COMP 545: Advanced topics in optimization
From simple to complex ML systems

Lecture 1

Overview

$$
\begin{aligned}
\min_{x} \quad & f(x) \\
\text{s.t.} \quad & x \in \mathcal{C}
\end{aligned}
$$

Overview

- Different objective classes
- Different strategies within each problem
- Different approaches based on computational capabilities
- Different approaches based on constraints

$$\min_{x} \quad f(x)$$
$$\text{s.t.} \quad x \in \mathcal{C}$$

Overview

- Different objective classes
- Different strategies within each problem
- Different approaches based on computational capabilities
- Different approaches based on constraints

$$\min_{x} \quad f(x)$$
$$\text{s.t.} \quad x \in \mathcal{C}$$

And, always having in mind applications in machine learning, AI and signal processing

Motivation (no fancy images included)

Provable efficiency

Motivation (no fancy images included)

Provable efficiency

Lots of data

Motivation (no fancy images included)

Provable efficiency

Harder problems

Lots of data

Motivation (no fancy images included)

More complicated models

Provable efficiency

Harder problems

Lots of data

Motivation (no fancy images included)

More complicated models

Better results

Provable efficiency

Harder problems

Lots of data

Motivation (no fancy images included)

More complicated models

Better results

Provable efficiency

Harder problems

Lots of data

Motivation (no fancy images included)

?

More complicated models

Better results

Provable efficiency

Harder problems

Lots of data

Motivation (no fancy images included)

Provable efficiency

"What shall we do?"

Motivation (no fancy images included)

Provable efficiency

"What shall we do?"

Set up algo nicely

Use prior knowledge

Converge faster

Exploit resources

Topics

- Continuous **optimization** (in general)
  - See syllabus
  - Both theory and practice

Topics

- Continuous **optimization** (in general)
  - See syllabus
  - Both theory and practice
- Recent applications that drive research

(We can discuss about it)

Topics

- Continuous **optimization** (in general)
  - See syllabus
  - Both theory and practice
- Recent applications that drive research

(We can discuss about it)

- When no theory applies, some intuition

Topics NOT covered in this course

- Extensive coverage of convex optimization

  (Duality, KKT conditions, self-concordance, interior point methods, Lagrange multipliers)

Topics NOT covered in this course

- Extensive coverage of convex optimization

  (Duality, KKT conditions, self-concordance, interior point methods, Lagrange multipliers)

- (Mixed) integer programming

Topics NOT covered in this course

- Extensive coverage of convex optimization

  (Duality, KKT conditions, self–concordance, interior point methods, Lagrange multipliers)

- (Mixed) integer programming
- Combinatorial optimization algorithms

  (E.g., Graph algorithms)

Topics NOT covered in this course

- Extensive coverage of convex optimization

  (Duality, KKT conditions, self−concordance, interior point methods, Lagrange multipliers)

- (Mixed) integer programming
- Combinatorial optimization algorithms

  (E.g., Graph algorithms)

- Randomized algorithms

  (See Anshu’s course)

Topics NOT covered in this course

- Extensive coverage of convex optimization

  (Duality, KKT conditions, self–concordance, interior point methods, Lagrange multipliers)

- (Mixed) integer programming
- Combinatorial optimization algorithms

  (E.g., Graph algorithms)

- Randomized algorithms

  (See Anshu's course)

- Online algorithms, like bandits

Topics NOT covered in this course

- Extensive coverage of convex optimization

  (Duality, KKT conditions, self–concordance, interior point methods, Lagrange multipliers)

- (Mixed) integer programming
- Combinatorial optimization algorithms

  (E.g., Graph algorithms)

- Randomized algorithms

  (See Anshu's course)

- Online algorithms, like bandits
- Bayesian algorithms

Who is this class for?

- Definitely, for PhD students

Who is this class for?

- Definitely, for PhD students

- Master or even undergraduates that want to start research

(but get in touch with me soon to assess your background)

Who is this class for?

- Definitely, for PhD students
- Master or even undergraduates that want to start research

(but get in touch with me soon to assess your background)

- Just auditing is fine by me

What is the vision for this course?

- For starters, have in mind that this is a first−time taught course

(Any feedback is more than welcome)

- My purpose and vision is to introduce a series of optimization courses in the CS (and Duncan Hall's in general) curriculum

(Feedback on GoogleDoc)

What is the vision for this course?

- For starters, have in mind that this is a first–time taught course

(Any feedback is more than welcome)

- My purpose and vision is to introduce a series of optimization courses in the CS (and Duncan Hall's in general) curriculum

- The vision is for this course to be part of a sequence of courses that will focus on the theory+practice of methods

(Feedback on GoogleDoc)

Course format

- Lectures (slides) + whiteboard + in-class code running

Course format

- Lectures (slides) + whiteboard + in−class code running
- Material broad enough to cover range of problems

(we will not get into much detail)

Course format

- Lectures (slides) + whiteboard + in−class code running
- Material broad enough to cover range of problems

(we will not get into much detail)

- Some material inspired from personal research

Course format

- Lectures (slides) + whiteboard + in−class code running
- Material broad enough to cover range of problems

  (we will not get into much detail)
- Some material inspired from personal research
- “Interlude” lectures to provide some background (if needed)

Course format

- Lectures (slides) + whiteboard + in−class code running
- Material broad enough to cover range of problems

  (we will not get into much detail)
- Some material inspired from personal research
- "Interlude" lectures to provide some background (if needed)
- Your workload: scribing, presentations,
  reviews, final project

Goals + outcomes

- Learn about research in related fields

Goals + outcomes

- Learn about research in related fields
- Make connections between areas,
  understand how research advances in such areas..

Goals + outcomes

- Learn about research in related fields

- Make connections between areas,
  understand how research advances in such areas..

- Consider possible extensions of these works (project)

Goals + outcomes

- Learn about research in related fields

- Make connections between areas,
  understand how research advances in such areas..

- Consider possible extensions of these works (project)

- Comprehend how optimization is key in ML/AI/SP

Goals + outcomes

- Learn about research in related fields
- Make connections between areas,
  understand how research advances in such areas..
- Consider possible extensions of these works (project)
- Comprehend how optimization is key in ML/AI/SP
- Read and review recent papers

Prerequisites

- If you have taken any ML class, you are good to go

Prerequisites

- If you have taken any ML class, you are good to go
- Basics of calculus, linear algebra, basic knowledge of ML topics

Prerequisites

- If you have taken any ML class, you are good to go
- Basics of calculus, linear algebra, basic knowledge of ML topics
- Programming skills are not necessary

(but might be required, depending on the project selected)

Prerequisites

- If you have taken any ML class, you are good to go
- Basics of calculus, linear algebra, basic knowledge of ML topics
- Programming skills are not necessary

  (but might be required, depending on the project selected)
- A quiz will be given today for self-assessment

Grading policy

- 5% participation
- 10% scribing notes
- 15% reviews
- 20% presentation
- 50% project

Grading policy

- 5% participation
- 10% scribing notes
- 15% reviews
- 20% presentation
- 50% project

Usually there is scaling in final grades. For me, a good grade is given based on the overall performance of the students: I value self-motivation, being proactive and enthusiasm.

Participation + scribing

- “Παν μετρον αριστον”

(Moderation is key)

Participation + scribing

- "Παν μετρον αριστον"
(Moderation is key)

- Questions over email is also fine
(I have set up an email for the course – see the syllabus)

Participation + scribing

- "Παν μετρον αριστον" (Moderation is key)
- Questions over email is also fine
  (I have set up an email for the course – see the syllabus)
- Scribing is useful for you to understand better the material
  (or even get a better understanding)

Participation + scribing

- "Παν μετρον αριστον" (Moderation is key)
- Questions over email is also fine (I have set up an email for the course – see the syllabus)
- Scribing is useful for you to understand better the material (or even get a better understanding)
- Individuals or groups (2–3) of volunteers for each lecture (it will depend on the attendance)

Participation + scribing

- “Παν μετρον αριστον” (Moderation is key)
- Questions over email is also fine

  (I have set up an email for the course – see the syllabus)
- Scribing is useful for you to understand better the material

  (or even get a better understanding)
- Individuals or groups (2–3) of volunteers for each lecture

  (it will depend on the attendance)
- Deliverable in LaTEX

  (template available online)

Reviews

(Almost every week where you choose a paper)

- Select papers from a pile of .pdfs that will be provided

(Reviews will be related to the topics currently taught)

Reviews

(Almost every week where you choose a paper)

- Select papers from a pile of .pdfs that will be provided

(Reviews will be related to the topics currently taught)

- Single page reviews, similar to NIPS/ICML standards:

(but not random as it usually is now)

  - Comment on novelty, clarity, importance
  - Main comments + your overall score

Reviews

(Almost every week where you choose a paper)

- Select papers from a pile of .pdfs that will be provided

  (Reviews will be related to the topics currently taught)

- Single page reviews, similar to NIPS/ICML standards:

  (but not random as it usually is now)

  - Comment on novelty, clarity, importance
  - Main comments + your overall score

- You select on Tuesday – deliverables next Tuesday

  (not always true if the subject is too generic for paper reading)

Presentations

(btw, this is not your project presentation)

- **When**: Twice during the semester (after lecture 4)

(it will depend on the attendance)

Presentations

(btw, this is not your project presentation)

- **When**: Twice during the semester
  (after lecture 4)

(it will depend on the attendance)

- (BTW, no midterms, no other homeworks)

Presentations

(btw, this is not your project presentation)

- **When**: Twice during the semester
  (after lecture 4)

(it will depend on the attendance)

- (BTW, no midterms, no other homeworks)

- There might be some discussions during the lectures
  Take advantage by asking questions

(depending on the size of the class)

Presentations

(btw, this is not your project presentation)

- **When**: Twice during the semester
  (after lecture 4)

  (it will depend on the attendance)

- (BTW, no midterms, no other homeworks)

- There might be some discussions during the lectures
  Take advantage by asking questions

  (depending on the size of the class)

- Presentation should be at most 10–12 minutes

  (tentative – depends on the class size)

Presentations

(btw, this is not your project presentation)

- **When**: Twice during the semester
  (after lecture 4)

(it will depend on the attendance)

- (BTW, no midterms, no other homeworks)

- There might be some discussions during the lectures
  Take advantage by asking questions

(depending on the size of the class)

- Presentation should be at most 10–12 minutes

(tentative – depends on the class size)

- Grading: slides quality, clarity of main ideas

Final Project

(see syllabus)

Final Project

(see syllabus)

Please come find me the earliest to discuss projects

Final Project

(see syllabus)

Please come find me the earliest to discuss projects

You should start reading papers soon, so that around mid-way you have a good project proposal

Any questions?

Quiz
(15 min. or NBO)

Setting up the background

Vectors

- Notation convention: vectors = lowercase, matrices = uppercase

Vectors

- Notation convention: vectors = lowercase, matrices = uppercase
- Vector in p-dimensions: $x \in \mathbb{R}^p$

$$x = \left[x_1, x_2, \ldots, x_p\right]^\top$$

Vectors

- Notation convention: vectors = lowercase, matrices = uppercase

- Vector in p−dimensions: $x \in \mathbb{R}^p$

$$x = \left[x_1, x_2, \ldots, x_p\right]^\top$$

- Some properties of vectors:

$$x + y = y + x, \quad x, y \in \mathbb{R}^p$$

(Commutative)

Vectors

- Notation convention: vectors = lowercase, matrices = uppercase

- Vector in p-dimensions: $x \in \mathbb{R}^p$

$$x = \left[x_1, x_2, \ldots, x_p\right]^\top$$

- Some properties of vectors:

$$x + y = y + x, \quad x, y \in \mathbb{R}^p \qquad \text{(Commutative)}$$

$$(x + y) + x = x + (y + z), \quad x, y, z \in \mathbb{R}^p \qquad \text{(Associative)}$$

Vectors

- Notation convention: vectors = lowercase, matrices = uppercase

- Vector in p−dimensions: $x \in \mathbb{R}^p$

$$x = \left[x_1, x_2, \ldots, x_p\right]^\top$$

- Some properties of vectors:

$$x + y = y + x, \quad x, y \in \mathbb{R}^p$$ (Commutative)

$$(x + y) + x = x + (y + z), \quad x, y, z \in \mathbb{R}^p$$ (Associative)

$$0 + x = x, \quad x \in \mathbb{R}^p$$

Vectors

- Notation convention: vectors = lowercase, matrices = uppercase

- Vector in p−dimensions: $x \in \mathbb{R}^p$

$$x = \left[x_1, x_2, \ldots, x_p\right]^\top$$

- Some properties of vectors:

$$x + y = y + x, \quad x, y \in \mathbb{R}^p$$ (Commutative)

$$(x + y) + x = x + (y + z), \quad x, y, z \in \mathbb{R}^p$$ (Associative)

$$0 + x = x, \quad x \in \mathbb{R}^p$$

$$\alpha(x + y) = \alpha x + \alpha y, \quad x, y \in \mathbb{R}^p$$ (Distributive)

Vectors

– Span of a set of vectors:

$$\operatorname{span}\{x_1, x_2, \ldots, x_k\} = \{\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k \mid \alpha_i \in \mathbb{R}, i = [1, k]\}$$

Vectors

- Span of a set of vectors:

$$\operatorname{span}\{x_1, x_2, \ldots, x_k\} = \{\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k \mid \alpha_i \in \mathbb{R}, i = [1, k]\}$$

- Linear independence:

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k = 0 \quad \Rightarrow \quad \alpha_i = 0,\ \forall i$$

Vectors

- Span of a set of vectors:

$$\operatorname{span}\{x_1, x_2, \ldots, x_k\} = \{\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k \mid \alpha_i \in \mathbb{R}, i = [1, k]\}$$

- Linear independence:

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k = 0 \quad \Rightarrow \quad \alpha_i = 0, \ \forall i$$

- How does k compare to p, the vector dimension?

Vectors

- Span of a set of vectors:

$$\text{span}\{x_1, x_2, \ldots, x_k\} = \{\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k \mid \alpha_i \in \mathbb{R}, i = [1, k]\}$$

- Linear independence:

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k = 0 \quad \Rightarrow \quad \alpha_i = 0, \ \forall i$$

- How does k compare to p, the vector dimension?

- Inner product:

$$x^\top y = \langle x, y \rangle = \sum_{i=1}^{p} x_i \cdot y_i$$

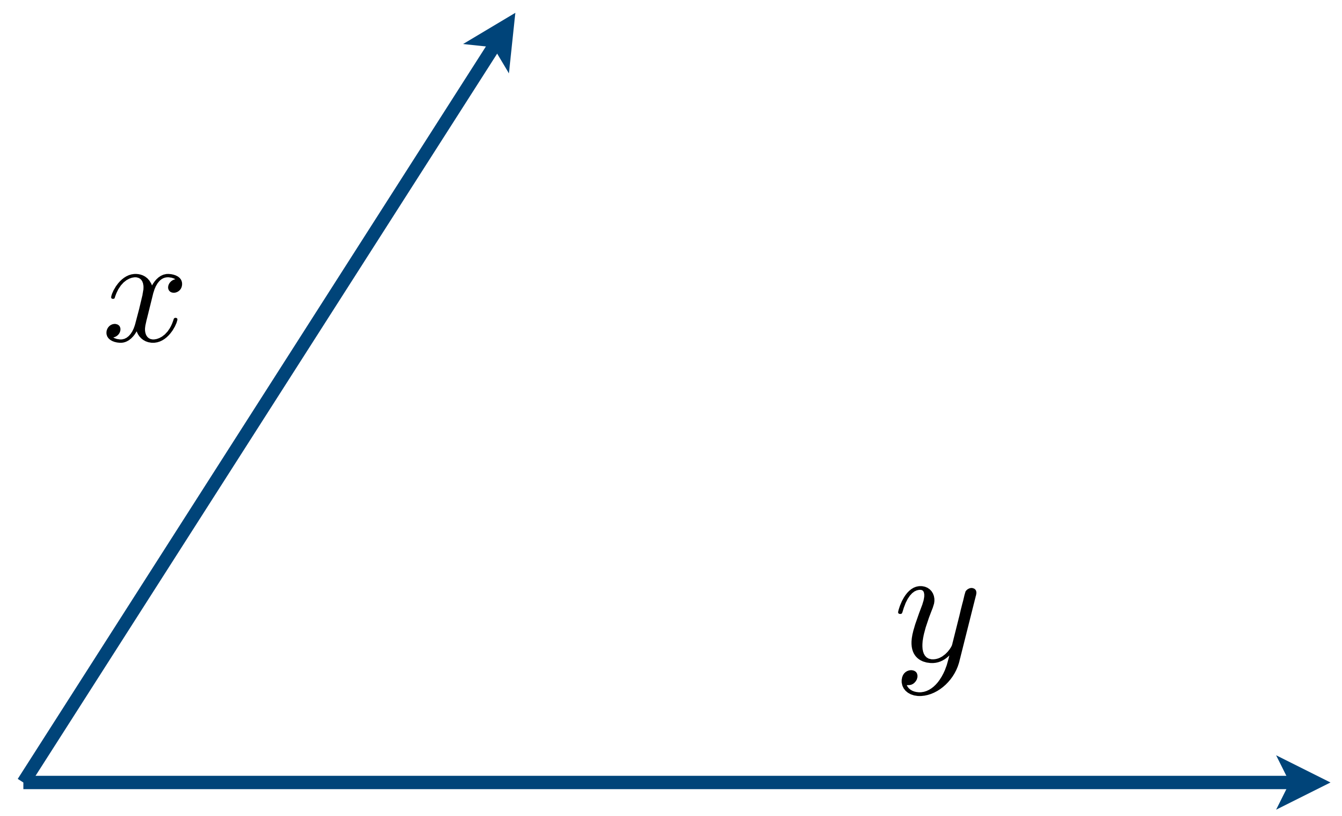
Vectors

- Inner production interpretation:

$$\langle x, y \rangle = \|x\| \cdot \|y\| \cdot \cos\theta$$

Vectors

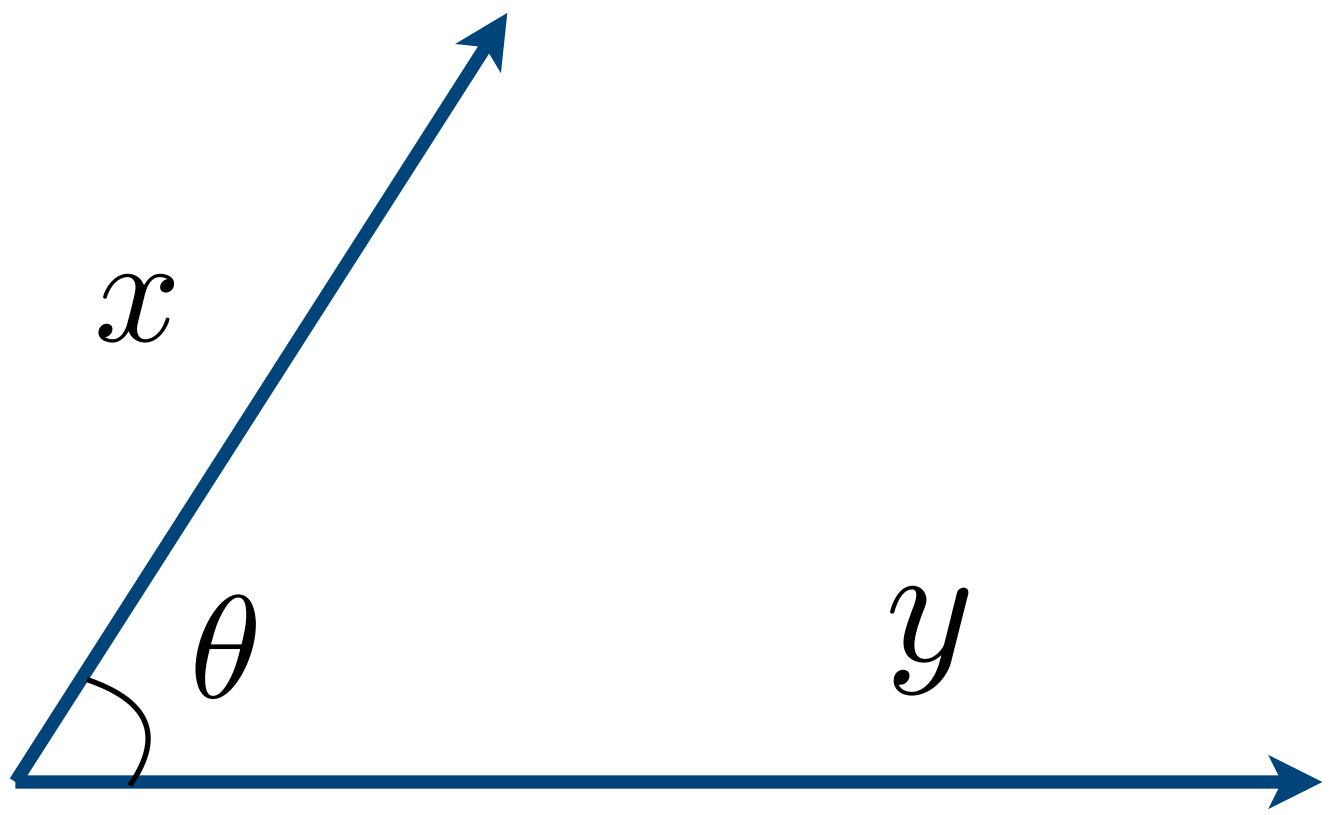
– Inner production interpretation:



$$\langle x, y \rangle = \|x\| \cdot \|y\| \cdot \cos\theta$$

Vectors

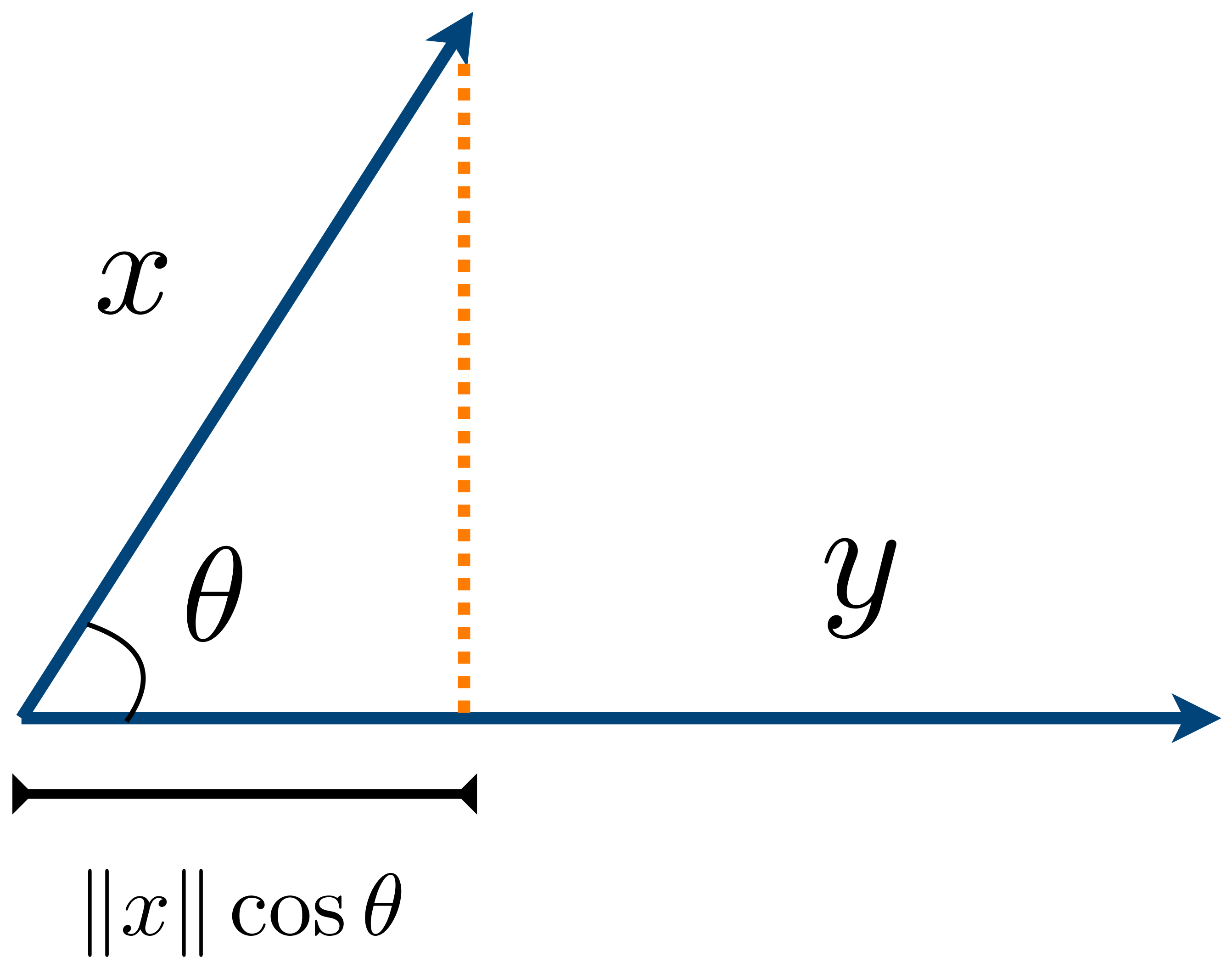
– Inner production interpretation:



$$\langle x, y \rangle = \|x\| \cdot \|y\| \cdot \cos\theta$$

Vectors

– Inner production interpretation:



$$\langle x, y \rangle = \|x\| \cdot \|y\| \cdot \cos\theta$$

Vectors

– Norms = notion of distance in multiple dimensions

Properties:

$$
\begin{aligned}
\|x\| &\geq 0, \forall x \in \mathbb{R}^p \\
\|x\| &= 0, \text{ iff } x = 0 \\
\|\alpha x\| &= |\alpha| \|x\|, \forall \alpha \in \mathbb{R} \\
\|x + y\| &\leq \|x\| + \|y\| \qquad \text{(Triangle inequality)} \\
|x^\top y| &\leq \|x\| \|y\| \qquad \text{(Cauchy–Schwarz)}
\end{aligned}
$$

Vectors

- Norms = notion of distance in multiple dimensions

Properties:

$$\|x\| \geq 0, \forall x \in \mathbb{R}^p$$

$$\|x\| = 0, \text{ iff } x = 0$$

$$\|\alpha x\| = |\alpha| \|x\|, \forall \alpha \in \mathbb{R}$$

$$\|x + y\| \leq \|x\| + \|y\|$$ (Triangle inequality)

$$|x^\top y| \leq \|x\| \|y\|$$ (Cauchy–Schwarz)

- Standard vector norms:

$$\|x\|_2 = \sqrt{\sum_i x_i^2} \qquad \|x\|_1 = \sum_i |x_i| \qquad \|x\|_\infty = \max_i |x_i|$$

Vectors

- Norms = notion of distance in multiple dimensions

Properties:

$$
\begin{aligned}
\|x\| &\geq 0, \forall x \in \mathbb{R}^p \\
\|x\| &= 0, \text{ iff } x = 0 \\
\|\alpha x\| &= |\alpha|\|x\|, \forall \alpha \in \mathbb{R} \\
\|x+y\| &\leq \|x\| + \|y\| && \text{(Triangle inequality)} \\
|x^\top y| &\leq \|x\|\|y\| && \text{(Cauchy–Schwarz)}
\end{aligned}
$$

- Standard vector norms:

$$\|x\|_2 = \sqrt{\sum_i x_i^2} \qquad \|x\|_1 = \sum_i |x_i| \qquad \|x\|_\infty = \max_i |x_i|$$

- Famous wanna-be norms: $\|x\|_0 = \operatorname{card}(x)$

Example – Sparse projection

– Find:

$$\widehat{x} \in \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \quad \|x - y\|_2^2 \qquad \text{for given } \; y \in \mathbb{R}^p$$

$$\text{s.t.} \quad \|x\|_0 \leq k < p$$

Example – Sparse projection

– Find: $\widehat{x} \in \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \quad \|x - y\|_2^2$ for given $y \in \mathbb{R}^p$

$\quad$ s.t. $\quad \|x\|_0 \leq k < p$

1. What is this problem?
2. How to solve this problem?

– Step 1: Assume we know $\mathcal{S} := \operatorname{supp}(\widehat{x}) \subset \{1, \ldots, p\}$

$$\widehat{x}_{\mathcal{S}} = y_{\mathcal{S}}$$

Example – Sparse projection

– Find:

$$\widehat{x} \in \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \quad \|x - y\|_2^2 \qquad \text{for given } y \in \mathbb{R}^p$$

$$\text{s.t.} \quad \|x\|_0 \leq k < p$$

– Step 1: Assume we know $\mathcal{S} := \operatorname{supp}(\widehat{x}) \subset \{1, \ldots, p\}$

$$\widehat{x}_{\mathcal{S}} = y_{\mathcal{S}}$$

– Step 2: Transform original problem

$$\mathcal{S} \in \underset{\operatorname{card}(\mathcal{S}) \leq k}{\operatorname{argmin}} \quad \|y_{\mathcal{S}} - y\|_2^2$$

Example – Sparse projection

– Find:

$$\widehat{x} \in \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \quad \|x - y\|_2^2 \qquad \text{for given } y \in \mathbb{R}^p$$
$$\text{s.t.} \quad \|x\|_0 \leq k < p$$

– Step 1: Assume we know $\mathcal{S} := \operatorname{supp}(\widehat{x}) \subset \{1, \ldots, p\}$

$$\widehat{x}_{\mathcal{S}} = y_{\mathcal{S}}$$

– Step 2: Transform original problem

$$\mathcal{S} \in \underset{\operatorname{card}(\mathcal{S}) \leq k}{\operatorname{argmin}} \quad \|y_{\mathcal{S}} - y\|_2^2 \qquad \equiv \qquad \mathcal{S} \in \underset{\operatorname{card}(\mathcal{S}) \leq k}{\operatorname{argmax}} \quad \|y_{\mathcal{S}}\|_2^2 := \sum_{i \in \mathcal{S}} y_i^2$$

Example – Sparse projection

– Find:
$$\widehat{x} \in \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \quad \|x - y\|_2^2 \qquad \text{for given } y \in \mathbb{R}^p$$
$$\text{s.t.} \quad \|x\|_0 \leq k < p$$

– Step 1: Assume we know $\mathcal{S} := \operatorname{supp}(\widehat{x}) \subset \{1, \dots, p\}$

$$\widehat{x}_{\mathcal{S}} = y_{\mathcal{S}}$$

– Step 2: Transform original problem

$$\mathcal{S} \in \underset{\operatorname{card}(\mathcal{S}) \leq k}{\operatorname{argmin}} \quad \|y_{\mathcal{S}} - y\|_2^2 \qquad \equiv \qquad \mathcal{S} \in \underset{\operatorname{card}(\mathcal{S}) \leq k}{\operatorname{argmax}} \quad \|y_{\mathcal{S}}\|_2^2 := \sum_{i \in \mathcal{S}} y_i^2$$

– Step 3: Find efficient ways to solve the last part!

Matrices

- Matrix in m, n dimensions: $A \in \mathbb{R}^{m \times n}$

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}$$

Matrices

- Matrix in m, n dimensions: $A \in \mathbb{R}^{m \times n}$

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}$$

- Names: Square, tall, fat, zero, identity, diagonal

Matrices

- Matrix in m, n dimensions: $A \in \mathbb{R}^{m\times n}$

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}$$

- Names: Square, tall, fat, zero, identity, diagonal
- Properties:

$$A + B = B + A,\ \forall A, B \in \mathbb{R}^{m\times n}$$
$$(A + B) + C = A + (B + C),\ \forall A, B, C \in \mathbb{R}^{m\times n}$$
$$A + 0 = 0 + A,\ \forall A \in \mathbb{R}^{m\times n}$$
$$(A + B)^\top = A^\top + B^\top,\ \forall A, B \in \mathbb{R}^{m\times n}$$

Matrices

– Matrix multiplication： $C = AB$ where $C \in \mathbb{R}^{m \times p}$, $A \in \mathbb{R}^{m \times n}$, and $B \in \mathbb{R}^{n \times p}$

$$AB = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} \cdot \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ B_{21} & B_{22} & \cdots & B_{2p} \\ \vdots & & \ddots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{np} \end{bmatrix}$$

Matrices

- Matrix multiplication: $C = AB$ where $C \in \mathbb{R}^{m\times p}$, $A \in \mathbb{R}^{m\times n}$, and $B \in \mathbb{R}^{n\times p}$

$$AB = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} \cdot \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ B_{21} & B_{22} & \cdots & B_{2p} \\ \vdots & & \ddots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{np} \end{bmatrix}$$

- Special cases: vector inner product, matrix–vector mult., outer product

Matrices

– Matrix multiplication： $C = AB$ where $C \in \mathbb{R}^{m \times p}$, $A \in \mathbb{R}^{m \times n}$, and $B \in \mathbb{R}^{n \times p}$

$$AB = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} \cdot \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ B_{21} & B_{22} & \cdots & B_{2p} \\ \vdots & & \ddots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{np} \end{bmatrix}$$

– Special cases： vector inner product, matrix–vector mult., outer product

– Properties：

$$(AB)C = A(BC),\ \forall A, B, C$$

$$\alpha(AB) = (\alpha A)B,\ \forall A, B$$

$$A(B + C) = AB + AC,\ \forall A, B, C$$

$$(AB)^\top = B^\top A^\top,\ \forall, A, B$$

$$AB \neq BA$$

Matrices

- Inner product:

$$\langle A, B \rangle = \mathrm{Tr}(A^\top B) = \mathrm{Tr}(B^\top A), \forall A, B \in \mathbb{R}^{m \times n}$$

Matrices

- Inner product:

$$\langle A, B \rangle = \mathrm{Tr}(A^\top B) = \mathrm{Tr}(B^\top A), \forall A, B \in \mathbb{R}^{m \times n}$$

- Rank of a matrix: maximum # of independent columns or rows

Matrices

- Inner product:

$$\langle A, B \rangle = \mathrm{Tr}(A^\top B) = \mathrm{Tr}(B^\top A), \forall A, B \in \mathbb{R}^{m \times n}$$

- Rank of a matrix: maximum # of independent columns or rows

- Nullspace of a matrix: $\{x \mid Ax = 0\}$

Matrices

- Inner product：

$$\langle A, B\rangle = \mathrm{Tr}(A^\top B) = \mathrm{Tr}(B^\top A), \forall A, B \in \mathbb{R}^{m\times n}$$

- Rank of a matrix：maximum # of independent columns or rows

- Nullspace of a matrix：　$\{x \mid Ax = 0\}$

- Positive semi-definite matrices：　$A \succeq 0$

1. $A \in \mathbb{R}^{n\times n}$
2. A is symmetric
3. $x^\top Ax \geq 0,\ \forall x \in \mathbb{R}^n,\ x \neq 0$

Matrices

- Matrix singular value decomposition: $A \in \mathbb{R}^{m \times n}$

$$A = U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i u_i v_i^\top,\ U \in \mathbb{R}^{m\times r}, \Sigma \in \mathbb{R}^{r\times r}, V \in \mathbb{R}^{n\times r} \qquad r \le \{m, n\}$$

Matrices

- Matrix singular value decomposition: $A \in \mathbb{R}^{m \times n}$

$$A = U \Sigma V^\top = \sum_{i=1}^{r} \sigma_i u_i v_i^\top,\ U \in \mathbb{R}^{m \times r}, \Sigma \in \mathbb{R}^{r \times r}, V \in \mathbb{R}^{n \times r} \qquad r \leq \{m, n\}$$

- $\mathrm{rank}(A) = r \leq \min\{m, n\}$

Matrices

- Matrix singular value decomposition: $A \in \mathbb{R}^{m \times n}$

$$A = U \Sigma V^\top = \sum_{i=1}^{r} \sigma_i u_i v_i^\top, \; U \in \mathbb{R}^{m \times r}, \Sigma \in \mathbb{R}^{r \times r}, V \in \mathbb{R}^{n \times r} \qquad r \leq \{m, n\}$$

- $\mathrm{rank}(A) = r \leq \min\{m, n\}$

- $u_i \in \mathbb{R}^m, v_i \in \mathbb{R}^n$ are the left and right singular vectors

Matrices

- Matrix singular value decomposition:　$A \in \mathbb{R}^{m\times n}$

$$A = U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i u_i v_i^\top,\ U \in \mathbb{R}^{m\times r}, \Sigma \in \mathbb{R}^{r\times r}, V \in \mathbb{R}^{n\times r} \qquad r \leq \{m, n\}$$

- $\mathrm{rank}(A) = r \leq \min\{m, n\}$
- $u_i \in \mathbb{R}^m, v_i \in \mathbb{R}^n$ are the left and right singular vectors
- $\Sigma = \texttt{diag}(\sigma_1, \ldots, \sigma_r)$ contains singular values where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$

Matrices

- Matrix singular value decomposition: $A \in \mathbb{R}^{m \times n}$

$$A = U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i u_i v_i^\top, \; U \in \mathbb{R}^{m\times r}, \Sigma \in \mathbb{R}^{r\times r}, V \in \mathbb{R}^{n\times r} \qquad r \leq \{m, n\}$$

- $\mathrm{rank}(A) = r \leq \min\{m, n\}$
- $u_i \in \mathbb{R}^m, v_i \in \mathbb{R}^n$ are the left and right singular vectors
- $\Sigma = \mathtt{diag}(\sigma_1, \ldots, \sigma_r)$ contains singular values where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$
- Left and right singular vectors are orthogonal: $U^\top U = I$ and $V^\top V = I$

Matrices

- Norms:

$$\|A\|_F = \sqrt{\sum_{ij} A_{ij}^2}$$

(Frobenius norm)

$$\|A\|_* = \sum_{i}^{r} \sigma_i$$

(Nuclear norm)

$$\|A\|_2 = \max_i \sigma_i$$

(Spectral norm)

..there are more norms to worry about (e.g., operator norms)
but we will skip them here..

Matrices

– Matrix inverses are defined on square matrices

Matrices

- Matrix inverses are defined on square matrices
- Matrix inverse definition as a collection of properties:
  - A is full rank

Matrices

- Matrix inverses are defined on square matrices
- Matrix inverse definition as a collection of properties:
  - A is full rank
  - A has empty nullspace

Matrices

- Matrix inverses are defined on square matrices
- Matrix inverse definition as a collection of properties:
  - A is full rank
  - A has empty nullspace
  - The equation $Ax = 0$ has only the trivial solution $x = 0$

Matrices

- Matrix inverses are defined on square matrices
- Matrix inverse definition as a collection of properties:
  - A is full rank
  - A has empty nullspace
  - The equation $Ax = 0$ has only the trivial solution $x = 0$
  - The linear system $Ax = b$ has a unique solution

Matrices

- Matrix inverses are defined on square matrices
- Matrix inverse definition as a collection of properties:
  - A is full rank
  - A has empty nullspace
  - The equation $Ax = 0$ has only the trivial solution $x = 0$
  - The linear system $Ax = b$ has a unique solution
  - The columns and rows of A are linearly independent

Matrices

- Matrix inverses are defined on square matrices
- Matrix inverse definition as a collection of properties:
  - A is full rank
  - A has empty nullspace
  - The equation $Ax = 0$ has only the trivial solution $x = 0$
  - The linear system $Ax = b$ has a unique solution
  - The columns and rows of A are linearly independent
  - There exists a square matrix, A^{-1}, such that $A^{-1}A = AA^{-1} = I$

Example – Low-rank projection

– Find:

$$\begin{aligned} \widehat{X} \in \min_{X} \quad & \frac{1}{2}\|X - Y\|_F^2 \\ \text{s.t.} \quad & \operatorname{rank}(X) \le r \end{aligned}$$

for given $Y \in \mathbb{R}^{m \times n}$

Example – Low-rank projection

- Find: $$\widehat{X} \in \min_X \quad \frac{1}{2}\|X - Y\|_F^2 \qquad \text{for given } Y \in \mathbb{R}^{m\times n}$$
$$\text{s.t.} \quad \operatorname{rank}(X) \le r$$

- Step 1: Observe $$\|X - Y\|_F^2 = \|X - U\Sigma V^\top\|_F^2$$
$$= \|U^\top X V - \Sigma\|_F^2$$

Example – Low–rank projection

– Find:

$$\widehat{X} \in \min_{X} \quad \frac{1}{2}\|X - Y\|_F^2 \qquad \text{for given } \; Y \in \mathbb{R}^{m\times n}$$

$$\text{s.t.} \quad \operatorname{rank}(X) \le r$$

– Step 1: Observe $\|X - Y\|_F^2 = \|X - U\Sigma V^\top\|_F^2$

$$= \|U^\top X V - \Sigma\|_F^2$$

– Step 2: Transform original problem

$$\min_{X} \quad \frac{1}{2}\|C - \Sigma\|_F^2 \qquad \text{where} \qquad C = U^\top X V$$

$$\text{s.t.} \quad \operatorname{rank}(C) \le r$$

Example – Low-rank projection

– Find: $$\widehat{X} \in \min_{X} \quad \frac{1}{2}\|X - Y\|_F^2 \qquad \text{s.t.} \quad \operatorname{rank}(X) \le r$$ for given $Y \in \mathbb{R}^{m\times n}$

– Step 1: Observe $$\|X - Y\|_F^2 = \|X - U\Sigma V^\top\|_F^2 = \|U^\top X V - \Sigma\|_F^2$$

– Step 2: Transform original problem

$$\min_{X} \quad \frac{1}{2}\|C - \Sigma\|_F^2 \qquad \text{s.t.} \quad \operatorname{rank}(C) \le r$$

where $C = U^\top X V$

– Step 3: Optimal C is diagonal with singular values from Σ

Example – Low-rank projection

- Find: $$\widehat{X} \in \min_{X} \quad \tfrac{1}{2}\|X - Y\|_F^2 \qquad \text{s.t.} \quad \operatorname{rank}(X) \leq r$$ for given $Y \in \mathbb{R}^{m\times n}$

- Step 1: Observe $$\|X - Y\|_F^2 = \|X - U\Sigma V^\top\|_F^2 = \|U^\top X V - \Sigma\|_F^2$$

- Step 2: Transform original problem
$$\min_{X} \quad \tfrac{1}{2}\|C - \Sigma\|_F^2 \qquad \text{s.t.} \quad \operatorname{rank}(C) \leq r$$
where $C = U^\top X V$

- Step 3: Optimal C is diagonal with singular values from Σ

How does this relate to sparse projection?

Demo

Papers to review – due next Thursday

(Select one of the following papers)

- “Efficient projections onto the ℓ_1–norm ball for learning in high dimensions”, Duchi et al., 2008.

- “Stay on path: PCA along graph paths”, Asteris et al., 2015.

- “CUR matrix decompositions for improved data analysis”, Mahoney and Drineas, 2008.

- “Simple and Deterministic Matrix Sketching”, Liberty, 2013.

(Rule: as you read, think of extensions – feel free to find me for more discussions)

Conclusion

- We have set up background and notation w.r.t. linear algebra
- We saw a toy example where non−convex operations happen

Conclusion

- We have set up background and notation w.r.t. linear algebra
- We saw a toy example where non-convex operations happen

Next lecture

- Brief introduction to convex optimization and related topics

Intuition for inner product