Randomized Numerical Linear Algebra
Lecture 1

F. M. Faulstich

01/09/2024

General info

- Course Website: https://fabianfaulstich.github.io/MATH6950_2024/
- Homework assignments:
  1. Submission through Gradescope.
  2. Gradescope will close on the due date at 11:59 p.m.
     No late submissions!
  3. Everyone get **one** joker
  4. zero-tolerance policy regarding cheating
- Lectures:
  1. hybrid slides and blackboard lecturing
  2. the slide part will be made available online
  3. Code presented and used in class will be made available online
- Programming assignments:
  1. P.A.s will carry substantial points in each homework
  2. To gain full credit you have to show exploration and clear reasoning

Planned course outlook

1. Review of numerical linear algebra and probability theory
2. Low-rank approximations and randomness

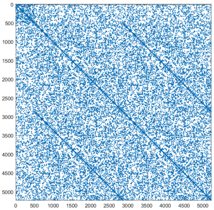
⋮

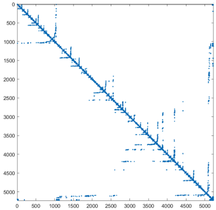
What to expect

- The course is centered primarily on computational aspects
- Aimed at equipping you with a robust set of computational skills
- Proofs will be included (mostly) at a high level
- The course aims to impart an understanding of the key ideas behind proofs rather than delving into exhaustive, fully worked-out proofs

numerical linear algebra

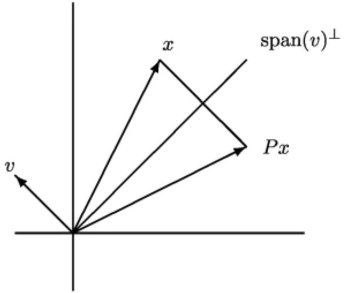
- solving dense and sparse linear systems





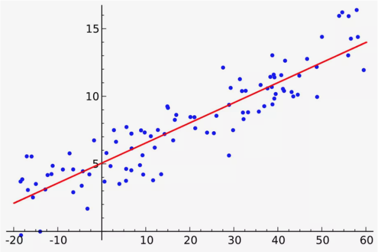
numerical linear algebra

- solving dense and sparse linear systems
- orthogonalization, least square & (Tikhonov) regularization



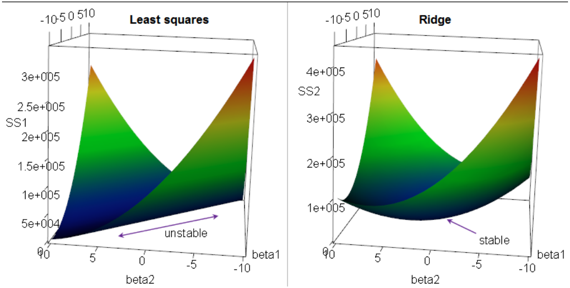
numerical linear algebra

- solving dense and sparse linear systems
- orthogonalization, least square & (Tikhonov) regularization



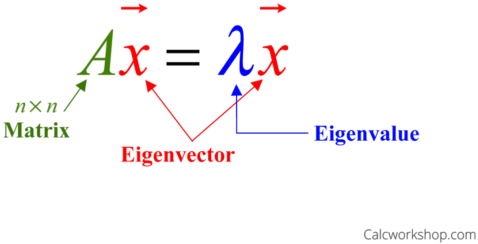
numerical linear algebra

- solving dense and sparse linear systems
- orthogonalization, least square & (Tikhonov) regularization



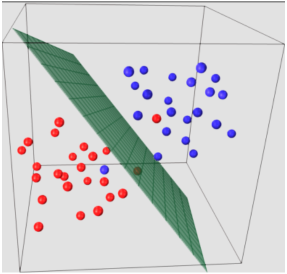
numerical linear algebra

- solving dense and sparse linear systems
- orthogonalization, least square & (Tikhonov) regularization
- determination of eigenvalues & eigenvectors, invariant subspaces



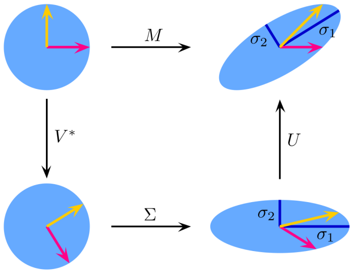
numerical linear algebra

- solving dense and sparse linear systems
- orthogonalization, least square & (Tikhonov) regularization
- determination of eigenvalues & eigenvectors, invariant subspaces



numerical linear algebra

- solving dense and sparse linear systems
- orthogonalization, least square & (Tikhonov) regularization
- determination of eigenvalues & eigenvectors, invariant subspaces
- singular value decomposition (SVD)



$$M = U \cdot \Sigma \cdot V^*$$

Numerical linear algebra

What is the problem?

$\Rightarrow$ Large scaling problems!

Example:
Given an $m \times n$ matrix $\mathbf{A}$ where both m and n are large, the singular value decomposition will require memory and time which is superlinear in m and n

$$\mathcal{O}(4mn^2 - \frac{4}{3}n^3) \qquad (GK)$$

Randomized algorithms bring this down to m and k, where k is the rank

$$\mathcal{O}(2kn^2 + 2n^3) \qquad (LHC)$$

Randomized algorithms

1. Monte Carlo (M.C.) algorithms (∼1940)
   - Last resort methods
   - M.C. converges slowly
   - Hesitation: Two runs should produce the same number

Randomized algorithms

1. Monte Carlo (M.C.) algorithms (∼1940)
   · Last resort methods
   · M.C. converges slowly
   · Hesitation: Two runs should produce the same number
2. Randomized algorithms in numerical linear algebra (∼1980)
   · Power method, random initialization (Dixon 1983)
   · M. C. methods for trace estimation (Girard 1989 & Hutchinson 1990)
   · Randomized transformations can avoid pivoting steps in Gaussian elimination (Parker 1995)

Randomized algorithms

1. Monte Carlo (M.C.) algorithms (∼1940)
   · Last resort methods
   · M.C. converges slowly
   · Hesitation: Two runs should produce the same number
2. Randomized algorithms in numerical linear algebra (∼1980)
   · Power method, random initialization (Dixon 1983)
   · M. C. methods for trace estimation (Girard 1989 & Hutchinson 1990)
   · Randomized transformations can avoid pivoting steps in Gaussian elimination (Parker 1995)
3. Practical randomized algorithms for low-rank matrix approximation and least-squares problems (mid-2000s)
   · First computational evidence that randomized algorithms outperform classical NLA algorithms for particular classes of problems

Review: Classical Numerical Linear Algebra

- L. N. Trefethen and D. Bau III, Numerical linear algebra, Vol. 50, SIAM
- G. Stewart, Matrix Algorithms Volume 1: Basic Decompositions, SIAM
- G. W. Stewart, Matrix algorithms volume 2: eigensystems, Vol. 2, SIAM
- G. H. Golub and C. F. Van Loan, Matrix computations
- R. A. Horn and C. R. Johnson, Matrix analysis
- R. Bhatia, Matrix analysis, Vol. 169 of Graduate Texts in Mathematics

Basics – Notation I

- Algebraic field: $\mathbb{C}$, $\mathbb{R}$, $\mathbb{F}$
- Scalars: $a, b, ...$ or $\alpha, \beta, ...$
- Vectors are elements of $\mathbb{F}^n$, $n \in \mathbb{N}$: $\mathbf{a}, \mathbf{b}, ...$ or $\boldsymbol{\alpha}, \boldsymbol{\beta}, ...$
- Special vectors $\mathbf{0}, \mathbf{1}, \boldsymbol{\delta}_i \in \mathbb{F}^n$:

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad \boldsymbol{\delta}_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- Vector element:

$$(\mathbf{a})_i = \mathbf{a}(i) \qquad i\text{th coordinate of } \mathbf{a}$$

- Colon notation: $(\mathbf{a})_{1:i} = \mathbf{a}(1 : i)$

Basics – Notation II

- A matrix is an element of $\mathbb{F}^{m\times n}$, $m, n \in \mathbb{N}$:
  $\mathbf{A}, \mathbf{B}, \ldots$ or $\boldsymbol{\Lambda}, \boldsymbol{\Delta}, \ldots$
- Matrix element:

$$(\mathbf{A})_{ij} = \mathbf{A}(ij) \qquad (i,j)\text{th coordinate of } \mathbf{A}$$

- Special matrices $\mathbf{0}, \mathbf{I} \in \mathbb{F}^{m\times n}$:

$$\mathbf{0} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \quad \mathbf{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

- Colon notation: $(\mathbf{A})_{i:} \equiv i$th row and $(\mathbf{A})_{:j}$ jth column of $\mathbf{A}$

Basics – Notation III

- $*$ is the conjugate transpose
- $\mathbb{H}_n = \mathbb{H}_n(\mathbb{F}) = \{\mathbf{A} \in \mathbb{F}^{n\times n} \ : \ \mathbf{A} = \mathbf{A}^*\}$
- $\dagger$ is the Moor–Penrose (pseudo)inverse:

Basics – Notation III

- $*$ is the conjugate transpose
- $\mathbb{H}_n = \mathbb{H}_n(\mathbb{F}) = \{\mathbf{A} \in \mathbb{F}^{n\times n} \;:\; \mathbf{A} = \mathbf{A}^*\}$
- $\dagger$ is the Moor–Penrose (pseudo)inverse:
  $\mathbf{A}^\dagger$ is the MP inverse of $\mathbf{A}$ iff

$$(i) \quad \mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A} \qquad\qquad (iii) \quad (\mathbf{A}\mathbf{A}^\dagger)^* = \mathbf{A}\mathbf{A}^\dagger$$

$$(ii) \quad \mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger \qquad\qquad (iv) \quad (\mathbf{A}^\dagger\mathbf{A})^* = \mathbf{A}^\dagger\mathbf{A}$$

Basics – Notation III

- $*$ is the conjugate transpose
- $\mathbb{H}_n = \mathbb{H}_n(\mathbb{F}) = \{\mathbf{A} \in \mathbb{F}^{n\times n} \;:\; \mathbf{A} = \mathbf{A}^*\}$
- $\dagger$ is the Moor–Penrose (pseudo)inverse:
  $\mathbf{A}^\dagger$ is the MP inverse of $\mathbf{A}$ iff

$$
\begin{array}{llll}
(i) & \mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A} & (iii) & (\mathbf{A}\mathbf{A}^\dagger)^* = \mathbf{A}\mathbf{A}^\dagger \\
(ii) & \mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger & (iv) & (\mathbf{A}^\dagger\mathbf{A})^* = \mathbf{A}^\dagger\mathbf{A}
\end{array}
$$

  If $\mathbf{A}$ has full column rank, then $\mathbf{A}^\dagger = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$
  If $\mathbf{A}$ attains an inverse then

$$\mathbf{A}^\dagger = \mathbf{A}^{-1}$$

Eigenvalues and singular values

- PSD$= \{\mathbf{A} \in \mathbb{H}_n \ : \ \mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0 \text{ for } \mathbf{x} \neq 0\}$
- PD$= \{\mathbf{A} \in \mathbb{H}_n \ : \ \mathbf{x}^\top \mathbf{A}\mathbf{x} > 0 \text{ for } \mathbf{x} \neq 0\}$
- $\preccurlyeq$ denotes the semidefinte order on $\mathbb{H}_n$, i.e.

$$\mathbf{A} \preccurlyeq \mathbf{B} \Leftrightarrow 0 \preccurlyeq \mathbf{B} - \mathbf{A}$$

- $\prec$ denotes the definte order on $\mathbb{H}_n$, i.e.

$$\mathbf{A} \prec \mathbf{B} \Leftrightarrow 0 \prec \mathbf{B} - \mathbf{A}$$

- Eigenvalues of $\mathbf{A} \in \mathbb{H}_n$: $\lambda_1 \geq \lambda_2 \geq ...$
- Singular values of $\mathbf{A} \in \mathbb{F}^{m \times n}$: $\sigma_1 \geq \sigma_2 \geq ...$
- Let $f : \mathbb{R} \to \mathbb{R}$. We extend f to spectral function $f : \mathbb{H}_n \to \mathbb{H}_n$

$$f(\mathbf{A}) := \sum_{i=1}^{n} f(\lambda_i)\mathbf{u}_i\mathbf{u}_i^* \quad \text{where} \quad \mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^*$$

Inner products and geometry I

- Equip $\mathbb{F}^n$ with standard scalar product and associated ℓ^2-norm. Let $\mathbf{a}, \mathbf{b} \in \mathbb{F}^n$ then

$$\langle \mathbf{a}, \mathbf{b} \rangle := \mathbf{a} \cdot \mathbf{b} = \mathbf{a}^* \mathbf{b} = \sum_{i=1}^{n} (\mathbf{a})_i^* (\mathbf{b})_i$$

  and

$$\|\mathbf{a}\|^2 := \langle \mathbf{a}, \mathbf{a} \rangle$$

- Unit sphere in $\mathbb{F}^n$: $\mathbb{S}^{n-1} = \mathbb{S}^{n-1}(\mathbb{F})$

Inner products and geometry II

- The trace of $\mathbf{A} \in \mathbb{F}^{n \times n}$:

$$\mathrm{Tr}(\mathbf{A}) = \mathrm{trace}(\mathbf{A}) = \sum_{i=1}^{n} (\mathbf{A})_{ii}$$

  Nonlinear functions bind before the trace.

- Equip $\mathbb{F}^{m \times n}$ with the standard trace inner product and Frobenius norm:
  Let $\mathbf{A}, \mathbf{B} \in \mathbb{F}^{m \times n}$ then

$$\langle \mathbf{A}, \mathbf{B} \rangle := \mathrm{Tr}(\mathbf{A}^* \mathbf{B})$$

  and

$$\|\mathbf{A}\|_F^2 = \langle \mathbf{A}, \mathbf{A} \rangle$$

- $\mathbf{U} \in \mathbb{F}^{m \times n}$ is orthonormal iff $\mathbf{U}^* \mathbf{U} = \mathbf{I}_n$.
  $\mathbf{U}$ is unitary ($\mathbb{F} = \mathbb{C}$) or orthogonal ($\mathbb{F} = \mathbb{R}$) if $m = n$.

Matrix norms I

- Let $\|\cdot\|_\alpha$ be a norm on $\mathbb{F}^n$ and $\|\cdot\|_\beta$ be a norm on $\mathbb{F}^m$. Then

$$\|\cdot\|_{\alpha,\beta} : \mathbb{F}^{m\times n} \to \mathbb{R}; \mathbf{A} \mapsto \sup_{\substack{\mathbf{x}\in\mathbb{F}^n \\ \|\mathbf{x}\|_\alpha \neq 0}} \frac{\|\mathbf{A}\mathbf{x}\|_\beta}{\|\mathbf{x}\|_\alpha}$$

  Induces a norm on $\mathbb{F}^{m\times n}$.
- Alternatively, we may define any function

$$\|\cdot\| : \mathbb{F}^{m\times n} \to \mathbb{R}$$

  that fulfills:
  1. $0 \leq \|\mathbf{A}\|, \ \forall \mathbf{A} \in \mathbb{F}^{m\times n}$ and $\|\mathbf{A}\| = 0 \Leftrightarrow \mathbf{A} = \mathbf{0}$
  2. $\|a\mathbf{A}\| = |a|\|\mathbf{A}\|, \ \forall \mathbf{A} \in \mathbb{F}^{m\times n}$, and $\forall a \in \mathbb{F}$
  3. $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|, \ \forall \mathbf{A}, \mathbf{B} \in \mathbb{F}^{m\times n}$

Matrix norms II

Several matrix norms will be used. Let $\mathbf{A} \in \mathbb{F}^{m \times n}$

- The unadorned norm $\|\cdot\|$ is the spectral norm

$$\|\mathbf{A}\| = \sigma_1 = \|\mathbf{A}\|_{\ell^2}$$

- $\|\cdot\|_*$ is the nuclear/trace norm

$$\|\mathbf{A}\|_* = \sum_{k=1}^{\min(m,n)} \sigma_k$$

- $\|\cdot\|_F$ is the Frobenius norm

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} |(A)_{ij}|^2 = \sum_{k=1}^{\min(m,n)} \sigma_k^2 = \mathrm{Tr}(\mathbf{A}^* \mathbf{A})$$

Matrix norms III

- $\|\cdot\|_p$ is the Schatten p-norm for $p \in [1, \infty]$

$$\|\mathbf{A}\|_p = \left(\sum_{k=1}^{\min(m,n)} \sigma_k^p \right)^{\frac{1}{p}}$$

- $\|\cdot\|_{K,p}$ is the Ky Fan p-norm for $p \leq \min(m, n)$

$$\|\mathbf{A}\|_{K,p} = \sum_{k=1}^{p} \sigma_k$$

Note:

$$\|\cdot\|_* = \|\cdot\|_{K,\min(m,n)} = \|\cdot\|_1$$

$$\|\cdot\|_F = \|\cdot\|_2$$

$$\|\cdot\| = \|\cdot\|_{K,1} = \|\cdot\|_\infty$$

Intrinsic Dimension

Let $\mathbf{A} \in \mathbb{H}_n$ be PSD. We define the intrinsic dimension as

$$\operatorname{intdim}(\mathbf{A}) := \frac{\operatorname{Tr}(\mathbf{A})}{\|\mathbf{A}\|}$$

Intrinsic Dimension

Let $\mathbf{A} \in \mathbb{H}_n$ be PSD. We define the intrinsic dimension as

$$\operatorname{intdim}(\mathbf{A}) := \frac{\operatorname{Tr}(\mathbf{A})}{\|\mathbf{A}\|}$$

Note that for $\mathbf{A}$ non-zero:

$$1 \leq \operatorname{intdim}(\mathbf{A}) \leq \operatorname{rank}(\mathbf{A})$$

The upper bound is saturated if $\mathbf{A}$ is an orthogonal projector

Intrinsic Dimension

Let $\mathbf{A} \in \mathbb{H}_n$ be PSD. We define the intrinsic dimension as

$$\operatorname{intdim}(\mathbf{A}) := \frac{\operatorname{Tr}(\mathbf{A})}{\|\mathbf{A}\|}$$

Note that for $\mathbf{A}$ non-zero:

$$1 \leq \operatorname{intdim}(\mathbf{A}) \leq \operatorname{rank}(\mathbf{A})$$

The upper bound is saturated if $\mathbf{A}$ is an orthogonal projector i.e.

1. $\mathbf{A} \in \mathbb{F}^{m \times m}$ and $\mathbf{A}^2 = \mathbf{A}$
2. $\mathbf{A}$ is projector and $\mathbf{A} \in \mathbb{H}_n$

The intrinsic rank can be interpreted as a continuous measure of the rank

Stable rank

Let $\mathbf{B} \in \mathbb{F}^{m \times n}$. We define the stable rank as

$$\operatorname{srank}(\mathbf{A}) := \operatorname{intdim}(\mathbf{B}^*\mathbf{B}) = \frac{\|\mathbf{B}\|_F^2}{\|\mathbf{B}\|^2}$$

Schur complement

Let

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \in \mathbb{F}^{m \times m}$$

with $\mathbf{A} \in \mathbb{F}^{n \times n}$.
If $\mathbf{D}$ is invertible the Schur complement of $\mathbf{D}$ in $\mathbf{M}$ is

$$\mathbf{M}/\mathbf{D} := \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$$

If $\mathbf{A}$ is invertible the Schur complement of $\mathbf{A}$ in $\mathbf{M}$ is

$$\mathbf{M}/\mathbf{A} := \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$$

The latter is used for Cholesky factorization ($\mathbf{M} \in \mathbb{H}_n$ and $\mathbf{A} \in \mathbb{F}^{1 \times 1}$).

What if $\mathbf{D}$ of $\mathbf{A}$ are singular or not square?

Generalized Schur complement

Let $\mathbf{M} \in \mathbb{F}^{m \times n}$ and

$$\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k) \subseteq [\![m]\!] \quad \text{and} \quad \boldsymbol{\alpha}^c = [\![m]\!] \setminus \boldsymbol{\alpha}$$

and

$$\boldsymbol{\beta} = (\beta_1, ..., \beta_\ell) \subseteq [\![n]\!] \quad \text{and} \quad \boldsymbol{\beta}^c = [\![n]\!] \setminus \boldsymbol{\beta}.$$

We denote

$$\mathbf{M}[\boldsymbol{\gamma}, \boldsymbol{\delta}]$$

the $(\boldsymbol{\gamma}, \boldsymbol{\delta})$-block in $\mathbf{M}$.
The Schur complement of $\mathbf{M}[\boldsymbol{\alpha}, \boldsymbol{\beta}]$ in $\mathbf{M}$ is

$$\mathbf{M}/\mathbf{M}[\boldsymbol{\alpha}, \boldsymbol{\beta}] = \mathbf{M}[\boldsymbol{\alpha}^c, \boldsymbol{\beta}^c] - \mathbf{M}[\boldsymbol{\alpha}^c, \boldsymbol{\beta}] \left(\mathbf{M}[\boldsymbol{\alpha}, \boldsymbol{\beta}]\right)^\dagger \mathbf{M}[\boldsymbol{\alpha}, \boldsymbol{\beta}^c]$$

F. Zhang, The Schur complement and its applications, Vol. 4 of Numerical Methods and Algorithms, Springer-Verlag, New York.

Approximation in the spectral norm

We will mostly establish spectral norm errors.

- Suppose $\mathbf{A} \in \mathbb{F}^{m\times n}$ and $\hat{\mathbf{A}} \in \mathbb{F}^{m\times n}$ is an approximation

$$\|\mathbf{A} - \hat{\mathbf{A}}\| \leq \varepsilon$$

  then

  1. $|\langle \mathbf{F}, \mathbf{A}\rangle - \langle \mathbf{F}, \hat{\mathbf{A}}\rangle| \leq \varepsilon \|\mathbf{F}\|_*$ for every matrix $\mathbf{F} \in \mathbb{F}^{m\times n}$
  2. $|\sigma_j(\mathbf{A}) - \sigma_j(\hat{\mathbf{A}})| \leq \varepsilon, \ \forall j$

How do spectral norm errors compare with Frobenius norm error measures?