San José State University

Math 253: Mathematical Methods for Data Visualization

Matrix norm and low-rank approximation

Dr. Guangliang Chen

Outline

- Matrix norm
- Condition number
- Low-rank matrix approximation
- Applications

Introduction

Recall that a **vector space** is a collection of objects, called "vectors", which are endowed with two kinds of operations, **vector addition** and **scalar multiplication**, subject to requirements such as *associativity*, *commutativity*, and *distributivity*.

Below are some examples of vector spaces:

- Euclidean spaces ($\mathbb{R}^n$)
- The collection of all matrices of a fixed size ($\mathbb{R}^{m\times n}$)
- The collection of all functions from $\mathbb{R}$ to $\mathbb{R}$
- The collection of all polynomials
- The collection of all infinite sequences

Vector norm

A **norm** on a vector space $\mathcal{V}$ is a function

$$\|\cdot\| : \mathcal{V} \to \mathbb{R}$$

that satisfies the following three conditions:

- $\|\mathbf{v}\| \geq 0$ for all $\mathbf{v} \in \mathcal{V}$ and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$
- $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$ for any scalar $k \in \mathbb{R}$ and vector $\mathbf{v} \in \mathbb{R}^d$
- $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ for any two vectors $\mathbf{v}, \mathbf{w} \in \mathcal{V}$

The norm of a vector can be thought of as the length or magnitude of the vector.

Example 0.1. Below are three different norms on the Euclidean space $\mathbb{R}^d$:

- **2-norm (or Euclidean norm)**:

$$\|\mathbf{x}\|_2 = \sqrt{\sum x_i^2} = \sqrt{\mathbf{x}^T\mathbf{x}}$$

- **1-norm (Taxicab norm or Manhattan norm)**:

$$\|\mathbf{x}\|_1 = \sum |x_i|$$

- ∞**-norm (maximum norm)**:

$$\|\mathbf{x}\|_\infty = \max |x_i|$$

When unspecified, it is understood as the Euclidean 2-norm.

Remark. More generally, for any fixed $p > 0$, the ℓ_p norm on $\mathbb{R}^d$ is defined as

$$\|\mathbf{x}\|_p = \left(\sum |x_i|^p\right)^{1/p}, \quad \text{for all } \mathbf{x} \in \mathbb{R}^d$$

Remark. Any norm on $\mathbb{R}^d$ can be used as a metric to measure the distance between two vectors:

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|, \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$$

For example, the Euclidean distance between $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ (corresponding to the Euclidean norm) is

$$\text{dist}_E(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum (x_i - y_i)^2}$$

Matrix norm

A matrix norm is a norm on $\mathbb{R}^{m\times n}$ as a vector space (consisting of all matrices of the fixed size).

More specifically, a matrix norm is a function

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

that satisfies the following three conditions:

- $\|\mathbf{A}\| \geq 0$ for all $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\|\mathbf{A}\| = 0$ if and only if $\mathbf{A} = \mathbf{O}$
- $\|k\mathbf{A}\| = |k| \cdot \|\mathbf{A}\|$ for any scalar $k \in \mathbb{R}$ and matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$
- $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$ for any two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$

The Frobenius norm

Def 0.1. The Frobenius norm of a matrix $\mathbf{A} \in \mathbb{R}^{n\times d}$ is defined as

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$$

Remark. The Frobenius norm on $\mathbb{R}^{n\times d}$ is equivalent to the Euclidean 2-norm on the space of vectorized matrices (i.e., $\mathbb{R}^{nd}$):

$$\|\mathbf{A}\|_F = \|\mathbf{A}(:)\|_2$$

Example 0.2. Let

$$\mathbf{X} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

By direct calculation, $\|\mathbf{X}\|_F = 2$.

Proposition 0.1. For any matrix $\mathbf{A} \in \mathbb{R}^{n\times d}$,

$$\|\mathbf{A}\|_F^2 = \mathrm{trace}(\mathbf{A}\mathbf{A}^T) = \mathrm{trace}(\mathbf{A}^T\mathbf{A})$$

Proof. We demonstrate the first identity for 2×2 matrices $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$:

$$\mathbf{A}\mathbf{A}^T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a^2+b^2 & * \\ * & c^2+d^2 \end{pmatrix}$$

(The full proof is just a direct generalization of the above case) □

Theorem 0.2. For any matrix $\mathbf{A} \in \mathbb{R}^{n\times d}$ (with singular values $\{\sigma_i\}$),

$$\|\mathbf{A}\|_F = \sqrt{\sum \sigma_i^2}$$

Proof. Let the full SVD of $\mathbf{A}$ be $\mathbf{A} = \mathbf{U\Sigma V}^T$. Then

$$\mathbf{AA}^T = \mathbf{U\Sigma V}^T \cdot \mathbf{V\Sigma}^T\mathbf{U}^T = \mathbf{U}(\mathbf{\Sigma\Sigma}^T)\mathbf{U}^T.$$

Applying the formula $\|\mathbf{A}\|_F^2 = \mathrm{trace}(\mathbf{AA}^T)$ gives that

$$\|\mathbf{A}\|_F^2 = \mathrm{trace}(\mathbf{U\Sigma\Sigma}^T\mathbf{U}^T) = \mathrm{trace}(\mathbf{\Sigma\Sigma}^T\mathbf{U}^T\mathbf{U}) = \mathrm{trace}(\mathbf{\Sigma\Sigma}^T) = \sum \sigma_i^2.$$

□

The Operator norm

A second matrix norm is the operator norm, which is induced by a vector norm on Euclidean spaces.

Theorem 0.3. For any norm $\|\cdot\|$ on Euclidean spaces, the following is a norm on $\mathbb{R}^{m\times n}$:

$$\|\mathbf{A}\| \overset{\text{def}}{=} \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\mathbf{u}\in\mathbb{R}^n:\|\mathbf{u}\|=1} \|\mathbf{A}\mathbf{u}\|$$

Proof. We need to verify the three conditions of a norm.

First, it is obvious that $\|\mathbf{A}\| \geq 0$ for any $\mathbf{A} \in \mathbb{R}^{m\times n}$. Suppose $\|\mathbf{A}\| = 0$. Then for any $\mathbf{x} \neq \mathbf{0}$, $\|\mathbf{A}\mathbf{x}\| = 0$, or equivalently, $\mathbf{A}\mathbf{x} = \mathbf{0}$. This implies that $\mathbf{A} = \mathbf{O}$. (The other direction is trivial)

Second, for any $k \in \mathbb{R}$,

$$\|k\mathbf{A}\| = \max_{\mathbf{u}\in\mathbb{R}^n:\|\mathbf{u}\|=1} \|(k\mathbf{A})\mathbf{u}\| = |k| \cdot \max_{\mathbf{u}\in\mathbb{R}^n:\|\mathbf{u}\|=1} \|\mathbf{A}\mathbf{u}\| = |k| \cdot \|\mathbf{A}\|.$$

Lastly, for any two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$,

$$\begin{aligned}
\|\mathbf{A}+\mathbf{B}\| &= \max_{\mathbf{u}\in\mathbb{R}^n:\|\mathbf{u}\|=1} \|(\mathbf{A}+\mathbf{B})\mathbf{u}\| = \max_{\mathbf{u}\in\mathbb{R}^n:\|\mathbf{u}\|=1} \|\mathbf{A}\mathbf{u}+\mathbf{B}\mathbf{u}\| \\
&\le \max_{\mathbf{u}\in\mathbb{R}^n:\|\mathbf{u}\|=1} (\|\mathbf{A}\mathbf{u}\| + \|\mathbf{B}\mathbf{u}\|) \\
&\le \max_{\mathbf{u}\in\mathbb{R}^n:\|\mathbf{u}\|=1} \|\mathbf{A}\mathbf{u}\| + \max_{\mathbf{u}\in\mathbb{R}^n:\|\mathbf{u}\|=1} \|\mathbf{B}\mathbf{u}\| \\
&= \|\mathbf{A}\| + \|\mathbf{B}\|.
\end{aligned}$$

$\square$

Theorem 0.4. For any norm on Euclidean spaces and its induced matrix operator norm, we have

$$\|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{x}\| \quad \text{for all } \mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{x} \in \mathbb{R}^n$$

Proof. For all $\mathbf{x} \neq \mathbf{0} \in \mathbb{R}^n$, by definition,

$$\frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|\mathbf{A}\|$$

This implies that

$$\|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{x}\|.$$

□

Remark. More generally, the matrix operator norm satisfies

$$\|\mathbf{AB}\| \le \|\mathbf{A}\| \cdot \|\mathbf{B}\|, \quad \text{for all } \mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times p}$$

Matrix norms with such a property are called sub-multiplicative matrix norms.

The proof of this result is left to you in the homework (you will need to apply the theorem on the preceding slide).

When the Euclidean norm (i.e., 2-norm) is used, the induced matrix operator norm is called the spectral norm.

Def 0.2. The **spectral norm** of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is defined as

$$\|\mathbf{A}\|_2 = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \max_{\mathbf{u}\in\mathbb{R}^n:\|\mathbf{u}\|_2=1} \|\mathbf{A}\mathbf{u}\|_2$$

Theorem 0.5. The spectral norm of any matrix coincides with its largest singular value:

$$\|\mathbf{A}\|_2 = \sigma_1(\mathbf{A}).$$

Proof.

$$\|\mathbf{A}\|_2^2 = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \lambda_1(\mathbf{A}^T\mathbf{A}) = \sigma_1(\mathbf{A})^2.$$

The maximizer is the largest right singular vector of $\mathbf{A}$, i.e. $\mathbf{v}_1$. □

Example 0.3. For the matrix

$$\mathbf{X} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix},$$

we have $\|\mathbf{X}\|_2 = \sqrt{3}$.

We note that the Frobenius and spectral norms of a matrix correspond to the 2- and ∞-norms of the vector of singular values ($\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_r)$):

$$\|\mathbf{A}\|_F = \|\boldsymbol{\sigma}\|_2, \qquad \|\mathbf{A}\|_2 = \|\boldsymbol{\sigma}\|_\infty$$

The 1-norm of the singular value vector is called the nuclear norm of $\mathbf{A}$, which is very useful in convex programming.

Def 0.3. The **nuclear norm** of a matrix $\mathbf{A} \in \mathbb{R}^{n\times d}$ is defined as

$$\|\mathbf{A}\|_* = \|\boldsymbol{\sigma}\|_1 = \sum \sigma_i.$$

Example 0.4. In the last example, $\|\mathbf{X}\|_* = \sqrt{3} + 1$.

MATLAB function for matrix/vector norm

norm – Matrix or vector norm.

norm(X,2) returns the 2-norm of X.

norm(X) is the same as norm(X,2).

norm(X,'fro') returns the Frobenius norm of X.

In addition, for vectors...

norm(V,P) returns the p-norm of V defined as SUM(ABS(V).^P)^(1/P).

norm(V,Inf) returns the largest element of ABS(V).

Condition number of a matrix

Briefly speaking, the condition number of a square, invertible matrix is a measure of its near-singularity.

Def 0.4. Let $\mathbf{A}$ be any square, invertible matrix. For a given matrix operator norm $\|\cdot\|$, the **condition number** of $\mathbf{A}$ is defined as

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\|$$

Remark. $\kappa(\mathbf{A}) \geq \|\mathbf{A}\mathbf{A}^{-1}\| = \|\mathbf{I}\| = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{I}\mathbf{x}\|}{\|\mathbf{x}\|} = 1.$

Remark. Under the matrix spectral norm,

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|_2\|\mathbf{A}^{-1}\|_2 = \sigma_{\max}(\mathbf{A}) \cdot \frac{1}{\sigma_{\min}(\mathbf{A})} = \frac{\sigma_{\max}(\mathbf{A})}{\sigma_{\min}(\mathbf{A})}$$

1st perspective (for understanding the matrix condition number): Let

$$M = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \|\mathbf{A}\|$$

$$m = \min_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \min_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{y}\|}{\|\mathbf{A}^{-1}\mathbf{y}\|} = \frac{1}{\max_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{A}^{-1}\mathbf{y}\|}{\|\mathbf{y}\|}} = \frac{1}{\|\mathbf{A}^{-1}\|}$$

which respectively represent how much the matrix $\mathbf{A}$ can stretch or shrink vectors. Then

$$\kappa(\mathbf{A}) = \frac{M}{m}$$

If $\mathbf{A}$ is singular, then there exists a nonzero $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{0}$. Thus, $m = 0$ and the condition number is infinity.

In general, a finite, large condition number means that the matrix is close to being singular. In this case, we say that the matrix $\mathbf{A}$ is ill-conditioned (for inversion).

2nd perspective: Consider solving a system of linear equations subject to measurement error:

$$\mathbf{A}\mathbf{x} = \mathbf{b} + \underbrace{\mathbf{e}}_{\text{error}} \qquad (\mathbf{A} \text{ is assumed to be exact})$$

We would like to see how the error term $\mathbf{e}$ (along with $\mathbf{b}$) affects the solution $\mathbf{x}$ (via the matrix $\mathbf{A}$).

Since $\mathbf{A}$ is invertible, the linear system has the following solution

$$\mathbf{x} = \mathbf{A}^{-1}(\mathbf{b} + \mathbf{e}) = \underbrace{\mathbf{A}^{-1}\mathbf{b}}_{\text{true solution}} + \underbrace{\mathbf{A}^{-1}\mathbf{e}}_{\text{error in solution}}$$

The ratio of the relative error in the solution to the relative error in $\mathbf{b}$ is

$$\frac{\|\mathbf{A}^{-1}\mathbf{e}\| \,/\, \|\mathbf{A}^{-1}\mathbf{b}\|}{\|\mathbf{e}\| \,/\, \|\mathbf{b}\|} = \frac{\|\mathbf{A}^{-1}\mathbf{e}\|}{\|\mathbf{e}\|} \cdot \frac{\|\mathbf{b}\|}{\|\mathbf{A}^{-1}\mathbf{b}\|}$$

where $\|\cdot\|$ represents a given vector norm.

The maximum possible value of the above ratio (for nonzero $\mathbf{b}, \mathbf{e}$) is then

$$\begin{aligned}
\max_{\mathbf{b}\neq\mathbf{0},\mathbf{e}\neq\mathbf{0}} \left(\frac{\|\mathbf{A}^{-1}\mathbf{e}\|}{\|\mathbf{e}\|} \cdot \frac{\|\mathbf{b}\|}{\|\mathbf{A}^{-1}\mathbf{b}\|} \right) &= \max_{\mathbf{e}\neq\mathbf{0}} \left(\frac{\|\mathbf{A}^{-1}\mathbf{e}\|}{\|\mathbf{e}\|} \right) \max_{\mathbf{b}\neq\mathbf{0}} \left(\frac{\|\mathbf{b}\|}{\|\mathbf{A}^{-1}\mathbf{b}\|} \right) \\
&= \max_{\mathbf{e}\neq\mathbf{0}} \left(\frac{\|\mathbf{A}^{-1}\mathbf{e}\|}{\|\mathbf{e}\|} \right) \max_{\mathbf{y}\neq\mathbf{0}} \left(\frac{\|\mathbf{A}\mathbf{y}\|}{\|\mathbf{y}\|} \right) \\
&= \|\mathbf{A}^{-1}\| \cdot \|\mathbf{A}\| \\
&= \kappa(\mathbf{A}).
\end{aligned}$$

Remark.

- If the condition number of $\mathbf{A}$ is large (i.e., ill-conditioned), then the relative error in the solution $\mathbf{x}$ is much larger than the relative error in $\mathbf{b}$, and thus even a small error in $\mathbf{b}$ may cause a large error in $\mathbf{x}$.
- On the other hand, if this number is small, then the relative error in $\mathbf{x}$ will not be much bigger than the relative error in $\mathbf{b}$.
- In the special case when the condition number is exactly one, the solution has the same relative error with the data $\mathbf{b}$.

See a MATLAB demonstration by C. Moler.[1]

[1] `https://blogs.mathworks.com/cleve/2017/07/17/what-is-the-condition-number-of-a-matrix/`

Low-rank approximation of matrices

Problem. For any matrix $\mathbf{A} \in \mathbb{R}^{n\times d}$ and integer $k \geq 1$, find the rank-k matrix $\mathbf{B}$ that is the closest to $\mathbf{A}$ (under a given norm such as Frobenius, or spectral):

$$\min_{\mathbf{B}\in\mathbb{R}^{n\times d}\,:\,\mathrm{rank}(\mathbf{B})=k} \|\mathbf{A}-\mathbf{B}\|$$

Remark. This problem arises in a number of tasks, e.g.,

- Orthogonal least squares fitting
- Data compression (and noise reduction)
- Recommender systems

Theorem 0.6 (Eckart–Young–Mirsky). Given $\mathbf{A} \in \mathbb{R}^{n\times d}$ and $1 \le k \le \operatorname{rank}(\mathbf{A})$, let $\mathbf{A}_k$ be the truncated SVD of $\mathbf{A}$ with the largest k terms: $\mathbf{A}_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$. Then $\mathbf{A}_k$ is the best rank-k approximation to $\mathbf{A}$ in terms of both the Frobenius and spectral norms:[2]

$$\min_{\mathbf{B}\,:\,\operatorname{rank}(\mathbf{B})=k} \|\mathbf{A}-\mathbf{B}\|_F = \|\mathbf{A}-\mathbf{A}_k\|_F = \sqrt{\sum_{i>k} \sigma_i^2}$$

$$\min_{\mathbf{B}\,:\,\operatorname{rank}(\mathbf{B})=k} \|\mathbf{A}-\mathbf{B}\|_2 = \|\mathbf{A}-\mathbf{A}_k\|_2 = \sigma_{k+1}.$$

Remark. The theorem still holds true if the equality constraint $\operatorname{rank}(\mathbf{B}) = k$ is relaxed to $\operatorname{rank}(\mathbf{B}) \le k$ (which will also include all the lower-rank matrices).

[2] Proof available at `https://en.wikipedia.org/wiki/Low-rank_approximation`

Example 0.5. For the matrix

$$\mathbf{X} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix},$$

the best rank-1 approximation is

$$\mathbf{X}_1 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T = \sqrt{3} \begin{pmatrix} \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

In this problem, the approximation error under either norm (spectral or Frobenius) is the same: $\|\mathbf{X} - \mathbf{X}_1\| = \sigma_2 = 1$.

Applications of low-rank approximation

- Orthogonal least-squares fitting
- Image compression

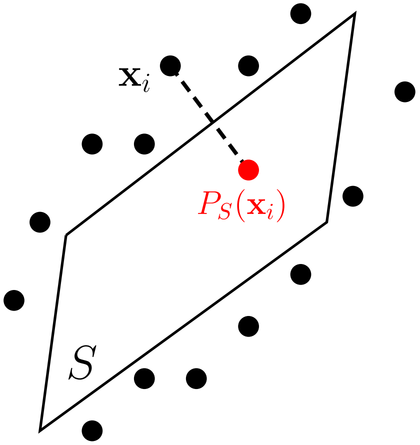
Orthogonal Best-Fit Subspace

Problem: Given data $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ and an integer $0 < k < d$, find the k-D orthogonal "best-fit" plane by solving

$$\min_S \sum_{i=1}^n \|\mathbf{x}_i - \mathcal{P}_S(\mathbf{x}_i)\|_2^2$$

Remark. This problem is different from ordinary linear regression:

- No predictor-response distinction
- Orthogonal (not vertical) fitting errors



Theorem 0.7. An orthogonal best-fit k-dimensional plane to the data $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^T \in \mathbb{R}^{n\times d}$ is given by

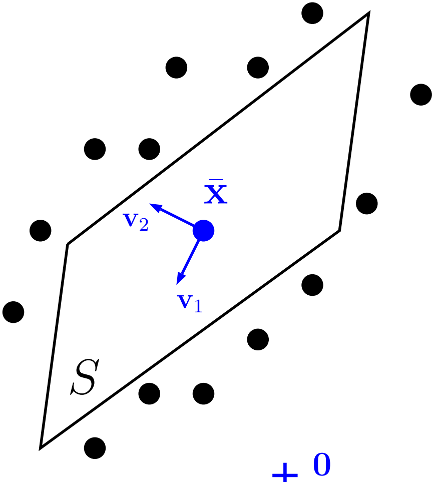
$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{V}_k \cdot \boldsymbol{\alpha}$$

where $\bar{\mathbf{x}}$ is the center of the data set

$$\bar{\mathbf{x}} = \frac{1}{n}\sum \mathbf{x}_i$$

and $\mathbf{V}_k = [\mathbf{v}_1 \dots \mathbf{v}_k]$ is a $d \times k$ matrix whose columns are the top k right singular vectors of the centered data matrix

$$\widetilde{\mathbf{X}} = [\mathbf{x}_1 - \bar{\mathbf{x}}, \dots, \mathbf{x}_n - \bar{\mathbf{x}}]^T = \mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T.$$



Proof. Suppose an arbitrary k-dimensional plane $\mathcal{S}$ is used to fit the data, with a fixed point $\mathbf{m} \in \mathbb{R}^d$, and an orthonormal basis

$$\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_k] \in \mathbb{R}^{d\times k}.$$

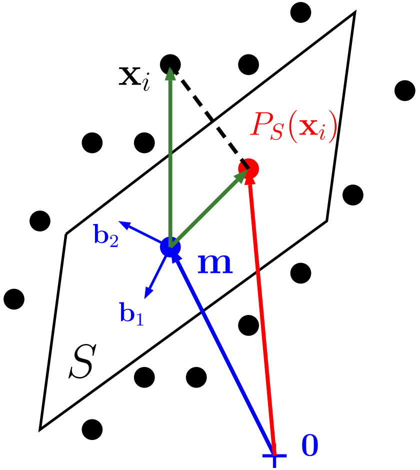
That is,

$$\mathbf{B}^T\mathbf{B} = \mathbf{I}_k,$$

$$\mathbf{B}\mathbf{B}^T : \text{orthogonal projection onto } S$$

The projection of each data point $\mathbf{x}_i$ onto the candidate plane is

$$\mathcal{P}_S(\mathbf{x}_i) = \mathbf{m} + \mathbf{B}\mathbf{B}^T(\mathbf{x}_i - \mathbf{m}).$$



Accordingly, we may rewrite the original problem as

$$\min_{\substack{\mathbf{m}\in\mathbb{R}^d,\,\mathbf{B}\in\mathbb{R}^{d\times k}\\ \mathbf{B}^T\mathbf{B}=\mathbf{I}_k}} \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{m} - \mathbf{B}\mathbf{B}^T(\mathbf{x}_i - \mathbf{m})\|^2$$

Using multivariable calculus, we can show that for any fixed $\mathbf{B}$ an optimal $\mathbf{m}$ is

$$\mathbf{m}^* = \frac{1}{n}\sum \mathbf{x}_i \stackrel{\text{def}}{=} \bar{\mathbf{x}}.$$

Plugging in $\bar{\mathbf{x}}$ for $\mathbf{m}$ and letting $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}}$ gives that

$$\min_{\mathbf{B}} \sum \|\tilde{\mathbf{x}}_i - \mathbf{B}\mathbf{B}^T\tilde{\mathbf{x}}_i\|^2.$$

In matrix notation, this becomes

$$\min_{\mathbf{B}} \|\widetilde{\mathbf{X}} - \widetilde{\mathbf{X}}\mathbf{B}\mathbf{B}^T\|_F^2, \qquad \text{where} \quad \widetilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n]^T \in \mathbb{R}^{n\times d}.$$

Let the full SVD of the centered data matrix $\widetilde{\mathbf{X}}$ be

$$\widetilde{\mathbf{X}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

Denote by $\widetilde{\mathbf{X}}_k$ the best rank-k approximation of $\widetilde{\mathbf{X}}$:

$$\widetilde{\mathbf{X}}_k = \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^T.$$

Then the minimum is attained when

$$\widetilde{\mathbf{X}}\mathbf{B}\mathbf{B}^T = \widetilde{\mathbf{X}}_k,$$

and a minimizer is the matrix consisting of the top k right singular vectors of $\widetilde{\mathbf{X}}$, i.e.,

$$\mathbf{B} = \mathbf{V}_k \equiv \mathbf{V}(:, 1:k).$$

Verify: If $\mathbf{B} = \mathbf{V}_k$, then

$$\begin{aligned}
\widetilde{\mathbf{X}}\mathbf{B}\mathbf{B}^T &= \widetilde{\mathbf{X}}\mathbf{V}_k\mathbf{V}_k^T \\
&= \widetilde{\mathbf{X}}[\mathbf{v}_1, \ldots, \mathbf{v}_k]\mathbf{V}_k^T \\
&= [\sigma_1\mathbf{u}_1, \ldots, \sigma_k\mathbf{u}_k]\mathbf{V}_k^T \\
&= [\mathbf{u}_1, \ldots, \mathbf{u}_k]\,\mathrm{diag}(\sigma_1, \ldots, \sigma_k)\mathbf{V}_k^T \\
&= \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^T \\
&= \widetilde{\mathbf{X}}_k.
\end{aligned}$$

Proof of $\mathbf{m}^* = \bar{\mathbf{x}}$:

First, rewrite the above objective function as

$$g(\mathbf{m}) = \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{m} - \mathbf{B}\mathbf{B}^T(\mathbf{x}_i - \mathbf{m})\|^2 = \sum_{i=1}^{n} \|(\mathbf{I} - \mathbf{B}\mathbf{B}^T)(\mathbf{x}_i - \mathbf{m})\|^2$$

and apply the formula

$$\frac{\partial}{\partial \mathbf{x}} \|\mathbf{A}\mathbf{x}\|^2 = 2\mathbf{A}^T\mathbf{A}\mathbf{x}$$

to find its gradient:

$$\nabla g(\mathbf{m}) = -\sum 2(\mathbf{I} - \mathbf{B}\mathbf{B}^T)^T(\mathbf{I} - \mathbf{B}\mathbf{B}^T)(\mathbf{x}_i - \mathbf{m})$$

Note that $\mathbf{I}-\mathbf{B}\mathbf{B}^T$ is also an orthogonal projection matrix (onto the complement). Thus,

$$(\mathbf{I}-\mathbf{B}\mathbf{B}^T)^T(\mathbf{I}-\mathbf{B}\mathbf{B}^T) = (\mathbf{I}-\mathbf{B}\mathbf{B}^T)^2 = \mathbf{I}-\mathbf{B}\mathbf{B}^T.$$

It follows that

$$\nabla g(\mathbf{m}) = -\sum 2(\mathbf{I}-\mathbf{B}\mathbf{B}^T)(\mathbf{x}_i - \mathbf{m}) = -2(\mathbf{I}-\mathbf{B}\mathbf{B}^T)\left(\sum \mathbf{x}_i - n\mathbf{m}\right)$$

Any minimizer $\mathbf{m}$ must satisfy

$$2(\mathbf{I}-\mathbf{B}\mathbf{B}^T)\left(\sum \mathbf{x}_i - n\mathbf{m}\right) = 0$$

This equation has infinitely many solutions, but the simplest one is

$$\sum \mathbf{x}_i - n\mathbf{m} = \mathbf{0} \qquad \longrightarrow \qquad \mathbf{m} = \frac{1}{n}\sum \mathbf{x}_i.$$

Example 0.6. Find the orthogonal best-fit line for a data set of three points $(1,1),(2,3),(3,2)$.

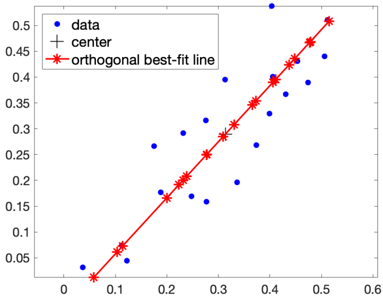
Solution. First, the center of the data is $\bar{\mathbf{x}} = \frac{1}{3}(1+2+3, 1+3+2) = (2,2)$. Thus, the centered data matrix is

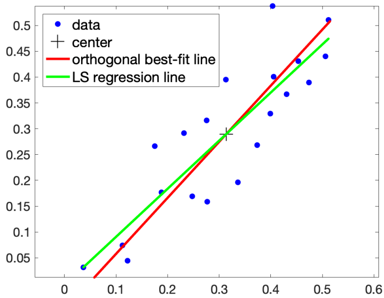
$$\bar{\mathbf{X}} = \begin{bmatrix} -1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \longrightarrow \quad \mathbf{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Therefore, the orthogonal best-fit line is $\mathbf{x}(t) = (2,2) + t(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, and the projections of the original data onto the best-fit line is

$$\mathbf{1}\bar{\mathbf{x}}^T + \widetilde{\mathbf{X}}\mathbf{v}_1\mathbf{v}_1^T = \begin{bmatrix} 2 & 2 \\ 2 & 2 \\ 2 & 2 \end{bmatrix} + \begin{bmatrix} -1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \frac{5}{2} & \frac{5}{2} \\ \frac{5}{2} & \frac{5}{2} \end{bmatrix}$$

Demonstration on another data set





Application to image compression

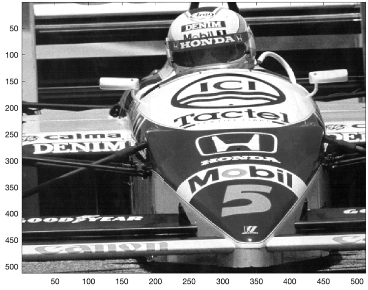
Digital images are stored as matrices, so we can apply SVD to obtain their low-rank approximations (and display them as images):

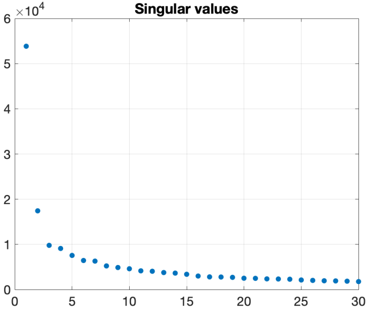
$$\mathbf{A}_{m\times n} \approx \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^T = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

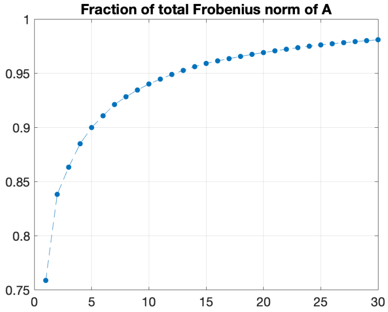
By storing $\mathbf{U}_k, \mathbf{\Sigma}_k, \mathbf{V}_k$ instead of $\mathbf{A}$, we can reduce the storage requirement from mn to

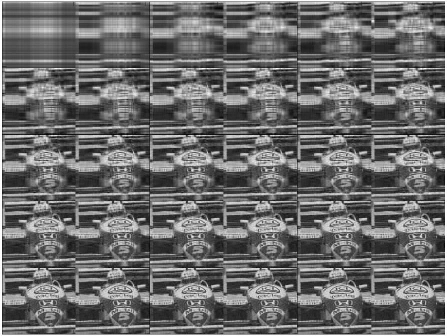
$$\underbrace{mk}_{\text{cost of } \mathbf{U}_k} + \underbrace{k}_{\text{cost of } \mathbf{\Sigma}_k} + \underbrace{nk}_{\text{cost of } \mathbf{V}_k} = k(m+n+1).$$

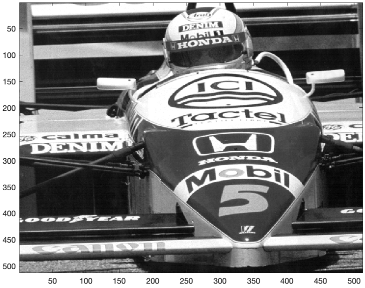
This is one magnitude smaller when $k \ll \min(m, n)$.

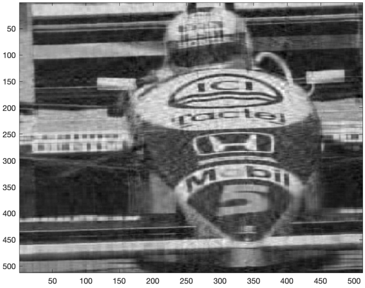

50
100
150
200
250
300
350
400
450
500
50
100
150
200
250
300
350
400
450
500
DENIM
HONDA
ICI
Tactel
calma
DENIM
HONDA
Mobil
5
GOODYEAR


Singular values
$\times 10^4$
6
5
4
3
2
1
0
0
5
10
15
20
25
30


Fraction of total Frobenius norm of A
1
0.95
0.9
0.85
0.8
0.75
0
5
10
15
20
25
30




50
100
150
200
250
300
350
400
450
500
50 100 150 200 250 300 350 400 450 500
ICI
Tactel
HONDA
Mobil
5
DENIM


50
100
150
200
250
300
350
400
450
500
50 100 150 200 250 300 350 400 450 500

Some practice problems

1. Show that for any $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times p}$,

$$\|\mathbf{AB}\|_F \le \|\mathbf{A}\|_F \|\mathbf{B}\|_F$$

Solution. Using the Cauchy-Schwarz inequality

$$(\mathbf{a}^T\mathbf{b})^2 = \left(\sum a_i b_i\right)^2 \le \left(\sum a_i^2\right)\left(\sum b_i^2\right) = \|\mathbf{a}\|^2\|\mathbf{b}\|^2$$

we have

$$\begin{aligned}
\|\mathbf{AB}\|_F^2 &= \sum_i \sum_j (\mathbf{A}(i,:)\mathbf{B}(:,j))^2 \\
&\le \sum_i \sum_j \|\mathbf{A}(i,:)\|^2 \|\mathbf{B}(:,j)\|^2 \\
&= \left(\sum_i \|\mathbf{A}(i,:)\|^2\right) \left(\sum_j \|\mathbf{B}(:,j)\|^2\right) \\
&= \|\mathbf{A}\|_F^2 \|\mathbf{B}\|_F^2.
\end{aligned}$$

2. Suppose $\mathbf{A} = \mathbf{x}\mathbf{y}^T$ where $\mathbf{x}, \mathbf{y}$ are two vectors (in column form). Show that

$$\|\mathbf{A}\|_2 = \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$$

Proof. First, the identity holds trivially true if $\mathbf{y} = \mathbf{0}$ (in which case $\mathbf{A} = \mathbf{O}$). Suppose $\mathbf{y} \neq \mathbf{0}$. Treating $\mathbf{x}, \mathbf{y}^T$ as matrices, we have

$$\|\mathbf{A}\|_2 = \|\mathbf{x}\mathbf{y}^T\|_2 \leq \|\mathbf{x}\|_2 \|\mathbf{y}^T\|_2 = \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$$

On the other hand, for the unit vector $\mathbf{v} = \frac{\mathbf{y}}{\|\mathbf{y}\|_2}$:

$$\|\mathbf{A}\|_2 \geq \|\mathbf{A}\mathbf{v}\|_2 = \left\| \mathbf{x}\mathbf{y}^T \frac{\mathbf{y}}{\|\mathbf{y}\|_2} \right\|_2 = \frac{1}{\|\mathbf{y}\|_2} \|\mathbf{x}(\mathbf{y}^T\mathbf{y})\|_2 = \frac{\|\mathbf{y}\|_2^2}{\|\mathbf{y}\|_2} \|\mathbf{x}\|_2 = \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$$

Consequently, we must have $\|\mathbf{A}\|_2 = \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$.

3. Let $\mathbf{A}$ be a square, symmetric matrix, with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Show that

$$\|\mathbf{A}\|_2 = \max(|\lambda_1|, |\lambda_n|) \qquad \text{and} \qquad \mathrm{Cond}(\mathbf{A}) = \frac{\max_i |\lambda_i|}{\min_i |\lambda_i|}$$

Proof. It suffices to show that the singular values of $\mathbf{A}$ are given by $|\lambda_1|, \ldots, |\lambda_n|$. To see this, consider the orthogonal diagonalization of $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$. It follows that

$$\mathbf{A}\mathbf{A}^T = \mathbf{A}^2 = \mathbf{Q}\mathbf{\Lambda}^2\mathbf{Q}^T$$

Therefore, the squared singular values of $\mathbf{A}$ (which are eigenvalues of $\mathbf{A}\mathbf{A}^T$) coincide with the eigenvalues of $\mathbf{A}^2$, i.e., $\sigma_1^2 = \lambda_1^2, \ldots, \sigma_n^2 = \lambda_n^2$ (not entirely sorted). This gives that $\sigma_1 = |\lambda_1|, \ldots, \sigma_n = |\lambda_n|$ (still not sorted, but the largest singular value must be the larger one of $|\lambda_1|, |\lambda_n|$.

4. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ be two unit vectors with angle θ. Show that

$$\|\mathbf{u}\mathbf{u}^T - \mathbf{v}\mathbf{v}^T\|_F^2 = 2\sin^2\theta$$

Solution. Using $\|\mathbf{A}\|_F^2 = \mathrm{trace}(\mathbf{A}\mathbf{A}^T)$ and observing that $\mathbf{u}\mathbf{u}^T, \mathbf{v}\mathbf{v}^T$ are both projection matrices (thus symmetric), we get

$$\begin{aligned}
\|\mathbf{u}\mathbf{u}^T - \mathbf{v}\mathbf{v}^T\|_F^2 &= \mathrm{trace}\left((\mathbf{u}\mathbf{u}^T - \mathbf{v}\mathbf{v}^T)(\mathbf{u}\mathbf{u}^T - \mathbf{v}\mathbf{v}^T)\right) \\
&= \mathrm{trace}\left(\mathbf{u}\mathbf{u}^T\mathbf{u}\mathbf{u}^T - \mathbf{u}\mathbf{u}^T\mathbf{v}\mathbf{v}^T - \mathbf{v}\mathbf{v}^T\mathbf{u}\mathbf{u}^T + \mathbf{v}\mathbf{v}^T\mathbf{v}\mathbf{v}^T\right) \\
&= \mathrm{trace}(\mathbf{u}\mathbf{u}^T) - \mathrm{trace}(\mathbf{u}\mathbf{u}^T\mathbf{v}\mathbf{v}^T) - \mathrm{trace}(\mathbf{v}\mathbf{v}^T\mathbf{u}\mathbf{u}^T) + \mathrm{trace}(\mathbf{v}\mathbf{v}^T) \\
&= \mathrm{trace}(\mathbf{u}^T\mathbf{u}) - \mathrm{trace}(\mathbf{u}^T\mathbf{v}\mathbf{v}^T\mathbf{u}) - \mathrm{trace}(\mathbf{v}^T\mathbf{u}\mathbf{u}^T\mathbf{v}) + \mathrm{trace}(\mathbf{v}^T\mathbf{v}) \\
&= 1 - (\mathbf{u}^T\mathbf{v})^2 - (\mathbf{v}^T\mathbf{u})^2 + 1 \\
&= 2 - 2(\mathbf{u}^T\mathbf{v})^2 \\
&= 2 - 2\cos^2\theta = 2\sin^2\theta.
\end{aligned}$$

5. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a symmetric matrix with eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ (and corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$). Solve the following problem:

$$\max_{\substack{\mathbf{x}\in\mathbb{R}^n:\, \mathbf{x}\neq\mathbf{0} \\ \mathbf{v}_1^T\mathbf{x}=0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

Solution. Write

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$$

and let

$$\mathbf{A}_2 = \mathbf{A} - \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T = \sum_{i=2}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$$

For any $\mathbf{x} \neq \mathbf{0} \in \mathbb{R}^n$ satisfying $\mathbf{v}_1^T\mathbf{x} = 0$, we have

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T(\mathbf{A} - \lambda_1\mathbf{v}_1\mathbf{v}_1^T)\mathbf{x} = \mathbf{x}^T\mathbf{A}_2\mathbf{x}$$

Accordingly, we can rewrite the original problem as

$$\max_{\substack{\mathbf{x}\in \operatorname{span}(\mathbf{v}_1)^{\perp}:\\ \mathbf{x}\neq \mathbf{0}}} \frac{\mathbf{x}^T\mathbf{A}_2\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

This is a regular Rayleigh quotient problem with a reduced domain (the orthogonal complement of $\operatorname{span}(\mathbf{v}_1)$ in $\mathbb{R}^n$) and the maximum is the largest eigenvalue of $\mathbf{A}_2$ over that domain which is λ_2, achieved at $\mathbf{x} = \mathbf{v}_2$.