Outline

- Monday 10-14, Liivi 2-225
- Topics
  - Introduction, statistical methods and their applications
  - Linear algebra and singular value decomposition
  - Basic optimization, Image processing and analysis
  - Clustering and applications
  - Compressed sensing
  - Text processing
  - Time series analysis and wavelets
Outline

- Motivation
- Linear Algebra
- Gram Schmidt Orthogonalization
- QR algorithm
- Singular Value Decomposition
- Higher order Singular Value Decomposition
Motivation

- Face recognition
- https://findface.ru/
- https://en.wikipedia.org/wiki/FindFace
- https://ru.wikipedia.org/wiki/FindFace
- https://github.com/kagan94/
  Face-recognition-via-SVD-and-PCA
Floating Point Arithmetic

- Many computers use binary to approximate decimal numbers
- Several standards in use, many use IEEE 754 floating point standard
- Format consists of exponent (e) and mantissa (m) so that a number is \((-1)^s \times b^e \times m\)
  - \(s = 0, 1, m = d_0 \cdot d_1 d_2...d_{p-1}\)
    - Single precision - binary 32 bit, \(b = 2, p = 24, \text{emax}=+127\)
    - Double precision - binary 64 bit, \(b = 2, p = 53, \text{emax}=+1023\)
    - Extended precision - binary 128 bit, \(b = 2, p = 113, \text{emax}=+16383\)
- There is also a specification for \(b = 10\)
- Different implementations of the standard will in most cases give comparable results. Can get differences which can sometimes be large.
- [http://www.ima.umn.edu/~arnold/disasters/patriot.html](http://www.ima.umn.edu/~arnold/disasters/patriot.html)
Condition Numbers

- Gives a way of understanding sensitivity of output to small changes in input.
- Important since exact input data is not possible to obtain in many problems, and if obtained, may be difficult to represent correctly in floating point format.
- Suppose data $d$ and output $x$, then a condition number $K$ is the maximum value of
  \[
  \frac{||\delta x||}{||\delta d||}
  \]
  where $\delta d$ is a perturbation in the allowed input data and $\delta x$ is the maximum resulting change in the output.
- In practice infinitesimal perturbations in the input data are examined.
Vectors and Matrices

- **Vector**

  \[ \mathbf{v}^T = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{bmatrix} \]

- **Dot product** \( \mathbf{a} \cdot \mathbf{b} = \sum_i a_i b_i \)

- **Matrix**

  \[ \mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{m1} & \cdots & \cdots & m_{mn} \end{bmatrix} \]
Vectors and Matrices

- **Vector Norms**
  - $\|v\| \geq 0$
  - $\|sv\| = |s|\|v\|$ for all scalars $s$
  - $\|v\| = 0$ if and only if $v = 0$
  - $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$
  - $\|v\|_p = \left(\sum |v_i|^p\right)^{1/p}$
  - $\|v\|_\infty = \sup_i\{|v_i|\}$
Vectors and Matrices

- **Matrix Norms**
  - $\| \mathbf{M} \| \geq 0$
  - $\| s \mathbf{M} \| = |s| \| \mathbf{M} \|$ for all scalars $s$
  - $\| \mathbf{M} \| = 0$ if and only if $\mathbf{M} = \mathbf{0}$
  - $\| \mathbf{M}_1 + \mathbf{M}_2 \| \leq \| \mathbf{M}_1 \| + \| \mathbf{M}_2 \|$

- **Induced Matrix Norm**

\[
\| \mathbf{M} \| = \sup_{\mathbf{v}} \frac{\| \mathbf{M} \mathbf{v} \|}{\| \mathbf{v} \|}
\]
Vectors and Matrices

- Matrix Inverse

\[ M^{-1}M = 1 \]

- Can calculate using Gaussian Elimination
- A matrix is invertible if determinant is non-zero
- Numerically, can have nearly invertible matrices for which care in choice of algorithms used is needed
Vectors and Matrices

• Eigenvalues and Eigenvectors

\[ \mathbf{Me} = \lambda \mathbf{e} \]

• Find eigenvalues by solving

\[ \det (\mathbf{M} - \lambda \mathbf{I}) = 0 \]

• Then find eigenvectors from eigenvalue eigenvector relationship

• Numerically, this can be very ill conditioned, in particular for large matrices
Vectors and Matrices

- Gram-Schmidt process for finding an orthogonal basis

\[ M = QR \]

- \( Q \in \mathbb{C}^{n \times n} \) is an orthogonal matrix, \( R \in \mathbb{C}^{n \times n} \) is an upper triangular matrix
Vectors and Matrices

- Singular Value Decomposition
  \[ M = USV \]

- \( M \in \mathbb{C}^{m \times n}, U \in \mathbb{C}^{m \times n}, S \in \mathbb{C}^{m \times n}, V \in \mathbb{C}^{m \times n}, \)

- Singular values are the positive square roots of the eigenvalues of \( M^T M \)

- \( U \) and \( V \) are orthogonal matrices, so \( U^TU = 1 \) and \( V^TV = 1 \)
**Theorem** Given an $m \times n$ matrix $A$, there are two orthonormal basis $\{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ and $\{u_1, \ldots, u_n\}$ of $\mathbb{R}^m$ and real numbers $s_1 \geq s_2 \geq \ldots \geq s_n \geq 0$ such that $Av_i = s_i u_i$ for $1 \leq i \leq \min\{m, n\}$. The columns of $V = [v_1 | \ldots | v_n]$, the right singular vectors are the set of orthonormal eigenvectors of $A^T A$; and the columns of $U = [u_1 | \ldots | u_m]$, the left singular vectors are the orthonormal eigenvectors of $AA^T$. Thus $AV = US$, where $S = \text{diag}(s_1, \ldots, s_n)$.

- This is used instead of the eigenvalue decomposition for non-square matrices
- This is used in principal component analysis
- This is used to compress images DEMO (though there are better algorithms)
Face Databases

- [http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html](http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html)
Higher order SVD

- Code [ds.cs.ut.ee/courses/course-files/codes.tar.gz](http://ds.cs.ut.ee/courses/course-files/codes.tar.gz)
References

- Quarteroni, Sacco and Saleri *Numerical Mathematics* Springer (2010), chapters 1 and 2
- Demmel *Applied Numerical Linear Algebra* SIAM (1997)
- Dashko “Face Recognition Project”
  https://github.com/kagan94/Face-recognition-via-SVD-and-PCA
  http://docs.oracle.com/cd/E19957-01/806-3568/ncg_goldberg.html
- Demmel, J. “Basic issues in Floating Point Arithmetic and Error Analysis” Lecture Notes
- IEEE 754-2008 standard http://dx.doi.org/10.1109/IEEESTD.2008.4610935
- Java Floating point www.cs.berkeley.edu/%7Ewtkahan/JAVAhurt.pdf