics interpretation of KKT conditions (4/4)

$$\begin{bmatrix} k_1x_1 - k_2(x_2 - x_1) \\ k_2(x_2 - x_1) - k_3(l - x_2) \end{bmatrix} + \lambda_1 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \lambda_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0$$



- The potential energy and kinematic constraint functions are convex.
- The refined form of Slater's constraint qualification holds provided $2w \leq l$, i.e., there is enough room between the walls to fit the two blocks.
- So we can conclude that the energy formulation of the equilibrium gives the same result as the force balance formulation, given by the KKT conditions.

Solving the primal problem via the dual (1/2)

- In the following, we illustrate cases where the primal problem can be solved by first solving the dual problem.
- We mentioned that if strong duality holds and a dual optimal solution (λ^*, ν^*) exists, then any primal optimal point is also a minimizer of $L(x, \lambda^*, \nu^*)$.
- This fact sometimes allows us to compute a primal optimal solution from a dual optimal solution. More precisely, suppose we have strong duality and an optimal (λ^*, ν^*) is known.
- Suppose that the minimizer of $L(x, \lambda^*, \nu^*)$, i.e., the solution of

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x),$$

is unique.

Solving the primal problem via the dual (2/2)

- Suppose that the minimizer of $L(x, \lambda^*, \nu^*)$, i.e., the solution of

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x),$$

  is unique.
- Then if the solution is primal feasible, it must be primal optimal; if it is not primal feasible, then the primal optimum is not attained.
- This observation is interesting when the dual problem is easier to solve than the primal problem, for example, because it can be solved analytically, or has some special structure that can be exploited.

Example – Entropy Maximization (1/2)

- We consider the entropy maximization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) = \sum_{i=1}^{n} x_i \log x_i \\ \text{subject to} & Ax \preceq b \\ & \mathbf{1}^T x = 1 \end{array}$$

with domain $\mathbf{R}^n_{++}$, and its dual problem

$$\begin{array}{ll} \text{maximize} & -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-a_i^T\lambda} \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

where a_i are the columns of A.
- We assume that the weak form of Slater's condition holds, i.e., there exists an $x \succ 0$ with $Ax \preceq b$ and $\mathbf{1}^T x = 1$.

Example – Entropy Maximization (2/2)

- We assume that the weak form of Slater's condition holds, i.e., there exists an $x \succ 0$ with $Ax \preceq b$ and $\mathbf{1}^T x = 1$, so strong duality holds and an optimal solution (λ^*, ν^*) exists. Suppose we have solved the dual problem.
- The Lagrangian at (λ^*, ν^*) is

$$L(x, \lambda^*, \nu^*) = \sum_{i=1}^{n} x_i \log x_i + {\lambda^*}^T (Ax - b) + \nu^* (\mathbf{1}^T x - 1)$$

which is strictly convex on $\mathcal{D}$ and bounded below, so it has a unique solution x^*, given by

$$x_i^* = 1/\exp(a_i^T \lambda^* + \nu^* + 1), i = 1, ..., n.$$

- If x^* is primal feasible, it must be the optimal solution of the primal problem; otherwise, the primal optimum is not attained.