Exercise Sheet 4

Singular Value Decomposition

1. Air Quality Tracking

A city plans to install m air quality stations to monitor pollution from n industrial sites. For a fixed pollutant, the concentration vector $\mathbf{c} \in \mathbb{R}^m$ at all stations is modeled by

$$\mathbf{c} \approx A\theta, \qquad A \in \mathbb{R}^{m\times n},\ \theta \in \mathbb{R}^n,$$

where θ_j is the emission rate of site j and the i-th row of A describes how emissions at all sites influence station i.

(i) Before installing any stations, the city has already obtained A in advance by running a dispersion simulation using historical wind statistics and the city's geography, but it has no measurements $\mathbf{c}$ yet. The city's administration would like to

  a) know if there are almost-redundant stations or not, and

  b) assess how sensitive the recovered θ would be to measurement noise in $\mathbf{c}$.

  Factorize the matrix A to solve this problem, and explain which parts of the decomposition you would inspect to find these answers.

(ii) For your chosen factorization, explain what each term encodes in the context of this measurement procedure (i.e., how one could interpret its entries in terms of stations, emission sources, or the ability to distinguish different scenarios).

(iii) The city is concerned that some pollution violations might go undetected. Characterize which emission patterns (i.e., which θ vectors) would produce the weakest sensor response and thus be hardest to detect.

Solution:

(i) The SVD $A = \mathbf{U}\Sigma\mathbf{V}^T$ answers both questions:

  a) **Redundant stations:** A sharp drop in the singular values (e.g., $\sigma_1 \approx \sigma_2 \gg \sigma_3 \approx 0$) indicates that A is approximately rank-deficient: its rows are nearly linearly dependent, meaning some stations can be approximated as linear combinations of others. The number of singular values before the drop tells you how many stations provide truly independent information.

  b) **Noise sensitivity:** Recovering θ from $\mathbf{c}$ requires inverting A. The condition number $\kappa = \sigma_1/\sigma_n$ controls how much measurement noise in $\mathbf{c}$ gets amplified in the recovered θ. If some singular values are tiny, certain emission patterns produce almost no change in $\mathbf{c}$, making those components of θ impossible to determine reliably.

(ii) The SVD identifies special emission scenarios and their corresponding measurements. If sources emit with rates given by $\theta = \mathbf{v}_\ell$ (the ℓ-th column of $\mathbf{V}$), the stations measure

$$\mathbf{c} = A\mathbf{v}_\ell = \sigma_\ell \mathbf{u}_\ell,$$

where $\mathbf{u}_\ell$ is the ℓ-th column of $\mathbf{U}$. Thus:

- $\mathbf{v}_\ell$ assigns an emission rate to each of the n sources.
- $\mathbf{u}_\ell$ gives the resulting reading at each of the m stations (up to the scale factor σ_ℓ).

- σ_ℓ measures how strongly this scenario affects the sensors: large σ_ℓ means easily detected, small σ_ℓ means nearly invisible.

The columns of $\mathbf{V}$ form an orthonormal basis for emission vectors, and the columns of $\mathbf{U}$ form an orthonormal basis for measurement vectors. Stations with similar rows in $\mathbf{U}$ respond similarly to all emission scenarios and are therefore redundant.

(iii) We want to find unit-norm emission patterns θ that minimize the sensor response $\|\mathbf{c}\| = \|A\theta\|$. From part (ii), if $\theta = \mathbf{v}_\ell$ then $\mathbf{c} = \sigma_\ell \mathbf{u}_\ell$, so $\|\mathbf{c}\| = \sigma_\ell$. The hardest-to-detect pattern is therefore

$$\theta = \mathbf{v}_n,$$

the right singular vector corresponding to the smallest singular value σ_n. This emission pattern produces the weakest possible sensor response $\|\mathbf{c}\| = \sigma_n$.

2. Total Least Squares

Ordinary linear regression in 2D fits the curve $y = mx+c$ by minimizing vertical errors. However, this assumes that the x-positions are exactly known, which may not be the case.

When *both* x- and y-values are contaminated by measurement errors, it is more principled to minimize *orthogonal* distances to the line. This is known as *total least squares* (TLS).

Writing the line as $ax + by = d$ with $a^2 + b^2 = 1$, the perpendicular distance of (x_i, y_i) is given by $|ax_i + by_i - d|$. For m data points (x_i, y_i), TLS then solves

$$\min_{a,b,d} \sum_{i=1}^{m} (ax_i + by_i - d)^2 \quad \text{s.t.} \quad a^2 + b^2 = 1.$$

(i) Let $\bar{x} = \frac{1}{m}\sum_i x_i$ and $\bar{y} = \frac{1}{m}\sum_i y_i$ be the centroid of the data. Show that at the optimum, $d = a\bar{x} + b\bar{y}$. (*Hint:* take the derivative of the objective with respect to d.)

(ii) Define centered coordinates $\tilde{x}_i = x_i - \bar{x}$ and $\tilde{y}_i = y_i - \bar{y}$. Using part (i), show that the objective simplifies to

$$\sum_{i=1}^{m} (a\tilde{x}_i + b\tilde{y}_i)^2.$$

(iii) Let $\mathbf{X} \in \mathbb{R}^{m\times 2}$ be the centered data matrix with rows $(\tilde{x}_i, \tilde{y}_i)$. Show that the objective from part (ii) equals $\|\mathbf{X}\mathbf{n}\|_2^2$ where $\mathbf{n} = (a, b)^T$.

(iv) Using the SVD of $\mathbf{X}$, explain why the optimal $\mathbf{n}$ is the right singular vector corresponding to the *smallest* singular value.

Solution:

(i) Taking the derivative with respect to d:

$$\frac{\partial}{\partial d} \sum_{i=1}^{m} (ax_i + by_i - d)^2 = -2 \sum_{i=1}^{m} (ax_i + by_i - d) = 0.$$

Solving: $\sum_i (ax_i + by_i) = md$, so $d = a\bar{x} + b\bar{y}$.

(ii) Substituting $d = a\bar{x} + b\bar{y}$ into the objective:

$$\begin{aligned}\sum_{i=1}^{m}(ax_i + by_i - d)^2 &= \sum_{i=1}^{m}(ax_i + by_i - a\bar{x} - b\bar{y})^2 \\ &= \sum_{i=1}^{m}(a(x_i - \bar{x}) + b(y_i - \bar{y}))^2 \\ &= \sum_{i=1}^{m}(a\tilde{x}_i + b\tilde{y}_i)^2.\end{aligned}$$

(iii) Writing out the squared norm:

$$\|\mathbf{X}\mathbf{n}\|_2^2 = \left\| \begin{pmatrix} \tilde{x}_1 & \tilde{y}_1 \\ \vdots & \vdots \\ \tilde{x}_m & \tilde{y}_m \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \right\|_2^2 = \left\| \begin{pmatrix} a\tilde{x}_1 + b\tilde{y}_1 \\ \vdots \\ a\tilde{x}_m + b\tilde{y}_m \end{pmatrix} \right\|_2^2 = \sum_{i=1}^{m}(a\tilde{x}_i + b\tilde{y}_i)^2.$$

(iv) To minimize $\|\mathbf{X}\mathbf{n}\|$ subject to $\|\mathbf{n}\| = 1$, we use that $\sigma_2 = \min_{\|\mathbf{n}\|=1} \|\mathbf{X}\mathbf{n}\|$, achieved by the corresponding right singular vector $\mathbf{v}_2$.