UNIVERSITÄT DES SAARLANDES

max planck institut informatik

ADFOCS 2021

Convex Optimization and Graph Algorithms
Primer Week - Day 1

Karl Bringmann

July 26, 2021

Speakers

Alina Ene

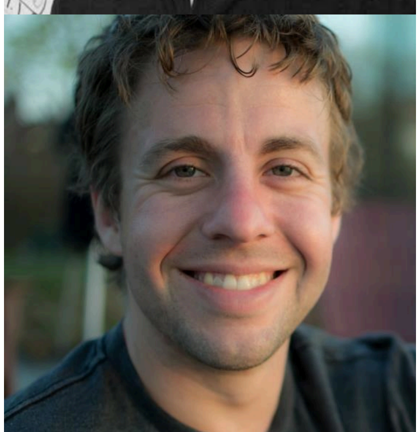
Boston University

Adaptive gradient descent

Rasmus Kyng

ETH Zurich

Graphs, sampling, and iterative methods



Aaron Sidford

Stanford University

Optimization Methods for Maximum Flow

Timetable

July 26-30: primer week

	Mon	Tue	Wed	Thu	Fri
16:00-18:00	**Primer talk**	**Primer talk**	**Primer talk**	**Primer talk**	**Primer talk**

August 2-6: speaker talks

	Mon	Tue	Wed	Thu	Fri
15:00-16:00		Alina Ex	Alina Ex		Rasmus Ex
16:00-18:00	**Alina**	**Alina**	**Alina**	**Rasmus**	**Rasmus**
18:00-19:00					

August 9-13: speaker talks

	Mon	Tue	Wed	Thu	Fri
15:00-16:00	Rasmus Ex				
16:00-18:00	**Aaron**	**Aaron**	**MPI talks**	**Rasmus**	**Aaron**
18:00-19:00		Aaron Ex	Aaron Ex		

Primer Week

July 26-30: primer week

	Mon	Tue	Wed	Thu	Fri
16:00-18:00	**Primer talk**	**Primer talk**	**Primer talk**	**Primer talk**	**Primer talk**
Speaker:	Karl Bringmann	Vasileios Nakos	Alejandro Cassis	Kurt Mehlhorn	Andreas Karrenbauer
Topic:	Math Background	Intro to Convex Optimization	Gradient Descent	Mirror Descent	Spectral Graph Theory

Goal: learn background of continuous optimization, assuming background in discrete TCS

Primer Day 1
Math Background

Graph Laplacian

undirected connected graph $G = (V, E)$, edge weights $w \in \mathbb{R}^{E}_{\geq 0}$

B = incidence matrix, $B_{v,(u,v)} = 1, B_{v,(v,u)} = -1, B_{v,e} = 0$ otherwise

W = diagonal matrix of edge weights, $W_{e,e} = w(e)$

A = weighted adjacency matrix, $A_{u,v} = w(u, v)$

D = diagonal matrix of weighted degrees, $D_{v,v} = \sum_{u \in V} w(u, v)$

Laplacian: $L_G = BWB^T = D - A$

step from u to (u, v)

multiply by $w(u, v)$

step from (u, v) to u or v

Laplacian Quadratic Form

A = weighted adjacency matrix, $A_{u,v} = w(u,v)$

D = diagonal matrix of weighted degrees, $D_{v,v} = \sum_{u \in V} w(u,v)$

Laplacian: $L = BWB^T = D - A$

$$x^T L x = \sum_{u,v \in V} x_u L_{u,v} x_v = \sum_{(u,v) \in E} w(u,v) \cdot (x_u - x_v)^2$$

$x^T L x \geq 0 \ \forall x$

L is a real symmetric $n \times n$ matrix with $\ker(L_G) = \mathbb{R} \cdot \mathbf{1}$

study such matrices in more generality

Eigenvalues

Always assume: $A \in \mathbb{R}^{n \times n}$ symmetric

Spectral Theorem: $\exists V, \Lambda \in \mathbb{R}^{n \times n}$ where Λ is diagonal and

- $A = V \Lambda V^T$
- $I = V^T V$

eigenvectors $\equiv$ columns of V

eigenvalues $\equiv$ diagonal entries of Λ

$$[\ A v_i = V \Lambda V^T v_i = V \Lambda e_i = V\left(\Lambda_{i,i} e_i\right) = \Lambda_{i,i} \cdot V e_i = \Lambda_{i,i} v_i\ ]$$

Eigenvalues vs Quadratic Form

Quadratic Form: $f(x) = x^T A x = \sum_{i,j=1}^{n} x_i A_{i,j} x_j$

Let $\lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of A

Courant-Fischer Theorem:

$$\lambda_1 = \min_{x \neq 0} \frac{x^T A x}{x^T x}$$

$$\lambda_n = \max_{x \neq 0} \frac{x^T A x}{x^T x}$$

more generally:

$$\lambda_i = \max_{\substack{\text{subspace } W \subseteq \mathbb{R}^n \\ \dim(W) = n+1-i}} \min_{\substack{x \neq 0 \\ x \in W}} \frac{x^T A x}{x^T x}$$

$$\lambda_i = \min_{\substack{\text{subspace } W \subseteq \mathbb{R}^n \\ \dim(W) = i}} \max_{\substack{x \neq 0 \\ x \in W}} \frac{x^T A x}{x^T x}$$

$A \in \mathbb{R}^{n \times n}$ symmetric

Positive Semi-Definite Matrices

A is **positive semi-definite** if $x^T A x \geq 0 \;\; \forall x$ → write $A \succcurlyeq 0$

A is **positive definite** if $x^T A x > 0 \;\; \forall x \neq 0$ → write $A \succ 0$

[$A \succcurlyeq 0 \;\Leftrightarrow\; f(x) = x^T A x$ is convex]

$A \succcurlyeq 0 \;\Leftrightarrow\;$ all eigenvalues are non-negative

$A \succ 0 \;\Leftrightarrow\;$ all eigenvalues are positive

Square Root: $A \succcurlyeq 0 \;\Leftrightarrow\;$ there exists unique $B \succcurlyeq 0$ with $A = B \cdot \, B$

$B = A^{1/2} =$ matrix with same eigenvectors, square root of eigenvalues

$A \in \mathbb{R}^{n \times n}$ symmetric, eigendecomposition $A = V \Lambda V^T$

Positive Semi-Definite Matrices

Examples: diagonal $D \succcurlyeq 0 \iff$ all entries ≥ 0

if $A \succcurlyeq 0$: $c \cdot A \succcurlyeq 0 \quad \forall c \geq 0$

$B^T AB \succcurlyeq 0 \quad \forall B \in \mathbb{R}^{n \times n}$

$A + B \succcurlyeq 0 \quad \forall B \succcurlyeq 0$

set of psd matrices is convex

$A \cdot B \succcurlyeq 0$ in general not true; not even symmetric!

but $B^{1/2} AB^{1/2} \succcurlyeq 0 \quad \forall B \succcurlyeq 0$

$A \in \mathbb{R}^{n \times n}$ symmetric, eigendecomposition $A = V \Lambda V^T$

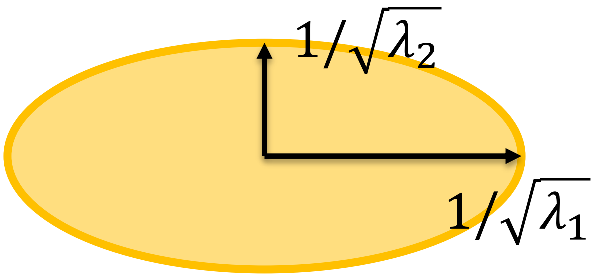
$A \succcurlyeq 0 \iff$ A is psd $\iff x^T Ax \geq 0 \ \forall x$

Geometric Interpretation

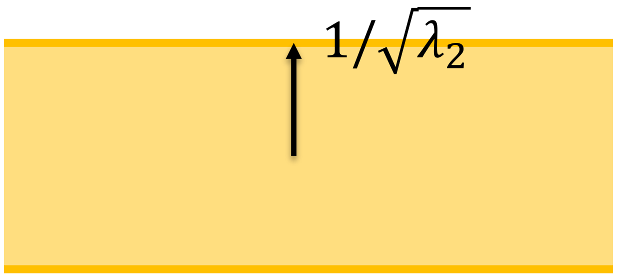
Ellipsoid: $A \succcurlyeq 0$ induces ellipsoid $E_A = \{\, x \in \mathbb{R}^n \mid x^T A x \leq 1 \,\}$

e.g. $x^T A x = \lambda_1 x_1^2 + \lambda_2 x_2^2 \leq 1$

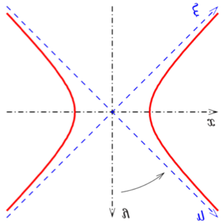
if $\lambda_1, \lambda_2 > 0$:



if $\lambda_1 = 0, \lambda_2 > 0$:



if $\lambda_1 > 0, \lambda_2 < 0$:



$A \in \mathbb{R}^{n \times n}$ symmetric, eigendecomposition $A = V \Lambda V^T$

$A \succcurlyeq 0 \iff$ A is psd $\iff x^T A x \geq 0 \;\; \forall x$

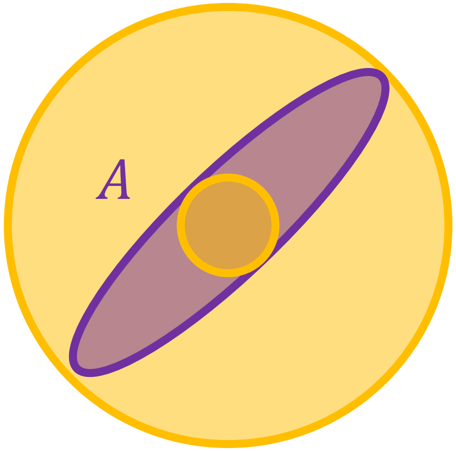
Loewner Order

$$A \succcurlyeq B \;:\Longleftrightarrow\; A - B \succcurlyeq 0$$

$$\Longleftrightarrow\; x^T(A-B)x \geq 0 \quad \forall x$$

$$\Longleftrightarrow\; x^T A x \geq x^T B x \quad \forall x$$

$$\Longleftrightarrow\; E_A \subseteq E_B$$



$$A \succcurlyeq 0 \implies \exists \alpha, \beta \geq 0: \quad \alpha \cdot I \preccurlyeq A \preccurlyeq \beta \cdot I$$

e.g. $\alpha = \lambda_1(A),\ \beta = \lambda_n(A)$

condition number $\kappa = \dfrac{\beta}{\alpha} = \dfrac{\lambda_n(A)}{\lambda_1(A)}$

$A \in \mathbb{R}^{n\times n}$ symmetric, eigendecomposition $A = V\Lambda V^T$

$A \succcurlyeq 0 \iff$ A is psd $\iff x^T A x \geq 0 \ \forall x$

Matrix Operations

Matrix Inverse: $A^{-1} = V\Lambda^{-1}V^T$ where $(\Lambda^{-1})_{i,i} = \left(\Lambda_{i,i}\right)^{-1}$

only exists if A is **non-singular**, i.e., $\Lambda_{i,i} \neq 0$ for all i

Pseudoinverse: $A^{+} = V\Lambda^{+}V^T$ where $(\Lambda^{+})_{i,i} = \left(\Lambda_{i,i}\right)^{-1}$ if $\Lambda_{i,i} \neq 0$

$= 0$ otherwise

Matrix function: $f(A) = Vf(\Lambda)V^T$ where $\left(f(\Lambda)\right)_{i,i} = f\left(\Lambda_{i,i}\right)$

e.g. $\exp(A)$, $A^{1/2}$

$A \in \mathbb{R}^{n\times n}$ symmetric, eigendecomposition $A = V\Lambda V^T$

Same-Matrix Inequalities

if $f(x) \le g(x)$ for all reals x with $\alpha \le x \le \beta$

then $f(A) \preccurlyeq g(A)$ for all symmetric $A \in \mathbb{R}^{n\times n}$ with $\alpha \cdot I \preccurlyeq A \preccurlyeq \beta \cdot I$

Example: $1 + x \le \exp(x) \le 1 + x + x^2$ for all $x \le 1$

$I + A \preccurlyeq \exp(A) \preccurlyeq I + A + A^2$ for all symmetric $A \preccurlyeq I$

Proof: $$g(A) - f(A) = V^T diag\big(\underbrace{g(\lambda_i) - f(\lambda_i)}_{\ge 0}\big)V \succcurlyeq 0$$

$A \in \mathbb{R}^{n\times n}$ symmetric, eigendecomposition $A = V\Lambda V^T$

$A \succcurlyeq 0 \iff$ A is psd $\iff x^T A x \ge 0 \;\forall x$

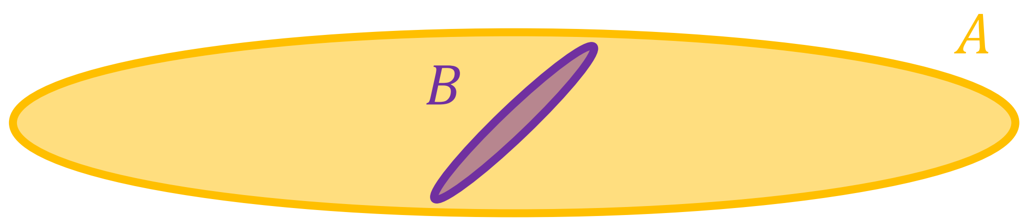
Different-Matrix Inequalities

let f map symmetric $n \times n$ matrices to symmetric $n \times n$ matrices

f is **monotone increasing** if $A \preccurlyeq B$ implies $f(A) \preccurlyeq f(B)$

For any $0 \prec A \preccurlyeq B$ we have $B^{-1} \preccurlyeq A^{-1}$ and $\log(A) \preccurlyeq \log(B)$

f
X^2
$\exp(X)$
$-X^{-1}$
$\log(X)$



$A \in \mathbb{R}^{n \times n}$ symmetric, eigendecomposition $A = V \Lambda V^T$

$A \succcurlyeq 0 \iff$ A is psd $\iff x^T A x \geq 0 \;\; \forall x$

Multivariate Calculus: Gradient

Assume: $f : \mathbb{R}^n \to \mathbb{R}$ is sufficiently differentiable

Gradient: $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)^T$

$$f(x + \delta) = f(x) + \nabla f(x)^T \delta + o(\|\delta\|_2) \quad \forall x, \delta$$

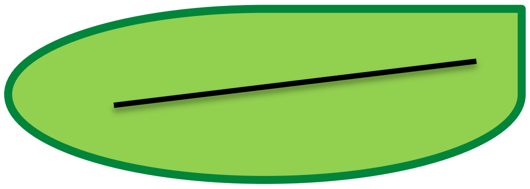
Taylor: $\forall x, y \; \exists z \in [x, y] := \{\lambda x + (1 - \lambda) y \mid 0 \leq \lambda \leq 1\}$:

$$f(y) = f(x) + \nabla f(z)^T (y - x)$$

x local extremum $\implies \nabla f(x) = 0$

f convex and $\nabla f(x) = 0 \implies x$ global minimum

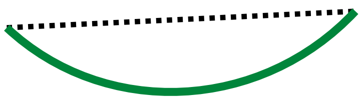
Convexity



Convex set: $S \subseteq \mathbb{R}^n$ is convex if $\forall x, y \in S\ \forall z \in [x, y]: z \in S$

Convex function: $\forall x, y \forall \lambda \in [0,1]$:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$



f is convex if and only if $f(y) \geq f(x) + \nabla f(x)^T (y - x)\ \ \forall x, y$

Maximum of convex functions is convex

Multivariate Calculus: Hessian

Assume: $f : \mathbb{R}^n \to \mathbb{R}$ is sufficiently differentiable

Hessian: $H_f(x) = \nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right)_{i,j} \in \mathbb{R}^{n \times n}$

$$f(x+\delta) = f(x) + \nabla f(x)^T \delta + \tfrac{1}{2}\delta^T H_f(x)\delta + o(\|\delta\|_2^2) \quad \forall x, \delta$$

Taylor: $\forall x, y\ \exists z \in [x, y]$:

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \tfrac{1}{2}(y-x)^T H_f(z)\,(y - x)$$

x local minimum $\implies$ $H_f(x) \succcurlyeq 0$

$\nabla f(x) = 0$ and $H_f(x) \succ 0 \implies$ x local minimum

f convex if and only if $H_f(x) \succcurlyeq 0\ \forall x$

Condition Number

f convex if and only if $H_f(x) \succcurlyeq 0 \;\; \forall x$

$H_f(x) \succ 0$ ... strictly convex

$H_f(x) \succcurlyeq \alpha \cdot I$ ... α-strictly convex

$H_f(x) \preccurlyeq \beta \cdot I$ ... β-smooth

$$f(x) + \nabla f(x)^T(y - x) + \frac{\alpha}{2}\|y - x\|_2^2$$
$$\leq f(y) \leq$$
$$f(x) + \nabla f(x)^T(y - x) + \frac{\beta}{2}\|y - x\|_2^2 \qquad \forall x, y$$

$\kappa \coloneqq \frac{\beta}{\alpha}$ ... condition number of f

See you in Gather!