CME 307 / MS&E 311: Optimization

Duality

Professor Udell

Management Science and Engineering
Stanford

May 3, 2023

Announcements

- meet with course staff to discuss project this week or next (see Ed)
- project 1 due this Friday 5/5

Outline

Duality

Lagrange duality

Duality

Definition (Dual space)

The **dual** $\mathcal{X}^*$ of a vector space $\mathcal{X}$ is the set of linear functionals on $\mathcal{X}$.

so if $x \in \mathcal{X}$ and you see someone write

$$w^T x, \qquad \langle w, x \rangle, \qquad \text{or} \qquad w \cdot x$$

you know that $w \in \mathcal{X}^*$ is a dual vector

Duality

Definition (Dual space)

The **dual** $\mathcal{X}^*$ of a vector space $\mathcal{X}$ is the set of linear functionals on $\mathcal{X}$.

so if $x \in \mathcal{X}$ and you see someone write

$$w^T x, \qquad \langle w, x \rangle, \qquad \text{or} \qquad w \cdot x$$

you know that $w \in \mathcal{X}^*$ is a dual vector

notation: solution to optimization problem $x^\star$ vs dual space $\mathcal{X}^*$

Careful of units!

example 1: suppose $y_i = w^T x_i$ where

$$x_i = \begin{bmatrix} \text{heart rate} \\ \text{blood pressure} \\ \text{age} \end{bmatrix}, \quad \text{with units} \quad \begin{bmatrix} \text{bpm} \\ \text{mmHg} \\ \text{years} \end{bmatrix}$$

and y_i is duration of stay in hospital (units: days)

Careful of units!

example 1: suppose $y_i = w^T x_i$ where

$$x_i = \begin{bmatrix} \text{heart rate} \\ \text{blood pressure} \\ \text{age} \end{bmatrix}, \quad \text{with units} \quad \begin{bmatrix} \text{bpm} \\ \text{mmHg} \\ \text{years} \end{bmatrix}$$

and y_i is duration of stay in hospital (units: days)

then w has units of

$$\begin{bmatrix} \text{days/bpm} \\ \text{days/mmHg} \\ \text{days/year} \end{bmatrix}$$

Careful of units!

example 2: $f(x) = \sum_{i=1}^{n} \left(1 + \exp\left(\underbrace{y_i w^T x_i}_{\text{input must be a scalar!}} \right) \right)$

Careful of units!

example 2: $f(x) = \sum_{i=1}^{n} \left(1 + \exp\left(\underbrace{y_i w^T x_i}_{\text{input must be a scalar!}} \right) \right)$

example 3: if $x \in \mathcal{X}$, gradient is a linear function on $\mathcal{X} \implies \nabla f(x_0) \in \mathcal{X}^*$

$$f(x) \approx f(x_0) + \nabla f(x_0)^T (x - x_0),$$

so gradient descent stepsize t has units

$$x^{k+1} = x^k - t \nabla f(x^k)$$

e.g., x (meters m), $\nabla f(x)$ (m^{-1}), and t (m^2)

Careful of units!

example 2: $f(x) = \sum_{i=1}^{n} \left(1 + \exp\left(\underbrace{y_i w^T x_i}_{\text{input must be a scalar!}}\right)\right)$

example 3: if $x \in \mathcal{X}$, gradient is a linear function on $\mathcal{X} \implies \nabla f(x_0) \in \mathcal{X}^*$

$$f(x) \approx f(x_0) + \nabla f(x_0)^T (x - x_0),$$

so gradient descent stepsize t has units

$$x^{k+1} = x^k - t \nabla f(x^k)$$

e.g., x (meters m), $\nabla f(x)$ (m^{-1}), and t (m^2)

- no wonder it's hard to choose the stepsize!
- basic recommendation: standardize your data

Dual of function space

- $f : [0,1] \to \mathbf{R}$ is a function
- $f(x)$ is a linear function of f, for any x:

$$(f+g)(x) = f(x) + g(x), \qquad (cf)(x) = cf(x)$$

- so is any integral:

$$\int_0^1 f(x) d\mu(x)$$

$\implies$ the dual of the space of functions on $[0,1]$ is the space of measures on $[0,1]$

Dual norm

Definition (Dual norm)

The **dual norm** of a norm $\|\cdot\|$ is

$$\|w\|_* = \sup_{\|x\| \le 1} \langle w, x \rangle$$

equivalently, $\|w\|_* = \sup_x \frac{\langle w,x \rangle}{\|x\|}$

Dual norm

Definition (Dual norm)

The **dual norm** of a norm $\|\cdot\|$ is

$$\|w\|_* = \sup_{\|x\|\leq 1} \langle w, x \rangle$$

equivalently, $\|w\|_* = \sup_x \frac{\langle w,x \rangle}{\|x\|}$

example: ℓ_1 norm dual is ℓ_∞ norm

$$\|w\|_1 = \sum_{i=1}^{n} |w_i|, \qquad \|w\|_\infty = \max_{i=1,\ldots,n} |w_i|$$

Dual norm

Definition (Dual norm)

The **dual norm** of a norm $\|\cdot\|$ is

$$\|w\|_* = \sup_{\|x\|\leq 1} \langle w, x \rangle$$

equivalently, $\|w\|_* = \sup_x \frac{\langle w,x \rangle}{\|x\|}$

example: ℓ_1 norm dual is ℓ_∞ norm

$$\|w\|_1 = \sum_{i=1}^{n} |w_i|, \qquad \|w\|_\infty = \max_{i=1,\dots,n} |w_i|$$

example: ℓ_2 norm dual is ℓ_2 norm $\implies$ ℓ_2 is **self-dual**

Dual norm

Definition (Dual norm)

The **dual norm** of a norm $\|\cdot\|$ is

$$\|w\|_* = \sup_{\|x\|\leq 1} \langle w, x\rangle$$

equivalently, $\|w\|_* = \sup_x \frac{\langle w,x\rangle}{\|x\|}$

example: ℓ_1 norm dual is ℓ_∞ norm

$$\|w\|_1 = \sum_{i=1}^{n} |w_i|, \qquad \|w\|_\infty = \max_{i=1,\dots,n} |w_i|$$

example: ℓ_2 norm dual is ℓ_2 norm $\implies$ ℓ_2 is **self-dual**

example: for $f : [0,1] \to \mathbf{R}$, if $\|f\| = \sup_{x\in[0,1]} |f(x)|$,

$$\|\mu\|_* = \sup_{\|f\|\leq 1} \int_0^1 f(x) d\mu(x) = \int_0^1 d|\mu|(x)$$

Self-dual norms

given primal space $\mathcal{X}$

- dual vector is a linear functional $w(x)$ on $x \in \mathcal{X}$
- we should define the dual norm on $\mathcal{X}^*$ as

$$\sup_{x \in \mathcal{X}, \|x\| \leq 1} w(x)$$

- but instead we used the inner product $\langle w, x \rangle$. why?

Self-dual norms

given primal space $\mathcal{X}$

- dual vector is a linear functional $w(x)$ on $x \in \mathcal{X}$
- we should define the dual norm on $\mathcal{X}^*$ as

$$\sup_{x \in \mathcal{X}, \|x\| \leq 1} w(x)$$

- but instead we used the inner product $\langle w, x \rangle$. why?

Theorem (Riesz representation)

Suppose $\mathcal{X} = H$ is a Hilbert (inner product) space. For any linear functional $\phi \in \mathcal{X}^*$, there is a unique vector $w \in H$ so that $w(x) = \langle w, x \rangle$ for all $x \in \mathcal{X} = H$. Moreover, $\|w\|_* = \|w\|$.

$\|\cdot\|$ is self-dual $\iff$ $\|\cdot\|$ is induced by an inner product

Self-dual norms

given primal space $\mathcal{X}$

- dual vector is a linear functional $w(x)$ on $x \in \mathcal{X}$
- we should define the dual norm on $\mathcal{X}^*$ as

$$\sup_{x \in \mathcal{X}, \|x\| \leq 1} w(x)$$

- but instead we used the inner product $\langle w, x \rangle$. why?

Theorem (Riesz representation)

Suppose $\mathcal{X} = H$ is a Hilbert (inner product) space. For any linear functional $\phi \in \mathcal{X}^*$, there is a unique vector $w \in H$ so that $w(x) = \langle w, x \rangle$ for all $x \in \mathcal{X} = H$. Moreover, $\|w\|_* = \|w\|$.

$\|\cdot\|$ is self-dual $\iff$ $\|\cdot\|$ is induced by an inner product

example: ℓ_2 norm is self-dual, induced by the inner product

$$\langle w, x \rangle = w^T x$$

Conjugate of linear operator

given $x \in \mathbf{R}^n$, $w \in \mathbf{R}^m$, and $A \in \mathbf{R}^{m \times n}$, conjugate of A is the linear operator A^* defined so that

$$\langle A^* w, x \rangle = \langle w, Ax \rangle$$

Conjugate of linear operator

given $x \in \mathbf{R}^n$, $w \in \mathbf{R}^m$, and $A \in \mathbf{R}^{m \times n}$, conjugate of A is the linear operator A^* defined so that

$$\langle A^* w, x \rangle = \langle w, Ax \rangle$$

example: $x \in \mathbf{R}^n$, $A \in \mathbf{R}^{m \times n}$ defined by

$$Ax = \begin{bmatrix} x_{i_1} \\ \vdots \\ x_{i_m} \end{bmatrix}$$

then $A^* \in \mathbf{R}^{n \times m}$ satisfies

$$\langle A^* w, x \rangle = \langle w, Ax \rangle = \sum_{j=1}^{m} w_j x_{i_j},$$

so A^* creates a sparse vector from w with

$$(A^* w)_{i_j} = w_j$$

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

$\Longrightarrow$ 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

Duality

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

Warmup: Farkas lemma

Theorem (Farkas lemma)

Given $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$, exactly one of the following is true:

- *there exists $x \in \mathbf{R}^n$ so that $Ax = b$ and $x \geq 0$*
- *there exists $y \in \mathbf{R}^m$ so that $A^T y \geq 0$ and $\langle b, y \rangle < 0$*

$\Longrightarrow$ can efficiently certify infeasibility of a linear program

Warmup: Farkas lemma

Theorem (Farkas lemma)

Given $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$, exactly one of the following is true:

- *there exists $x \in \mathbf{R}^n$ so that $Ax = b$ and $x \geq 0$*
- *there exists $y \in \mathbf{R}^m$ so that $A^T y \geq 0$ and $\langle b, y \rangle < 0$*

$\implies$ can efficiently certify infeasibility of a linear program

proof: suppose we have $x \in \mathbf{R}^n$ so that $Ax = b$ and $x \geq 0$. then for any $y \in \mathbf{R}^m$,

$$\begin{aligned} 0 &= \langle y, b - Ax \rangle = \langle y, b \rangle - \langle A^T y, x \rangle \\ \langle y, b \rangle &= \langle A^T y, x \rangle \end{aligned}$$

so if $A^T y \geq 0$, then use $x \geq 0$ to conclude $\langle y, b \rangle \geq 0$.

Warmup: Farkas lemma

Theorem (Farkas lemma)

Given $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$, exactly one of the following is true:

- *there exists $x \in \mathbf{R}^n$ so that $Ax = b$ and $x \geq 0$*
- *there exists $y \in \mathbf{R}^m$ so that $A^T y \geq 0$ and $\langle b, y \rangle < 0$*

$\implies$ can efficiently certify infeasibility of a linear program

proof: suppose we have $x \in \mathbf{R}^n$ so that $Ax = b$ and $x \geq 0$.
then for any $y \in \mathbf{R}^m$,

$$\begin{aligned} 0 &= \langle y, b - Ax \rangle = \langle y, b \rangle - \langle A^T y, x \rangle \\ \langle y, b \rangle &= \langle A^T y, x \rangle \end{aligned}$$

so if $A^T y \geq 0$, then use $x \geq 0$ to conclude $\langle y, b \rangle \geq 0$.

(opposite direction is similar)

Lagrange duality

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

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

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

Lagrange 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**

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

Lagrange 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**

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

Lagrange 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**

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

$g(y)$ is called the **dual function**

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 (pf later)
- 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 $(\mathcal{P})$ or (1) 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 $p^\star$:

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

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

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 $p^\star$:

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

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

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

Lagrange duality with inequality constraints

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

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

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

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

as before, this holds for all y, so we have weak duality

$$p^\star \geq \underbrace{\sup_y g(y)}_{\mathcal{D}} =: d^\star$$

SVM dual

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

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

SVM dual

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

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

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! coordinate descent on α (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
- $K = \mathbf{R}^m_+$, $K^* = \mathbf{R}^m_+$
- $K = \{x \in \mathbf{R}^n \mid \|x\| \leq x_0\}$, $K^* = \{y \in \mathbf{R}^n \mid \|y\| \leq y_0\}$
- $K = \{X \in \mathbf{S}^n \mid X \succeq 0\}$, $K^* = \{Y \in \mathbf{S}^n \mid Y \succeq 0\}$

Conic duality

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

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

Conic duality

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

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

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 $p^\star$:

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

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 $p^\star$:

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

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

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

Conic duality

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

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

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

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

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

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

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

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

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

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

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

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

which is $-\infty$ unless $c + A^* y = 0$, so define the **dual problem**

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

Dual of the dual

- 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

picture

Strong duality for LPs

primal and dual LP in standard form:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & A^T y \leq c \end{array}$$

claim: if primal LP has a bounded feasible solution $x^\star$, then strong duality holds
i.e., dual LP has a bounded feasible solution $y^\star$ and $p^\star = d^\star$

Proof of strong duality for LPs

consider the following system with variables $x' \in \mathbf{R}^n$, $\tau \in \mathbf{R}$

$$Ax' - b\tau = 0, \quad c^T x' = p^\star \tau - 1, \quad (x', \tau) \geq 0$$

Proof of strong duality for LPs

consider the following system with variables $x' \in \mathbf{R}^n$, $\tau \in \mathbf{R}$

$$Ax' - b\tau = 0, \quad c^T x' = p^\star \tau - 1, \quad (x', \tau) \geq 0$$

claim: this system has no solution. pf by contradiction:

- if $\tau > 0$, then x'/τ is feasible for LP and $c^T x'/\tau < p^\star$
- if $\tau = 0$, then $x^\star + x'$ is feasible for LP and $c^T(x^\star + x') < p^\star$

Proof of strong duality for LPs

consider the following system with variables $x' \in \mathbf{R}^n$, $\tau \in \mathbf{R}$

$$Ax' - b\tau = 0, \quad c^T x' = p^\star \tau - 1, \quad (x', \tau) \geq 0$$

claim: this system has no solution. pf by contradiction:

- if $\tau > 0$, then x'/τ is feasible for LP and $c^T x'/\tau < p^\star$
- if $\tau = 0$, then $x^\star + x'$ is feasible for LP and $c^T(x^\star + x') < p^\star$

so use Farkas' lemma:

$$Ax + b = 0, \; x \geq 0 \qquad \text{or} \qquad A^T y \geq 0, \quad b^T y < 0$$

Proof of strong duality for LPs

consider the following system with variables $x' \in \mathbf{R}^n$, $\tau \in \mathbf{R}$

$$Ax' - b\tau = 0, \quad c^T x' = p^\star \tau - 1, \quad (x', \tau) \geq 0$$

claim: this system has no solution. pf by contradiction:

- if $\tau > 0$, then x'/τ is feasible for LP and $c^T x'/\tau < p^\star$
- if $\tau = 0$, then $x^\star + x'$ is feasible for LP and $c^T(x^\star + x') < p^\star$

so use Farkas' lemma:

$$Ax + b = 0, \; x \geq 0 \qquad \text{or} \qquad A^T y \geq 0, \quad b^T y < 0$$

$$\begin{bmatrix} A & -b \\ c^T & -p^\star \end{bmatrix} \begin{bmatrix} x \\ \tau \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \qquad \text{or} \qquad \begin{bmatrix} A^T & c \\ -b^T & -p^\star \end{bmatrix} \begin{bmatrix} y \\ \sigma \end{bmatrix} \geq 0, \; \sigma > 0$$

Proof of strong duality for LPs

consider the following system with variables $x' \in \mathbf{R}^n$, $\tau \in \mathbf{R}$

$$Ax' - b\tau = 0, \quad c^T x' = p^\star \tau - 1, \quad (x', \tau) \geq 0$$

claim: this system has no solution. pf by contradiction:

- if $\tau > 0$, then x'/τ is feasible for LP and $c^T x'/\tau < p^\star$
- if $\tau = 0$, then $x^\star + x'$ is feasible for LP and $c^T(x^\star + x') < p^\star$

so use Farkas' lemma:

$$Ax + b = 0, \;\; x \geq 0 \qquad \text{or} \qquad A^T y \geq 0, \quad b^T y < 0$$

$$\begin{bmatrix} A & -b \\ c^T & -p^\star \end{bmatrix} \begin{bmatrix} x \\ \tau \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \qquad \text{or} \qquad \begin{bmatrix} A^T & c \\ -b^T & -p^\star \end{bmatrix} \begin{bmatrix} y \\ \sigma \end{bmatrix} \geq 0, \;\; \sigma > 0$$

use second system to show y/σ is dual feasible and optimal

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

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

KKT conditions

KKT conditions give **necessary** conditions for optimality of convex problem.

Theorem (KKT conditions)

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

- **stationarity.** *$x^\star$ and $y^\star$ are a min/max saddle point of the Lagrangian*

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

- **feasibility.** *$x^\star$ is primal feasible; $y^\star$ is dual feasible*
- **complementary slackness.** *$x^\star$ and $y^\star$ are complementary:*

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

KKT conditions turn optimization problem into a system of equations

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

Lagrangian:

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

KKT conditions:

1. stationarity: $\nabla L(x, y, \lambda, \mu) = 0$
2. feasibility:
   - primal: $x + y \leq 1$ and $x - y = 0$
   - dual: $\lambda \geq 0$
3. complementary slackness: $\lambda(x + y - 1) = 0$

Taking the gradient of L wrt x, y, λ, and μ, we get:

$$\frac{\partial \mathcal{L}}{\partial x} = 2x + \lambda + \mu = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2y + \lambda - \mu = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = x + y - 1 = 0$$
$$\frac{\partial \mathcal{L}}{\partial \mu} = x - y = 0$$

solve!

Lagrangian:

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

KKT conditions:

1. stationarity: $\nabla L(x, y, \lambda, \mu) = 0$
2. feasibility:
   - primal: $x + y \leq 1$ and $x - y = 0$
   - dual: $\lambda \geq 0$
3. complementary slackness: $\lambda(x + y - 1) = 0$

Taking the gradient of L wrt x, y, λ, and μ, we get:

$$\frac{\partial \mathcal{L}}{\partial x} = 2x + \lambda + \mu = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2y + \lambda - \mu = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = x + y - 1 = 0$$
$$\frac{\partial \mathcal{L}}{\partial \mu} = x - y = 0$$

solve! $\rightarrow x^* = 0.5,\ y^* = 0.5,\ \lambda^* = 0,\ \mu^* = 1$

Semidefinite program (SDP)

consider primal SDP with decision variable $X \in \mathbf{S}^n$:

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CX) \\ \text{subject to} & \mathcal{A}X = b \\ & X \succeq 0, \end{array} \qquad (\mathcal{P})$$

problem data:

- cost matrix $C \in \mathbf{S}^n$
- linear map $\mathcal{A} : \mathbf{R}^{n \times n} \to \mathbf{R}^m$
- righthand side $b \in \mathbf{R}^m$

First order methods use too much storage

suppose $(\mathcal{P})$ has

- **compact specification**: problem data use $\mathcal{O}(n)$ storage
- **compact solution**: solution $X_\star$ has constant rank $r_\star$
  $\Longrightarrow$ solution uses $\mathcal{O}(n)$ storage

First order methods use too much storage

suppose $(\mathcal{P})$ has

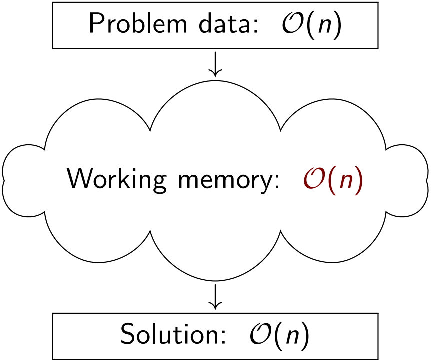
- **compact specification**: problem data use $\mathcal{O}(n)$ storage
- **compact solution**: solution $X_\star$ has constant rank $r_\star$
  $\Longrightarrow$ solution uses $\mathcal{O}(n)$ storage

$(\mathcal{P})$, using any first order method:

$$\boxed{\text{Problem data: } \mathcal{O}(n)}$$
$$\downarrow$$
$$\boxed{\text{Working memory: } \mathcal{O}(n^2)}$$
$$\downarrow$$
$$\boxed{\text{Solution: } \mathcal{O}(n)}$$

Approximate complementarity requires less storage

($\mathcal{P}$), using approximate complementarity:



Motivation: large scale SDP

Max-Cut [goemans1995improved]	Matrix Completion [srebro2005rank]
$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(-LX) \\ \text{subject to} & \mathbf{diag}(X) = \mathbf{1} \\ & X \succeq 0 \end{array}$$	$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(W_1) + \mathbf{tr}(W_2) \\ \text{subject to} & X_{ij} = M_{ij},\ (i,j) \in \Omega \\ & \begin{bmatrix} W_1 & X \\ X^* & W_2 \end{bmatrix} \succeq 0 \end{array}$$

Motivation: large scale SDP

Max-Cut [goemans1995improved]	Matrix Completion [srebro2005rank]
$\begin{array}{ll} \text{minimize} & \mathbf{tr}(-LX) \\ \text{subject to} & \mathbf{diag}(X) = \mathbf{1} \\ & X \succeq 0 \end{array}$	$\begin{array}{ll} \text{minimize} & \mathbf{tr}(W_1) + \mathbf{tr}(W_2) \\ \text{subject to} & X_{ij} = M_{ij},\ (i,j) \in \Omega \\ & \begin{bmatrix} W_1 & X \\ X^* & W_2 \end{bmatrix} \succeq 0 \end{array}$

- Matrix completion:
  - 10^9 users, 10^9 products
  - $\implies$ SDP with 10^{18} variables

Motivation: large scale SDP

Max-Cut [goemans1995improved]	**Matrix Completion** [srebro2005rank]
$\begin{array}{ll} \text{minimize} & \mathbf{tr}(-LX) \\ \text{subject to} & \mathbf{diag}(X) = \mathbf{1} \\ & X \succeq 0 \end{array}$	$\begin{array}{ll} \text{minimize} & \mathbf{tr}(W_1) + \mathbf{tr}(W_2) \\ \text{subject to} & X_{ij} = M_{ij},\ (i,j) \in \Omega \\ & \begin{bmatrix} W_1 & X \\ X^* & W_2 \end{bmatrix} \succeq 0 \end{array}$

- Matrix completion:
  - 10^9 users, 10^9 products
  - $\implies$ SDP with 10^{18} variables
- MaxCut:
  - 10^9 people in social network
  - $\implies$ SDP with 10^{18} variables

Motivation: large scale SDP

Max-Cut [goemans1995improved]	**Matrix Completion** [srebro2005rank]
$\begin{array}{ll} \text{minimize} & \mathbf{tr}(-LX) \\ \text{subject to} & \mathbf{diag}(X) = \mathbf{1} \\ & X \succeq 0 \end{array}$	$\begin{array}{ll} \text{minimize} & \mathbf{tr}(W_1) + \mathbf{tr}(W_2) \\ \text{subject to} & X_{ij} = M_{ij}, \ (i,j) \in \Omega \\ & \begin{bmatrix} W_1 & X \\ X^* & W_2 \end{bmatrix} \succeq 0 \end{array}$

- Matrix completion:
  - 10^9 users, 10^9 products
  - $\implies$ SDP with 10^{18} variables
- MaxCut:
  - 10^9 people in social network
  - $\implies$ SDP with 10^{18} variables
- Phase retrieval:
  - $10^3 \times 10^3$ discretization of sample
  - $\implies$ SDP with 10^{12} variables

Optimal Storage

What kind of storage bounds can we hope for?

- Assume black-box implementation of

$$u \mapsto Cu, \qquad (u, v) \mapsto \mathcal{A}(uv^*), \text{and} \qquad (u, y) \mapsto (\mathcal{A}^* y)u,$$

  where u, $v \in \mathbf{R}^n$, $y \in \mathbf{R}^m$.

Optimal Storage

What kind of storage bounds can we hope for?

- Assume black-box implementation of

$$u \mapsto Cu, \qquad (u, v) \mapsto \mathcal{A}(uv^*), \text{and} \qquad (u, y) \mapsto (\mathcal{A}^* y)u,$$

  where u, $v \in \mathbf{R}^n$, $y \in \mathbf{R}^m$.
- Need $\Omega(m + n)$ storage to apply problem data.

Optimal Storage

What kind of storage bounds can we hope for?

- Assume black-box implementation of

$$u \mapsto Cu, \qquad (u, v) \mapsto \mathcal{A}(uv^*), \text{and} \qquad (u, y) \mapsto (\mathcal{A}^* y)u,$$

  where u, $v \in \mathbf{R}^n$, $y \in \mathbf{R}^m$.
- Need $\Omega(m + n)$ storage to apply problem data.
- Need $\Theta(rn)$ storage for a rank-r approximate solution.

Optimal Storage

What kind of storage bounds can we hope for?

- Assume black-box implementation of

  $$u \mapsto Cu, \qquad (u, v) \mapsto \mathcal{A}(uv^*), \text{and} \qquad (u, y) \mapsto (\mathcal{A}^* y)u,$$

  where u, $v \in \mathbf{R}^n$, $y \in \mathbf{R}^m$.
- Need $\Omega(m+n)$ storage to apply problem data.
- Need $\Theta(rn)$ storage for a rank-r approximate solution.

Definition. An algorithm to return a rank r (approximate) solution to $(\mathcal{P})$ has **optimal storage** if it uses working storage

$$\Theta(m + rn).$$

Dual problem

consider dual SDP with decision variable $y \in \mathbf{R}^m$:

$$\begin{array}{ll} \text{maximize} & b^* y \\ \text{subject to} & C - \mathcal{A}^* y \succeq 0 \end{array} \qquad (\mathcal{D})$$

Dual problem

consider dual SDP with decision variable $y \in \mathbf{R}^m$:

$$\begin{array}{ll} \text{maximize} & b^* y \\ \text{subject to} & C - \mathcal{A}^* y \succeq 0 \end{array} \qquad (\mathcal{D})$$

for sufficiently large $\alpha > 0$, equivalent to **penalized dual**

$$\text{maximize} \quad b^* y + \alpha \min(\lambda_{\min}(C - \mathcal{A}^* y), 0)$$

Dual problem

consider dual SDP with decision variable $y \in \mathbf{R}^m$:

$$\begin{array}{ll} \text{maximize} & b^* y \\ \text{subject to} & C - \mathcal{A}^* y \succeq 0 \end{array} \qquad (\mathcal{D})$$

for sufficiently large $\alpha > 0$, equivalent to **penalized dual**

$$\text{maximize} \quad b^* y + \alpha \min(\lambda_{\min}(C - \mathcal{A}^* y), 0)$$

- any first order method for $(\mathcal{D})$ uses optimal storage
- (just compute minimal eigenvalue with iterative method)
- *e.g.*, subgradient method, AdaGrad [**duchi2011adaptive**], AdaNGD [**levy2017online**], AccelGrad [**levy2018online**], ...

Assumptions

Assumption (Genericity)

- primal SDP attains its solution(s)
- dual SDP has a unique solution $y_\star$
- primal and dual SDP satisfy strong duality

$$0 = \mathbf{tr}(CX_\star) - b^* y_\star = \mathbf{tr}(X_\star(C - \mathcal{A}^* y_\star)) = X_\star(C - \mathcal{A}^* y_\star)$$

- (for storage optimality) strict complementary slackness

$$\mathrm{Rank}(X_\star) + \mathrm{Rank}(C - \mathcal{A}^* y_\star) = n$$

$\implies$ solution $X_\star$ of primal SDP is unique [**lemon2016low**]

Assumptions

Assumption (Genericity)

- primal SDP attains its solution(s)
- dual SDP has a unique solution $y_\star$
- primal and dual SDP satisfy strong duality

$$0 = \mathbf{tr}(CX_\star) - b^* y_\star = \mathbf{tr}(X_\star(C - \mathcal{A}^* y_\star)) = X_\star(C - \mathcal{A}^* y_\star)$$

- (for storage optimality) strict complementary slackness

$$\mathrm{Rank}(X_\star) + \mathrm{Rank}(C - \mathcal{A}^* y_\star) = n$$

$\implies$ solution $X_\star$ of primal SDP is unique [**lemon2016low**]

Note: these conditions hold generically [**alizadeh1997complementarity**]

Primal recovery via exact complementarity

suppose $y_\star \in \mathbf{R}^m$ solves $(\mathcal{D})$

Primal recovery via exact complementarity

suppose $y_\star \in \mathbf{R}^m$ solves $(\mathcal{D})$

- define dual slack matrix $Z_\star = C - \mathcal{A}^* y_\star$

Primal recovery via exact complementarity

suppose $y_\star \in \mathbf{R}^m$ solves $(\mathcal{D})$

- define dual slack matrix $Z_\star = C - \mathcal{A}^* y_\star$
- strict complementarity holds between $X_\star$ and $Z_\star$

$$\mathrm{Rank}(X_\star) + \mathrm{Rank}(Z_\star) = n \quad \Longrightarrow \quad \mathrm{range}(X_\star) = \textbf{nullspace}(Z_\star)$$

Primal recovery via exact complementarity

suppose $y_\star \in \mathbf{R}^m$ solves $(\mathcal{D})$

- define dual slack matrix $Z_\star = C - \mathcal{A}^* y_\star$
- strict complementarity holds between $X_\star$ and $Z_\star$

$$\mathrm{Rank}(X_\star) + \mathrm{Rank}(Z_\star) = n \quad \Longrightarrow \quad \mathrm{range}(X_\star) = \textbf{nullspace}(Z_\star)$$

- let $V_\star$ be a basis for nullspace of dual slack matrix $Z_\star$

Primal recovery via exact complementarity

suppose $y_\star \in \mathbf{R}^m$ solves $(\mathcal{D})$

- define dual slack matrix $Z_\star = C - \mathcal{A}^* y_\star$
- strict complementarity holds between $X_\star$ and $Z_\star$

$$\mathrm{Rank}(X_\star) + \mathrm{Rank}(Z_\star) = n \quad \Longrightarrow \quad \mathrm{range}(X_\star) = \textbf{nullspace}(Z_\star)$$

- let $V_\star$ be a basis for nullspace of dual slack matrix $Z_\star$
- constrain $X = V_\star S V_\star^*$ in primal SDP for some $S \in \mathbf{S}_\star^{r_\star}$.
  $\Longrightarrow$ solution is preserved!

SDP via exact complementarity: algorithm

Algorithm SDP via exact complementarity

Given: problem data C, $\mathcal{A}$, b

SDP via exact complementarity: algorithm

Algorithm SDP via exact complementarity

Given: problem data C, $\mathcal{A}$, b

1. compute solution $y_\star \in \mathbf{R}^m$ to $(\mathcal{D})$

SDP via exact complementarity: algorithm

Algorithm SDP via exact complementarity

Given: problem data C, $\mathcal{A}$, b

1. compute solution $y_\star \in \mathbf{R}^m$ to $(\mathcal{D})$
2. compute basis $V_\star$ for nullspace of dual slack matrix $Z_\star = C - \mathcal{A}^* y_\star$

SDP via exact complementarity: algorithm

Algorithm SDP via exact complementarity

Given: problem data C, $\mathcal{A}$, b

1. compute solution $y_\star \in \mathbf{R}^m$ to $(\mathcal{D})$
2. compute basis $V_\star$ for nullspace of dual slack matrix $Z_\star = C - \mathcal{A}^* y_\star$
3. solve **reduced primal** $(\mathcal{P}_{V_\star})$ with variable $S \in \mathbf{S}_+^{r_\star}$

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CV_\star S V_\star^*) \\ \text{subject to} & \mathcal{A}(V_\star S V_\star^*) = b \\ & S \succeq 0 \end{array} \qquad (\mathcal{P}_{V_\star})$$

to find primal solution $X_\star = V_\star S V_\star^*$

SDP via exact complementarity: algorithm

Algorithm SDP via exact complementarity

Given: problem data C, $\mathcal{A}$, b

1. compute solution $y_\star \in \mathbf{R}^m$ to $(\mathcal{D})$
2. compute basis $V_\star$ for nullspace of dual slack matrix $Z_\star = C - \mathcal{A}^* y_\star$
3. solve **reduced primal** $(\mathcal{P}_{V_\star})$ with variable $S \in \mathbf{S}_+^{r_\star}$

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(C V_\star S V_\star^*) \\ \text{subject to} & \mathcal{A}(V_\star S V_\star^*) = b \\ & S \succeq 0 \end{array} \qquad (\mathcal{P}_{V_\star})$$

to find primal solution $X_\star = V_\star S V_\star^*$

Reduced primal is easy: *e.g.*, when $r_\star = 1$, $(\mathcal{P}_{V_\star})$ solves a 1D problem over $S \in \mathbf{R}_+$