CME 307 / MS&E 311: Optimization

Convex duality

Professor Udell

Management Science and Engineering
Stanford

February 28, 2024

Announcements

facts:

- CME 307 has a qual (for ICME PhD students), and
- you want more lectures

Announcements

facts:

- CME 307 has a qual (for ICME PhD students), and
- you want more lectures

new plan for course:

1. KKT conditions and IPMs
2. first order methods
3. Bayesian optimization
4. two sessions of project presentations

Announcements

facts:

- CME 307 has a qual (for ICME PhD students), and
- you want more lectures

new plan for course:

1. KKT conditions and IPMs
2. first order methods
3. Bayesian optimization
4. two sessions of project presentations

- Friday sessions will be research paper presentations
- next paper signups will open by this coming Friday.

Fenchel dual

Definition (Fenchel dual)

The **Fenchel dual** of a function $f : \mathcal{X} \to \mathbf{R}$ is

$$f^*(w) = \sup_{x \in \mathcal{X}} \langle w, x \rangle - f(x)$$

also called the **conjugate function**. draw picture!
https://remilepriol.github.io/dualityviz/

Fenchel dual

Definition (Fenchel dual)

The **Fenchel dual** of a function $f : \mathcal{X} \to \mathbf{R}$ is

$$f^*(w) = \sup_{x \in \mathcal{X}} \langle w, x \rangle - f(x)$$

also called the **conjugate function**. draw picture!
https://remilepriol.github.io/dualityviz/

example: $f(x) = \|x\|_1$, $x \in \mathbf{R}^n$

$$f^*(w) = \sup_{x \in \mathbf{R}^n} \langle w, x \rangle - \|x\|_1 = \begin{cases} 0 & \|w\|_\infty \leq 1 \\ \infty & \text{otherwise} \end{cases}$$

$\implies$ fenchel dual of ℓ_1 norm is indicator of ℓ_∞ ball

Biconjugate

Definition (Biconjugate)

The **biconjugate** of a function $f : \mathcal{X} \to \mathbf{R}$ is

$$f^{**}(x) = \sup_{w \in \mathcal{X}^*} \langle w, x \rangle - f^*(w)$$

- for convex $f : \mathbf{R} \to \mathbf{R}$, $f^{**} = f$
- for nonconvex f, f^{**} is convex hull of f

$\Longrightarrow$ biconjugate is a convexification operation

Biconjugate

Definition (Biconjugate)

The **biconjugate** of a function $f : \mathcal{X} \to \mathbf{R}$ is

$$f^{**}(x) = \sup_{w \in \mathcal{X}^*} \langle w, x \rangle - f^*(w)$$

- for convex $f : \mathbf{R} \to \mathbf{R}$, $f^{**} = f$
- for nonconvex f, f^{**} is convex hull of f

$\implies$ biconjugate is a convexification operation

example: consider $f : \mathbf{R} \to \mathbf{R}$ defined by

$$f(x) = \begin{cases} 0 & x \in \{-1, 1\} \\ \infty & \text{otherwise} \end{cases}$$

what is f^*? f^{**}?

Outline

Lagrange duality

Why duality?

- certify optimality
  - turn $\forall$ into $\exists$
  - use dual lower bound to derive stopping conditions
- new algorithms based on the dual
  - solve dual, then recover primal solution

Nonlinear duality

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & Ax = b: & \quad \text{dual } y \\ \text{variable} & x \in \mathbf{R}^n & \end{array} \qquad (\mathcal{P})$$

if x is feasible, then $Ax = b$, so $\langle y, Ax - b \rangle = 0$.

Nonlinear duality

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & Ax = b : & \quad \text{dual } y \\ \text{variable} & x \in \mathbf{R}^n & \end{array} \qquad (\mathcal{P})$$

if x is feasible, then $Ax = b$, so $\langle y, Ax - b\rangle = 0$.
define the **Lagrangian**

$$\mathcal{L}(x, y) \quad := \quad f(x) - \langle y, b - Ax\rangle$$

Nonlinear duality

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & Ax = b : & \qquad \text{dual } y \\ \text{variable} & x \in \mathbf{R}^n & \end{array} \qquad (\mathcal{P})$$

if x is feasible, then $Ax = b$, so $\langle y, Ax - b \rangle = 0$.
define the **Lagrangian**

$$\begin{aligned} \mathcal{L}(x, y) &:= f(x) - \langle y, b - Ax \rangle \\ p^\star &= \inf_{x: Ax = b} \mathcal{L}(x, y) \geq \inf_x \mathcal{L}(x, y) \end{aligned}$$

Nonlinear duality

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & Ax = b: & \quad \text{dual } y \\ \text{variable} & x \in \mathbf{R}^n & \end{array} \qquad (\mathcal{P})$$

if x is feasible, then $Ax = b$, so $\langle y, Ax - b\rangle = 0$.
define the **Lagrangian**

$$\begin{aligned}
\mathcal{L}(x,y) &:= f(x) - \langle y, b - Ax\rangle \\
p^\star &= \inf_{x:Ax=b} \mathcal{L}(x,y) \geq \inf_x \mathcal{L}(x,y) \\
&= \inf_x f(x) + \langle y, -b + Ax\rangle \\
&= \langle y, -b\rangle + \inf_x \left(f(x) + \langle A^T y, x\rangle\right) \\
&= \langle y, -b\rangle - \sup_x \left(\langle -A^T y, x\rangle - f(x)\right) \\
&= \langle y, -b\rangle - f^*(-A^T y) = g(y)
\end{aligned}$$

Lagrange duality

inequality holds for any $y \in \mathbf{R}^m$, so we have proved **weak duality**

$$\begin{aligned} p^\star &\geq g(y) \quad \forall y \in \mathbf{R}^m \\ &\geq \underbrace{\sup_y g(y)}_{\mathcal{D}} =: d^\star \end{aligned} \tag{1}$$

dual optimal value $d^\star \leq p^\star$

Strong duality

Definition (Duality gap)

The **duality gap** for a primal-dual pair (x, y) is $f(x) - g(y)$

by weak duality, duality gap is always nonnegative

Strong duality

Definition (Duality gap)

The **duality gap** for a primal-dual pair (x, y) is $f(x) - g(y)$

by weak duality, duality gap is always nonnegative

Definition (Strong duality)

A primal-dual pair $(x^\star, y^\star)$ satisfies **strong duality** if

$$p^\star = d^\star \iff f(x^\star) - g(y^\star) = 0$$

Strong duality

Definition (Duality gap)

The **duality gap** for a primal-dual pair (x, y) is $f(x) - g(y)$

by weak duality, duality gap is always nonnegative

Definition (Strong duality)

A primal-dual pair $(x^\star, y^\star)$ satisfies **strong duality** if

$$p^\star = d^\star \iff f(x^\star) - g(y^\star) = 0$$

strong duality holds

- for feasible LPs
- for convex problems under **constraint qualification** aka **Slater's condition**. feasible region has an **interior point** x so that all inequality constraints hold strictly

strong duality fails if either primal or dual problem is infeasible or unbounded

Lagrange duality with inequality constraints

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \leq b : \quad y \geq 0 \\ \text{variable} & x \in \mathbf{R}^n \end{array} \qquad (\mathcal{P})$$

Lagrange duality with inequality constraints

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \leq b : \quad y \geq 0 \\ \text{variable} & x \in \mathbf{R}^n \end{array} \tag{$\mathcal{P}$}$$

to construct Lagrangian $\mathcal{L}(x, y) = f(x) - \langle y, b - Ax \rangle$, ensure value is **better** (lower) when x and y are feasible

Lagrange duality with inequality constraints

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \leq b : \quad y \geq 0 \\ \text{variable} & x \in \mathbf{R}^n \end{array} \qquad (\mathcal{P})$$

to construct Lagrangian $\mathcal{L}(x,y) = f(x) - \langle y, b - Ax\rangle$, ensure value is **better** (lower) when x and y are feasible

$$\mathcal{L}(x,y) \quad := \quad f(x) - \langle y, b - Ax\rangle$$

Lagrange duality with inequality constraints

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \leq b : \quad y \geq 0 \\ \text{variable} & x \in \mathbf{R}^n \end{array} \tag{$\mathcal{P}$}$$

to construct Lagrangian $\mathcal{L}(x, y) = f(x) - \langle y, b - Ax \rangle$, ensure value is **better** (lower) when x and y are feasible

$$\begin{aligned} \mathcal{L}(x, y) &:= f(x) - \langle y, b - Ax \rangle \\ p^\star &\geq \inf_{x \text{ feas}} f(x) - \langle y, b - Ax \rangle \\ &\geq \inf_x f(x) - \langle y, b - Ax \rangle \\ &= \langle y, -b \rangle - f^*(-A^* y) =: g(y) \end{aligned}$$

Lagrange duality with inequality constraints

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \leq b : \quad y \geq 0 \\ \text{variable} & x \in \mathbf{R}^n \end{array} \tag{$\mathcal{P}$}$$

to construct Lagrangian $\mathcal{L}(x, y) = f(x) - \langle y, b - Ax \rangle$, ensure value is **better** (lower) when x and y are feasible

$$\begin{aligned} \mathcal{L}(x, y) &:= f(x) - \langle y, b - Ax \rangle \\ p^\star &\geq \inf_{x \text{ feas}} f(x) - \langle y, b - Ax \rangle \\ &\geq \inf_x f(x) - \langle y, b - Ax \rangle \\ &= \langle y, -b \rangle - f^*(-A^* y) =: g(y) \end{aligned}$$

this holds for all $y \geq 0$, so we have weak duality

$$p^\star \geq \sup_y g(y) =: d^\star$$

SVM dual

support vector machine: for $x_i \in \mathbf{R}^n$, $y_i \in \{-1, 1\}$, $i = 1, \ldots, m$

$$\begin{array}{lll} \text{minimize} & \frac{1}{2}\|w\|^2 + 1^T s & \\ \text{subject to} & y_i w^T x_i + s_i \geq 1 \quad i = 1, \ldots, m : & \alpha \geq 0 \\ & s \geq 0 : & \mu \geq 0 \end{array} \qquad \text{(SVM)}$$

SVM dual

support vector machine: for $x_i \in \mathbf{R}^n$, $y_i \in \{-1, 1\}$, $i = 1, \ldots, m$

$$\begin{array}{lll} \text{minimize} & \frac{1}{2}\|w\|^2 + 1^T s & \\ \text{subject to} & y_i w^T x_i + s_i \geq 1 \quad i = 1, \ldots, m : & \alpha \geq 0 \\ & s \geq 0 : & \mu \geq 0 \end{array} \tag{SVM}$$

verify Slater's condition. strong duality holds!

SVM dual

support vector machine: for $x_i \in \mathbf{R}^n$, $y_i \in \{-1, 1\}$, $i = 1, \ldots, m$

$$
\begin{array}{lll}
\text{minimize} & \frac{1}{2}\|w\|^2 + 1^T s & \\
\text{subject to} & y_i w^T x_i + s_i \geq 1 \quad i = 1, \ldots, m: & \alpha \geq 0 \\
& s \geq 0: & \mu \geq 0
\end{array}
\qquad \text{(SVM)}
$$

verify Slater's condition. strong duality holds! Lagrangian: for $\alpha \geq 0$, $\mu \geq 0$,

$$
\mathcal{L}(w, s, \alpha, \mu) = \frac{1}{2}\|w\|^2 + 1^T s - \sum_{i=1}^{m} \alpha_i (y_i w^T x_i + s_i - 1) - \mu^T s
$$

- minimize $\mathcal{L}(w, s, \alpha, \mu)$ over w:

$$
w = \sum_{i=1}^{m} \alpha_i y_i x_i
$$

- minimize $\mathcal{L}(w, s, \alpha, \mu)$ over $s \implies \alpha + \mu = 1$

SVM dual

so simplify:

$$\begin{aligned} g(\alpha) &= \inf_{w,s} \mathcal{L}(w, s, \alpha, 1-\alpha) \\ &= \frac{1}{2}\|w\|^2 - w^T \sum_{i=1}^m \alpha_i y_i x_i + 1^T \alpha \\ &= -\frac{1}{2}\|\sum_{i=1}^m \alpha_i y_i x_i\|^2 + 1^T \alpha \end{aligned}$$

SVM dual

so simplify:

$$\begin{aligned} g(\alpha) &= \inf_{w,s} \mathcal{L}(w, s, \alpha, 1-\alpha) \\ &= \frac{1}{2}\|w\|^2 - w^T \sum_{i=1}^{m} \alpha_i y_i x_i + 1^T \alpha \\ &= -\frac{1}{2}\| \sum_{i=1}^{m} \alpha_i y_i x_i \|^2 + 1^T \alpha \end{aligned}$$

define $K \in \mathbf{R}^m$ so $K_{ij} = y_i y_j x_i^T x_j$. then

$$\| \sum_{i=1}^{m} \alpha_i y_i x_i \|^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j x_i^T x_j = \alpha^T K \alpha$$

SVM dual

so simplify:

$$\begin{aligned} g(\alpha) &= \inf_{w,s} \mathcal{L}(w, s, \alpha, 1-\alpha) \\ &= \frac{1}{2}\|w\|^2 - w^T \sum_{i=1}^m \alpha_i y_i x_i + 1^T \alpha \\ &= -\frac{1}{2}\|\sum_{i=1}^m \alpha_i y_i x_i\|^2 + 1^T \alpha \end{aligned}$$

define $K \in \mathbf{R}^m$ so $K_{ij} = y_i y_j x_i^T x_j$. then

$$\|\sum_{i=1}^m \alpha_i y_i x_i\|^2 = \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j = \alpha^T K \alpha$$

dual problem:

$$\begin{array}{ll} \text{maximize} & -\frac{1}{2}\alpha^T K \alpha + 1^T \alpha \\ \text{subject to} & \alpha \geq 0 \end{array} \qquad \text{(SVM-dual)}$$

SVM dual

so simplify:

$$\begin{aligned} g(\alpha) &= \inf_{w,s} \mathcal{L}(w, s, \alpha, 1-\alpha) \\ &= \frac{1}{2}\|w\|^2 - w^T \sum_{i=1}^m \alpha_i y_i x_i + 1^T \alpha \\ &= -\frac{1}{2}\|\sum_{i=1}^m \alpha_i y_i x_i\|^2 + 1^T \alpha \end{aligned}$$

define $K \in \mathbf{R}^m$ so $K_{ij} = y_i y_j x_i^T x_j$. then

$$\|\sum_{i=1}^m \alpha_i y_i x_i\|^2 = \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j = \alpha^T K \alpha$$

dual problem:

$$\begin{array}{ll} \text{maximize} & -\frac{1}{2}\alpha^T K \alpha + 1^T \alpha \\ \text{subject to} & \alpha \geq 0 \end{array} \qquad \text{(SVM-dual)}$$

new solution ideas! proj grad, coord descent (SMO), kernel trick

Generalize Lagrangian duality

Generalize Lagrangian duality

- **nonlinear duality:** replace

$$0 \geq Ax - b \quad \text{with} \quad 0 \geq g(x)$$

(harder to derive explicit form for dual problem)

Generalize Lagrangian duality

- **nonlinear duality:** replace

$$0 \geq Ax - b \quad \text{with} \quad 0 \geq g(x)$$

  (harder to derive explicit form for dual problem)

- **conic duality:** for cone K, replace

$$b - Ax \geq 0 \quad \text{with} \quad b - Ax \in K$$

  define **slack vector** $s = b - Ax \in K$
  for weak duality, dual y must satisfy

$$\langle y, s \rangle \geq 0 \quad \forall s \in K$$

Dual cones

this inequality defines the **dual cone** K^*:

Definition (dual cone)

the dual cone K^* of a cone K is the set of vectors y such that

$$\langle y, s\rangle \geq 0 \quad \forall s \in K$$

Dual cones

this inequality defines the **dual cone** K^*:

Definition (dual cone)

the dual cone K^* of a cone K is the set of vectors y such that

$$\langle y, s \rangle \geq 0 \quad \forall s \in K$$

examples of cones and their duals:

- K acute, K^* obtuse
- $K = \mathbf{R}^m_+$, $K^* = \mathbf{R}^m_+$
- $K = \{x \in \mathbf{R}^n \mid \|x\| \leq x_0\}$, $K^* = \{y \in \mathbf{R}^n \mid \|y\| \leq y_0\}$
- $K = \{X \in \mathbf{S}^n \mid X \succeq 0\}$, $K^* = \{Y \in \mathbf{S}^n \mid Y \succeq 0\}$

inner product $\langle X, Y \rangle = \mathbf{tr}(X^T Y) = \sum_{ij} X_{ij} Y_{ij}$ for $X, Y \in \mathbf{S}^n$

Conic duality

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{ll} \text{minimize} & \langle c, x\rangle \\ \text{subject to} & b - Ax \in K : \quad y \in K^* \\ \text{variable} & x \in \mathbf{R}^n \end{array} \qquad (\mathcal{P})$$

Conic duality

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{ll} \text{minimize} & \langle c, x\rangle \\ \text{subject to} & b - Ax \in K : \quad y \in K^* \\ \text{variable} & x \in \mathbf{R}^n \end{array} \qquad (\mathcal{P})$$

for $y \in K^*$, construct Lagrangian $\mathcal{L}(x, y) = \langle c, x\rangle - \langle y, b - Ax\rangle$, ensure value is **better** (lower) when x and y are feasible

Conic duality

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{ll} \text{minimize} & \langle c, x \rangle \\ \text{subject to} & b - Ax \in K : \quad y \in K^* \\ \text{variable} & x \in \mathbf{R}^n \end{array} \qquad (\mathcal{P})$$

for $y \in K^*$, construct Lagrangian $\mathcal{L}(x, y) = \langle c, x \rangle - \langle y, b - Ax \rangle$, ensure value is **better** (lower) when x and y are feasible

$$\mathcal{L}(x, y) \quad := \quad \langle c, x \rangle - \langle y, b - Ax \rangle$$

Conic duality

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{ll} \text{minimize} & \langle c, x\rangle \\ \text{subject to} & b - Ax \in K : \quad y \in K^* \\ \text{variable} & x \in \mathbf{R}^n \end{array} \tag{$\mathcal{P}$}$$

for $y \in K^*$, construct Lagrangian $\mathcal{L}(x, y) = \langle c, x\rangle - \langle y, b - Ax\rangle$, ensure value is **better** (lower) when x and y are feasible

$$\begin{array}{rcl} \mathcal{L}(x, y) & := & \langle c, x\rangle - \langle y, b - Ax\rangle \\ p^\star & \geq & \inf_{x \text{ feas}} \langle c, x\rangle - \langle y, b - Ax\rangle \\ & \geq & \inf_x \langle c, x\rangle - \langle y, b - Ax\rangle \\ & = & \langle y, -b\rangle + \inf_x \langle c + A^* y, x\rangle \end{array}$$

Conic duality

primal problem with solution $x^\star \in \mathbf{R}^n$, optimal value $p^\star$:

$$\begin{array}{ll} \text{minimize} & \langle c, x\rangle \\ \text{subject to} & b - Ax \in K : \quad y \in K^* \\ \text{variable} & x \in \mathbf{R}^n \end{array} \qquad (\mathcal{P})$$

for $y \in K^*$, construct Lagrangian $\mathcal{L}(x, y) = \langle c, x\rangle - \langle y, b - Ax\rangle$, ensure value is **better** (lower) when x and y are feasible

$$\begin{aligned} \mathcal{L}(x, y) &:= \langle c, x\rangle - \langle y, b - Ax\rangle \\ p^\star &\geq \inf_{x \text{ feas}} \langle c, x\rangle - \langle y, b - Ax\rangle \\ &\geq \inf_x \langle c, x\rangle - \langle y, b - Ax\rangle \\ &= \langle y, -b\rangle + \inf_x \langle c + A^* y, x\rangle \end{aligned}$$

which is $-\infty$ unless $c + A^* y = 0$, so ...

Conic duality

define the **dual problem**

$$\begin{array}{ll} \text{maximize} & \langle y, -b \rangle \\ \text{subject to} & c + A^* y = 0 \\ \text{variable} & y \in K^* \end{array} \qquad (\mathcal{D})$$

Conic duality

define the **dual problem**

$$\begin{array}{ll} \text{maximize} & \langle y, -b \rangle \\ \text{subject to} & c + A^* y = 0 \\ \text{variable} & y \in K^* \end{array} \qquad (\mathcal{D})$$

again, we have weak duality $p^\star \geq d^\star$ and,
(under a constraint qualification) strong duality $p^\star = d^\star$

Conic duality

define the **dual problem**

$$\begin{array}{ll} \text{maximize} & \langle y, -b \rangle \\ \text{subject to} & c + A^* y = 0 \\ \text{variable} & y \in K^* \end{array} \qquad (\mathcal{D})$$

again, we have weak duality $p^\star \geq d^\star$ and,
(under a constraint qualification) strong duality $p^\star = d^\star$

- if $(\mathcal{P})$ is convex, then the dual of (1) is $(\mathcal{P})$
- otherwise, the dual of the dual is the **convexification** of the primal

Strong duality and complementary slackness

Definition (complementary slackness)

The primal-dual pair x and y are **complementary** if

$$\langle y, b - Ax \rangle = 0$$

They satisfy **strict complementary slackness** if $y_i(b_i - a_i^T x) = 0$ for $i = 1, \ldots, n$.

for conic problem, strong duality $\iff$ complementary slackness

$$\begin{aligned}\langle y, s \rangle &= \langle y, b - Ax \rangle \\ &= \langle y, b \rangle - \langle A^* y, x \rangle \\ &= \langle y, b \rangle - \langle c, x \rangle\end{aligned}$$

First-order optimality condition

The KKT conditions are first-order **necessary** conditions for optimality of optimization problem.

Theorem (KKT conditions)

Suppose $x^\star$ and $y^\star$ are primal and dual optimal, respectively. Then

- **stationarity.** *$x^\star$ minimizes the Lagrangian at $y^\star$. If $\mathcal{L}$ is differentiable, then*

$$\nabla_x \mathcal{L}(x^\star, y^\star) = 0.$$

- **feasibility.** *$x^\star$ is primal feasible; $y^\star$ is dual feasible.*
- **complementary slackness.** *dual variable $y_i^\star$ is nonzero only if the ith constraint is active at $x^\star$.*

- KKT conditions are named after Karush, Kuhn, and Tucker.
- KKT conditions turn optimization problem into a system of equations.
- If the problem is convex, then the KKT conditions are also **sufficient** for optimality.

KKT conditions: example

nonlinear optimization with inequality constraints:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \leq b : \quad y \geq 0 \end{array}$$

Lagrangian $\mathcal{L}(x, y) = f(x) - \langle y, Ax - b \rangle$.

Suppose $x^\star$ and $y^\star$ are primal and dual optimal, respectively. Then

- **stationarity.** $x^\star$ minimizes the Lagrangian at $y^\star$:

$$\nabla_x \mathcal{L}(x^\star, y^\star) = 0 \implies \nabla f(x^\star) = A^T y^\star$$

- **feasibility.** $Ax^\star \leq b$ is primal feasible; $y^\star \geq 0$ is dual feasible.
- **complementary slackness.** dual variable $y_i^\star$ is nonzero only if the ith constraint is active at $x^\star$:

$$\langle y^\star, b - Ax^\star \rangle = 0$$

KKT Example

Consider the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & x^2 + y^2 \\ \text{subject to} \quad & x + y \leq -1 : \quad \lambda \geq 0 \\ & x - y = 0 : \quad \mu \end{aligned}$$

Lagrangian:

KKT Example

Consider the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & x^2 + y^2 \\ \text{subject to} \quad & x + y \leq -1 : \quad \lambda \geq 0 \\ & x - y = 0 : \quad \mu \end{aligned}$$

Lagrangian:

$$\mathcal{L}(x, y, \lambda, \mu) = x^2 + y^2 + \lambda(x + y + 1) + \mu(x - y)$$

Lagrangian:

$$\mathcal{L}(x, y, \lambda, \mu) = x^2 + y^2 + \lambda(x + y + 1) + \mu(x - y)$$

Lagrangian:
$$\mathcal{L}(x, y, \lambda, \mu) = x^2 + y^2 + \lambda(x + y + 1) + \mu(x - y)$$

KKT conditions:

1. stationarity: $\nabla_x L(x, y, \lambda, \mu) = 0$, $\nabla_y L(x, y, \lambda, \mu) = 0$, ie,
$$\frac{\partial \mathcal{L}}{\partial x} = 2x + \lambda + \mu = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2y + \lambda - \mu = 0$$
2. feasibility:
   - primal: $x + y \leq -1$ and $x - y = 0$
   - dual: $\lambda \geq 0$
3. complementary slackness: $\lambda = 0$ or $x + y = -1$ (or both)

Lagrangian:

$$\mathcal{L}(x, y, \lambda, \mu) = x^2 + y^2 + \lambda(x + y + 1) + \mu(x - y)$$

KKT conditions:

1. stationarity: $\nabla_x L(x, y, \lambda, \mu) = 0$, $\nabla_y L(x, y, \lambda, \mu) = 0$, ie,

$$\frac{\partial \mathcal{L}}{\partial x} = 2x + \lambda + \mu = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2y + \lambda - \mu = 0$$

2. feasibility:
   - primal: $x + y \leq -1$ and $x - y = 0$
   - dual: $\lambda \geq 0$
3. complementary slackness: $\lambda = 0$ or $x + y = -1$ (or both)

solve!

Lagrangian:
$$\mathcal{L}(x, y, \lambda, \mu) = x^2 + y^2 + \lambda(x + y + 1) + \mu(x - y)$$

KKT conditions:

1. stationarity: $\nabla_x L(x, y, \lambda, \mu) = 0$, $\nabla_y L(x, y, \lambda, \mu) = 0$, ie,
$$\frac{\partial \mathcal{L}}{\partial x} = 2x + \lambda + \mu = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2y + \lambda - \mu = 0$$
2. feasibility:
   - primal: $x + y \leq -1$ and $x - y = 0$
   - dual: $\lambda \geq 0$
3. complementary slackness: $\lambda = 0$ or $x + y = -1$ (or both)

solve!

- primal feasibility (PF) $\implies x = y$

Lagrangian:
$$\mathcal{L}(x, y, \lambda, \mu) = x^2 + y^2 + \lambda(x + y + 1) + \mu(x - y)$$

KKT conditions:

1. stationarity: $\nabla_x L(x, y, \lambda, \mu) = 0$, $\nabla_y L(x, y, \lambda, \mu) = 0$, ie,
$$\frac{\partial \mathcal{L}}{\partial x} = 2x + \lambda + \mu = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2y + \lambda - \mu = 0$$
2. feasibility:
   - primal: $x + y \leq -1$ and $x - y = 0$
   - dual: $\lambda \geq 0$
3. complementary slackness: $\lambda = 0$ or $x + y = -1$ (or both)

solve!

- primal feasibility (PF) $\implies$ $x = y$
- if $\lambda^\star = 0$, stationarity $\implies$ $\lambda^\star + \mu^\star = 1$ and $\lambda^\star - \mu^\star = 1$. impossible!

Lagrangian:

$$\mathcal{L}(x, y, \lambda, \mu) = x^2 + y^2 + \lambda(x + y + 1) + \mu(x - y)$$

KKT conditions:

1. stationarity: $\nabla_x L(x, y, \lambda, \mu) = 0$, $\nabla_y L(x, y, \lambda, \mu) = 0$, ie,

$$\frac{\partial \mathcal{L}}{\partial x} = 2x + \lambda + \mu = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2y + \lambda - \mu = 0$$

2. feasibility:
    - primal: $x + y \leq -1$ and $x - y = 0$
    - dual: $\lambda \geq 0$
3. complementary slackness: $\lambda = 0$ or $x + y = -1$ (or both)

solve!

- primal feasibility (PF) $\implies x = y$
- if $\lambda^\star = 0$, stationarity $\implies \lambda^\star + \mu^\star = 1$ and $\lambda^\star - \mu^\star = 1$. impossible!
- if $\lambda^\star \neq 0$, PF + CS requires $x = y = -\frac{1}{2}$.

Lagrangian:

$$\mathcal{L}(x, y, \lambda, \mu) = x^2 + y^2 + \lambda(x + y + 1) + \mu(x - y)$$

KKT conditions:

1. stationarity: $\nabla_x L(x, y, \lambda, \mu) = 0$, $\nabla_y L(x, y, \lambda, \mu) = 0$, ie,

$$\frac{\partial \mathcal{L}}{\partial x} = 2x + \lambda + \mu = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2y + \lambda - \mu = 0$$

2. feasibility:
   - primal: $x + y \leq -1$ and $x - y = 0$
   - dual: $\lambda \geq 0$
3. complementary slackness: $\lambda = 0$ or $x + y = -1$ (or both)

solve!

- primal feasibility (PF) $\implies$ $x = y$
- if $\lambda^\star = 0$, stationarity $\implies$ $\lambda^\star + \mu^\star = 1$ and $\lambda^\star - \mu^\star = 1$. impossible!
- if $\lambda^\star \neq 0$, PF + CS requires $x = y = -\frac{1}{2}$.
- so use stationarity to solve for optimal dual: $\lambda^\star = \frac{1}{2}$, $\mu^\star = 0$

Summary

- Duality provides lower bounds on the optimal value of an optimization problem.
- Construct the Lagrangian for any optimization problem by
  1. adding a linear combination of the constraints to the objective,
  2. restricting the associated dual variables to ensure Lagrangian provides a lower bound when primal is feasible.
- Duality can be used to certify optimality or as a stopping condition.
- KKT conditions give necessary (and for convex problems, sufficient) conditions for optimality,
  - ...and hence new ways to solve the problem by solving the KKT system.
  - Solving KKT conditions reduces to a linear system for problems with equality constraints,
  - but more complex for problems with inequality (or conic) constraints.