Convex Optimization
Primer: Lecture II

Vasileios Nakos

July 27, 2021

Convex Programs

Definition

Given a convex set $K \subseteq \mathbb{R}^n$ and a convex $f : K \to \mathbb{R}$, a convex program is the following optimization problem

$$\inf_{x \in K} f(x).$$

Convex Programs

Definition

Given a convex set $K \subseteq \mathbb{R}^n$ and a convex $f : K \to \mathbb{R}$, a convex program is the following optimization problem

$$\inf_{x \in K} f(x).$$

- unconstrained when $K = \mathbb{R}^n$

Convex Programs

Definition

Given a convex set $K \subseteq \mathbb{R}^n$ and a convex $f : K \to \mathbb{R}$, a convex program is the following optimization problem

$$\inf_{x \in K} f(x).$$

- unconstrained when $K = \mathbb{R}^n$
- smooth when f is differentiable with a continuous derivative

Convex Programs

Definition

Given a convex set $K \subseteq \mathbb{R}^n$ and a convex $f : K \to \mathbb{R}$, a convex program is the following optimization problem

$$\inf_{x \in K} f(x).$$

- unconstrained when $K = \mathbb{R}^n$
- smooth when f is differentiable with a continuous derivative
- nonsmooth otherwise.

Consider

$$\inf\{\frac{1}{x} : x \in (0, \infty)\}.$$

Consider

$$\inf\{\frac{1}{x} : x \in (0, \infty)\}.$$

No $x \in K$ attains the infimum.
If $K \subseteq \mathbb{R}^n$ is closed and bounded then the minimum is attained by some $x \in K$.

Some examples of convex programs

Linear Regression. $\min_{x \in \mathbb{R}^n} \|Ax - b\|_2$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^n$.

Some examples of convex programs

Linear Regression. $\min_{x \in \mathbb{R}^n} \|Ax - b\|_2$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^n$. In that case

$$f(x) = \|Ax - b\|_2^2 = x^T A^T A x - 2b^T A x + b^T b,$$

and $\nabla^2 f(x) = 2A^T A \gtrsim 0$.

Some examples of convex programs

Linear Regression. $\min_{x\in\mathbb{R}^n}\|Ax-b\|_2$, where $A\in\mathbb{R}^{m\times n}$ and $b\in\mathbb{R}^n$. In that case

$$f(x)=\|Ax-b\|_2^2=x^TA^TAx-2b^TAx+b^Tb,$$

and $\nabla^2 f(x)=2A^TA\gtrsim 0$.

Linear programming.

$$\min\ c^Tx \quad \text{s.t.} \quad Ax\leq b$$

Computational models

Oracle Model. Often we allow oracle access to $f, \nabla f, \nabla^2 f$ and bound the number of oracle calls or iterations that the (usually iterative) algorithm performs.

word RAM model. Addition,subtraction, multiplication etc take exactly 1 time step, for numbers that can be stored in a word; usually the input of the problem shall consist of numbers that can fit in a word.

Membership for convex sets

Given a point $x \in \mathbb{R}^n$ and $K \subseteq \mathbb{R}^n$, does $x \in K$?

Membership for convex sets

Given a point $x \in \mathbb{R}^n$ and $K \subseteq \mathbb{R}^n$, does $x \in K$?

$\star$ Halfspaces: Let $K := \{y \in \mathbb{R}^n : \langle a, y \rangle \leq b\}$ where $a \in \mathbb{R}^n, b \in \mathbb{R}$. We need to write down a, b, x using finite number of bits to perform membership in K.

Membership for convex sets

Given a point $x \in \mathbb{R}^n$ and $K \subseteq \mathbb{R}^n$, does $x \in K$?

⋆ Halfspaces: Let $K := \{y \in \mathbb{R}^n : \langle a, y \rangle \leq b\}$ where $a \in \mathbb{R}^n, b \in \mathbb{R}$. We need to write down a, b, x using finite number of bits to perform membership in K.

⋆ Ellipsoids. Let $K := \{y \in \mathbb{R}^n : y^T A y \leq 1\}$ for a PD matrix $A \in \mathbb{Q}^{n \times n}$.

Membership for convex sets

Given a point $x \in \mathbb{R}^n$ and $K \subseteq \mathbb{R}^n$, does $x \in K$?

⋆ Halfspaces: Let $K := \{y \in \mathbb{R}^n : \langle a, y \rangle \leq b\}$ where $a \in \mathbb{R}^n, b \in \mathbb{R}$. We need to write down a, b, x using finite number of bits to perform membership in K.

⋆ Ellipsoids. Let $K := \{y \in \mathbb{R}^n : y^T A y \leq 1\}$ for a PD matrix $A \in \mathbb{Q}^{n \times n}$.

⋆ Intersection of halfspaces (polytopes). $K := \{\langle a_i, y \rangle \leq b_i, i = 1, \ldots, m\}$.

Membership for convex sets

$\star$ ℓ_1 ball: $K := \{x \in \mathbb{R}^n : \sum_{i=1}^n |x|_i \leq r\}$.

Membership for convex sets

$\star$ ℓ_1 ball: $K := \{x \in \mathbb{R}^n : \sum_{i=1}^n |x|_i \leq r\}$. It is an intersection of 2^n hyperplanes, of all $\{y : \langle y, s \rangle\}, s \in \{-, 1+1\}^n$. No hyperplane is redundant.

Membership for convex sets

⋆ ℓ_1 ball: $K := \{x \in \mathbb{R}^n : \sum_{i=1}^n |x|_i \le r\}$. It is an intersection of 2^n hyperplanes, of all $\{y : \langle y, s\rangle\}, s \in \{-, 1+1\}^n$. No hyperplane is redundant.

⋆ PSD matrices. Given $X \in \mathbb{R}^n$, does $y^T X y \ge 0$ for all $y \in \mathbb{R}^n$? Equivalent to checking whether $\lambda_1(X) \ge 0$. Can only approximately check.

Membership for convex sets

$\star$ ℓ_1 ball: $K := \{x \in \mathbb{R}^n : \sum_{i=1}^n |x|_i \leq r\}$. It is an intersection of 2^n hyperplanes, of all $\{y : \langle y, s \rangle\}, s \in \{-, 1+1\}^n$. No hyperplane is redundant.

$\star$ PSD matrices. Given $X \in \mathbb{R}^n$, does $y^T X y \geq 0$ for all $y \in \mathbb{R}^n$? Equivalent to checking whether $\lambda_1(X) \geq 0$. Can only approximately check.

Some operations that preserve convexity:

- Intersection
- Scaling
- Translation
- Affine transformation
- Set sum

Separation Oracles for convex sets

Theorem (Convexity implies Separating Hyperplane)

For all closed and convex $K \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n \setminus K$ there exists $a \in \mathbb{R}^n, b \in \mathbb{R}$ such that

$$\langle a, x \rangle > b \quad \text{and} \quad \langle a, y \rangle \leq b, \forall y \in K.$$

Separation Oracles for convex sets

Theorem (Convexity implies Separating Hyperplane)

For all closed and convex $K \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n \setminus K$ there exists $a \in \mathbb{R}^n, b \in \mathbb{R}$ such that

$$\langle a, x \rangle > b \quad \text{and} \quad \langle a, y \rangle \leq b, \forall y \in K.$$

Theorem (Separating hyperplanes implies convexity)

Let $K \subseteq \mathbb{R}^n$ be a convex set. If for every $x \in \mathbb{R}^n \setminus K$ there exists a hyperplane separating x from K, then K is convex.

Separation Oracle

A separation oracle for a convex set $K \subseteq \mathbb{R}^n$ is a primitive which:

1. given $x \in K$, answers YES

Separation Oracle

A separation oracle for a convex set $K \subseteq \mathbb{R}^n$ is a primitive which:

1. given $x \in K$, answers YES
2. given $x \notin K$, answers NO and returns $a \in \mathbb{Q}^n, b \in \mathbb{Q}$ such that the hyperplane $\{y : \langle a, y \rangle = b\}$ separates x from K.

Separation Oracle

A separation oracle for a convex set $K \subseteq \mathbb{R}^n$ is a primitive which:

1. given $x \in K$, answers YES
2. given $x \notin K$, answers NO and returns $a \in \mathbb{Q}^n, b \in \mathbb{Q}$ such that the hyperplane $\{y : \langle a, y \rangle = b\}$ separates x from K.

Separation vs. optimization. Constructing efficient (polynomial time) separation oracles for a given family of convex sets is equivalent to constructing algorithms to optimize linear functions over convex sets in this family.

Back to solving convex programs

Given $c \in \mathbb{Q}$, find whether $\min_{x \in K} f(x) = c$.

Back to solving convex programs

Given $c \in \mathbb{Q}$, find whether $\min_{x \in K} f(x) = c$.

Consider $f(x) = \frac{2}{x} + x, K = [1, \infty)$. What happens?

Back to solving convex programs

Given $c \in \mathbb{Q}$, find whether $\min_{x \in K} f(x) = c$.

Consider $f(x) = \frac{2}{x} + x, K = [1, \infty)$. What happens?

Refined goal: Given $\epsilon > 0$ compute $c \in \mathbb{Q}$ such that

$$\min_{x \in K} f(x) \in [c - \epsilon, c + \epsilon].$$

Distance to optimum versus distance to value

Consider a convex program which has a unique optimal solution $x^\star \in K$. Then we can ask for either

1. proximity in value, $f(x) \leq f(x^*) + \epsilon$, or
2. proximity in optimum, $\|x - x^*\|_2 \leq \epsilon$.

Representing Functions

- Linear and affine. $f(x) = \langle a, x \rangle + b$.

Representing Functions

- Linear and affine. $f(x) = \langle a, x \rangle + b$.
- Quadratic. $f(x) = x^T A x + \langle b, x \rangle + c$ for PSD matrix $A \in \mathbb{Q}^n$.

Representing Functions

- Linear and affine. $f(x) = \langle a, x \rangle + b$.
- Quadratic. $f(x) = x^T Ax + \langle b, x \rangle + c$ for PSD matrix $A \in \mathbb{Q}^n$.
- Linear matrix functions. $f(X) = Tr(XA)$, where $A \in \mathbb{Q}^{n \times n}$ and $X \in \mathbb{R}^{n \times n}$ is a symmetric matrix variable.

Representing Functions

- Linear and affine. $f(x) = \langle a, x \rangle + b$.
- Quadratic. $f(x) = x^T Ax + \langle b, x \rangle + c$ for PSD matrix $A \in \mathbb{Q}^n$.
- Linear matrix functions. $f(X) = Tr(XA)$, where $A \in \mathbb{Q}^{n \times n}$ and $X \in \mathbb{R}^{n \times n}$ is a symmetric matrix variable.

Models of accessing f (again)

Value Oracle: Given $x \in K$, compute $f(x)$.

Models of accessing f (again)

Value Oracle: Given $x \in K$, compute $f(x)$.
Gradient Oracle: Given $x \in K$, compute $\nabla f(x), \nabla^2 f(x), \nabla^3 f(x), \ldots$.

Models of accessing f (again)

Value Oracle: Given $x \in K$, compute $f(x)$.
Gradient Oracle: Given $x \in K$, compute $\nabla f(x), \nabla^2 f(x), \nabla^3 f(x), \ldots$.
Measure number of oracle calls, ideally $\mathrm{poly}(n, \log(1/\epsilon))$.

Examples of how (iterative) algorithms look like

Gradient descent. $x_{t+1} := x_t - \eta \nabla f(x_t)$.

Examples of how (iterative) algorithms look like

Gradient descent. $x_{t+1} := x_t - \eta \nabla f(x_t)$.

Examples of how (iterative) algorithms look like

Gradient descent. $x_{t+1} := x_t - \eta \nabla f(x_t)$.

Projected Gradient Descent. $x_{t+1} := \Pi_K(x_t - \eta \nabla f(x_t))$.

Newton's method. $x_{t+1} := x_t - (\nabla^2 f(x_t))^{-1} \nabla f(x_t)$.

Examples of how (iterative) algorithms look like

Gradient descent. $x_{t+1} := x_t - \eta \nabla f(x_t)$.

Projected Gradient Descent. $x_{t+1} := \Pi_K(x_t - \eta \nabla f(x_t))$.

Newton's method. $x_{t+1} := x_t - (\nabla^2 f(x_t))^{-1} \nabla f(x_t)$.

Mind the gap: Run one step of Newton's method on $f(x) = \frac{1}{2} x^T M x + bx$.

Examples of how (iterative) algorithms look like

Gradient descent. $x_{t+1} := x_t - \eta \nabla f(x_t)$.

Projected Gradient Descent. $x_{t+1} := \Pi_K(x_t - \eta \nabla f(x_t))$.

Newton's method. $x_{t+1} := x_t - (\nabla^2 f(x_t))^{-1} \nabla f(x_t)$.

Mind the gap: Run one step of Newton's method on $f(x) = \frac{1}{2} x^T M x + bx$. For all quadratic functions one step of Newton's method lands on the optimum!

Unit capacity Max flow and convex optimization

Problem. Given unit capacity graph of $G = (V, E)$, vertices $s, t \in G$ route the maximum amount of flow from s to t.

Unit capacity Max flow and convex optimization

Problem. Given unit capacity graph of $G = (V, E)$, vertices $s, t \in G$ route the maximum amount of flow from s to t.

Let $x \in \mathbb{R}^E, \|x\|_\infty \leq 1$ with constrains for all $u \in V$:

Unit capacity Max flow and convex optimization

Problem. Given unit capacity graph of $G = (V, E)$, vertices $s, t \in G$ route the maximum amount of flow from s to t.

Let $x \in \mathbb{R}^E, \|x\|_\infty \leq 1$ with constrains for all $u \in V$:

- $\sum_{(u,v)\in E} x_{uv} - \sum_{(v,u)\in E} x_{uv} = 0$ if $u \neq s, t$

Unit capacity Max flow and convex optimization

Problem. Given unit capacity graph of $G = (V, E)$, vertices $s, t \in G$ route the maximum amount of flow from s to t.

Let $x \in \mathbb{R}^E, \|x\|_\infty \leq 1$ with constrains for all $u \in V$:

- $\sum_{(u,v)\in E} x_{uv} - \sum_{(v,u)\in E} x_{uv} = 0$ if $u \neq s, t$
- $\sum_{(u,v)\in E} x_{uv} - \sum_{(v,u)\in E} x_{uv} = F*$ if $u = s$

Unit capacity Max flow and convex optimization

Problem. Given unit capacity graph of $G = (V, E)$, vertices $s, t \in G$ route the maximum amount of flow from s to t.

Let $x \in \mathbb{R}^E, \|x\|_\infty \leq 1$ with constrains for all $u \in V$:

- $\sum_{(u,v)\in E} x_{uv} - \sum_{(v,u)\in E} x_{uv} = 0$ if $u \neq s, t$
- $\sum_{(u,v)\in E} x_{uv} - \sum_{(v,u)\in E} x_{uv} = F*$ if $u = s$
- $\sum_{(u,v)\in E} x_{uv} - \sum_{(v,u)\in E} x_{uv} = -F*$ if $u = t$

Unit capacity Max flow and convex optimization

Problem. Given unit capacity graph of $G = (V, E)$, vertices $s, t \in G$ route the maximum amount of flow from s to t.

Let $x \in \mathbb{R}^E, \|x\|_\infty \leq 1$ with constrains for all $u \in V$:

- $\sum_{(u,v)\in E} x_{uv} - \sum_{(v,u)\in E} x_{uv} = 0$ if $u \neq s, t$
- $\sum_{(u,v)\in E} x_{uv} - \sum_{(v,u)\in E} x_{uv} = F*$ if $u = s$
- $\sum_{(u,v)\in E} x_{uv} - \sum_{(v,u)\in E} x_{uv} = -F*$ if $u = t$

$$\boxed{\min \|x\|_\infty \text{ s.t. } Bx = \mathbb{1}_s - \mathbb{1}_t.}$$

Unit capacity Max flow and convex optimization

Problem. Given unit capacity graph of $G = (V, E)$, vertices $s, t \in G$ route the maximum amount of flow from s to t.

Let $x \in \mathbb{R}^E, \|x\|_\infty \le 1$ with constrains for all $u \in V$:

- $\sum_{(u,v)\in E} x_{uv} - \sum_{(v,u)\in E} x_{uv} = 0$ if $u \neq s, t$
- $\sum_{(u,v)\in E} x_{uv} - \sum_{(v,u)\in E} x_{uv} = F*$ if $u = s$
- $\sum_{(u,v)\in E} x_{uv} - \sum_{(v,u)\in E} x_{uv} = -F*$ if $u = t$

$\min \|x\|_\infty$ s.t. $Bx = \mathbb{1}_s - \mathbb{1}_t$.

$\min \ \eta \log(\sum_i e^{x_i/\eta} + e^{-x_i/\eta})$ s.t. $Bx = \mathbb{1}_s - \mathbb{1}_t$.

Langragian Duality

$\inf_{x \in \mathbb{R}^n} f(x)$
s.t. $f_j(x) \leq 0$, for $j \in [m]$
$h_i(x) = 0$, for $i \in [p]$

Langragian Duality

$$\begin{aligned}&\inf_{x\in\mathbb{R}^n} f(x)\\ &\text{s.t. } f_j(x)\le 0, \text{ for } j\in[m]\\ &h_i(x)=0, \text{ for } i\in[p]\end{aligned}$$

Let $L(x,\lambda,\mu) := f(x) + \sum_{j\in[m]} \lambda_j f_j(x) + \sum_{i\in[p]} \mu_i h_i(x)$.

Langragian Duality

$$\inf_{x\in\mathbb{R}^n} f(x)$$
s.t. $f_j(x) \le 0$, for $j \in [m]$
$h_i(x) = 0$, for $i \in [p]$

Let $L(x,\lambda,\mu) := f(x) + \sum_{j\in[m]} \lambda_j f_j(x) + \sum_{i\in[p]} \mu_i h_i(x)$.

- $x \in K \Rightarrow$

Langragian Duality

$$\boxed{\begin{array}{l}\inf_{x\in\mathbb{R}^n} f(x)\\ \text{s.t. } f_j(x)\le 0, \text{ for } j\in[m]\\ h_i(x)=0, \text{ for } i\in[p]\end{array}}$$

Let $L(x,\lambda,\mu) := f(x) + \sum_{j\in[m]} \lambda_j f_j(x) + \sum_{i\in[p]} \mu_i h_i(x)$.

- $x \in K \Rightarrow L(x,\lambda,\mu) \le f(x)$.
- $\sup_{\lambda\ge 0,\mu} L(x,\lambda,\mu) = f(x), x \in K$

Langragian Duality

$$\begin{array}{|l|}\hline \inf_{x\in\mathbb{R}^n} f(x) \\ \text{s.t. } f_j(x) \leq 0, \text{ for } j \in [m] \\ h_i(x) = 0, \text{ for } i \in [p] \\ \hline \end{array}$$

Let $L(x, \lambda, \mu) := f(x) + \sum_{j\in[m]} \lambda_j f_j(x) + \sum_{i\in[p]} \mu_i h_i(x)$.

- $x \in K \Rightarrow L(x, \lambda, \mu) \leq f(x)$.
- $\sup_{\lambda\geq 0,\mu} L(x, \lambda, \mu) = f(x), x \in K$
- $\sup_{\lambda,\geq 0,\mu} L(x, \lambda, \mu) = \infty$, otherwise

Langragian Duality

$\inf_{x\in\mathbb{R}^n} f(x)$
s.t. $f_j(x) \le 0$, for $j \in [m]$
$h_i(x) = 0$, for $i \in [p]$

Let $L(x,\lambda,\mu) := f(x) + \sum_{j\in[m]} \lambda_j f_j(x) + \sum_{i\in[p]} \mu_i h_i(x)$.

- $x \in K \Rightarrow L(x,\lambda,\mu) \le f(x)$.
- $\sup_{\lambda\ge 0,\mu} L(x,\lambda,\mu) = f(x), x \in K$
- $\sup_{\lambda,\ge 0,\mu} L(x,\lambda,\mu) = \infty$, otherwise

$$y^\star = \inf_{x\in K} \sup_{\lambda\ge 0,\mu} L(x,\lambda,\mu) = \inf_{x\in\mathbb{R}^n} \sup_{\lambda\ge 0,\mu} L(x,\lambda,\mu)$$

Langragian Duality

$$y^\star = \inf_{x \in K} \sup_{\lambda \geq 0, \mu} L(x, \lambda, \mu) = \inf_{x \in \mathbb{R}^n} \sup_{\lambda \geq 0, \mu} L(x, \lambda, \mu)$$

Langragian Duality

$$y^\star = \inf_{x \in K} \sup_{\lambda \geq 0, \mu} L(x, \lambda, \mu) = \inf_{x \in \mathbb{R}^n} \sup_{\lambda \geq 0, \mu} L(x, \lambda, \mu)$$

Let

$$g(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu)$$

Definition (Dual Program)

$\sup_{\lambda \geq 0, \mu} g(\lambda, \mu)$.

Langragian Duality

$$y^\star = \inf_{x\in K}\sup_{\lambda\geq 0,\mu} L(x,\lambda,\mu) = \inf_{x\in\mathbb{R}^n}\sup_{\lambda\geq 0,\mu} L(x,\lambda,\mu)$$

Let

$$g(\lambda,\mu) = \inf_{x\in\mathbb{R}^n} L(x,\lambda,\mu)$$

Definition (Dual Program)

$\sup_{\lambda\geq 0,\mu} g(\lambda,\mu)$.

Theorem (Weak Duality)

$\sup_{\lambda\geq 0,\mu} g(\lambda,\mu) \leq \inf_{x\in K} f(x)$.

Strong Duality

Slater's condition. There exists $\bar{x}$ such that $h_j(\bar{x}) = 0$ and $f_i(\bar{x}) < 0$.

Strong Duality

Slater's condition. There exists $\bar{x}$ such that $h_j(\bar{x}) = 0$ and $f_i(\bar{x}) < 0$.

Theorem (Slater's gives strong guality)

If all f_j, h_i are affine and Slater's condition holds, then

$$\sup_{\lambda \geq 0, \mu} g(\lambda, \mu) = \inf_{x \in K} f(x)$$

Examples

Linear programming. $\min\ c^T x \quad \text{s.t.} \quad Ax \geq b$

Examples

Linear programming. $\min\ c^T x \quad \text{s.t.} \quad Ax \geq b$

Examples

Linear programming. $\min c^T x \quad \text{s.t.} \quad Ax \geq b$
Let $L(x, \lambda) = c^T x + \lambda^T (b - Ax) = \langle x, c - A^T \lambda \rangle + \langle b, \lambda \rangle$

Examples

Linear programming. $\min\ c^T x \quad \text{s.t.} \quad Ax \geq b$
Let $L(x, \lambda) = c^T x + \lambda^T (b - Ax) = \langle x, c - A^T \lambda \rangle + \langle b, \lambda \rangle$
What is $g(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda)$?

$$\max \langle b, \lambda \rangle$$
$$\text{s.t.} \quad A^T \lambda = c, \lambda \geq 0$$

Examples

Linear programming. $\min c^T x \quad \text{s.t.} \quad Ax \geq b$
Let $L(x, \lambda) = c^T x + \lambda^T (b - Ax) = \langle x, c - A^T \lambda \rangle + \langle b, \lambda \rangle$
What is $g(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda)$?

$$\boxed{\begin{aligned} &\max \langle b, \lambda \rangle \\ \text{s.t.} \quad &A^T \lambda = c, \lambda \geq 0 \end{aligned}}$$

It is known that in the setting of linear programming strong duality holds.

Examples

Linear programming. $\min c^T x \quad \text{s.t.} \quad Ax \geq b$
Let $L(x, \lambda) = c^T x + \lambda^T (b - Ax) = \langle x, c - A^T \lambda \rangle + \langle b, \lambda \rangle$
What is $g(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda)$?

$$\boxed{\begin{aligned} &\max \langle b, \lambda \rangle \\ \text{s.t.} \quad &A^T \lambda = c, \lambda \geq 0 \end{aligned}}$$

It is known that in the setting of linear programming strong duality holds.
Fact. There exist some convex programs for which strong duality fails, but such programs are not commonly encountered in practice.

Thank you!