- Whitening and Canonical Correlation Analysis.

  - Now consider covariance matrix
    \[ M = \sum_{i=1}^{m} x_i x_i^T \]

  - Often want to perform linear transformation
    \[ z_i = W^T x_i \quad \text{such that} \]
    \[ M_Z = \sum_{i=1}^{m} z_i z_i^T = I \]
    - Motivation: result invariant to linear transformation.

  - How to find W?
    \[ \text{SVD of } M : \quad M = U D U^T \]
    \[ W = U D^{-\frac{1}{2}} \]
    \[ M_Z = \sum_{i=1}^{m} (W^T x_i)(W^T x_i)^T = W^T M W = D^{-\frac{1}{2}} U^T (U^T D U) U D^{-\frac{1}{2}} = I \]
    - Whiteness matrix is not unique (RW also works when RR^T = I)

- Canonical Correlation analysis
  - Given data \((X_1, Y_1), (X_2, Y_2), \ldots, (X_m, Y_m)\)
  - want to find direction \((u, v)\) such that \((X, Y)\) are most correlated in direction \((u, v)\).

  - More precisely: find \(u, v\) such that
    \[ \mathbb{E}[uu^T] = \mathbb{E}[vv^T] = 1 \]
    \[ \mathbb{E}[uu^T vv^T] \text{ maximized} \]

  - Solve using SVD
    \[ \text{Let } M_x = \mathbb{E}[xx^T], \quad M_y = \mathbb{E}[yy^T] \]
    \[ M_{xy} = \mathbb{E}[xy^T] \]
    \[ \text{Let } W_x, W_y \text{ be whitening matrix of } M_x, M_y. \]
    - Claim: If \(|\tilde{u}| = 1\), then \(u = W_x \tilde{u}\) satisfies
\[ U^T M X U = I. \]

**Proof:** just by \( W_x^T M_x W_x = I. \)

Therefore, we are trying to find \( \| \vec{u} \| = 1 \) s.t.

\[ E[<W_x \vec{u}, x> <W_y \vec{v}, y>] \] is maximized

but \[ E[<W_x \vec{u}, x> <W_y \vec{v}, y>] \]

\[ = \overline{u}^T (W_x^T X Y^T W_y) \overline{v} \]

\[ = \overline{u} W_x^T M_{xy} W_y \overline{v} \]

so we can simply do SVD of \( W_x^T M_{xy} W_y \)

---

**Matrix Perturbations**

- **Typical Scenario**

  "true matrix" \( A \) (often low rank)

  noise \( N \)

  how does singular values/vectors change from \( A \to A+N \)?

- **Main intuition:** Singular vectors are stable if

  \( \| N \| \ll \) spectral gap

- **Weyl's Theorem:** \( \forall i \)

  \[ \sigma_i(A) - \| N \| \leq \sigma_i(A+N) \leq \sigma_i(A) + \| N \| \]

To get intuition, think of \( A \) as diagonal

(can also do this w.l.o.g. in proof)

- **Wedin's Theorem:**

  If \( A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} D_1 & \mathbf{0} \\ \mathbf{0} & D_2 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}^T \)

  \( A+N = \begin{bmatrix} \hat{U}_1 & \hat{U}_2 \end{bmatrix} \begin{bmatrix} \hat{D}_1 & \mathbf{0} \\ \mathbf{0} & \hat{D}_2 \end{bmatrix} \begin{bmatrix} \hat{V}_1 \\ \hat{V}_2 \end{bmatrix}^T \)

  if \( \| N \| < \frac{1}{4} \cdot \min_{i,j} \| P(i,i) - \hat{D}_2(j,j) \| \)
if \[ \|N\| < \frac{1}{4} \cdot \min_{i,j} |P_{(i,j)} - \hat{D}_z(i,j)| \]

then \[ \sin \Theta(U, \hat{U}) \leq \frac{\text{gap } \tau}{2 \|N\|} \]

Def: \( \Theta(U, \hat{U}) = \max_{u \in U} \min_{\hat{u} \in \hat{U}} \Theta(u, \hat{u}) \)

"Principal angle between two subspaces"

\[ \sin \Theta(U, \hat{U}) = \| (I - U^T \hat{U}) \hat{U} \| = \| (I - \hat{U}^T \hat{U}) U \| \]

\[ = \| U^T - \hat{U}^T \| \]

Example: \( A \) is rank \( r \), \( \sigma_r(A) = \sigma \Rightarrow \|N\| \)

Weyl’s Theorem

\[ \text{singular value} \]

\[ \hat{\sigma} \]

index

\[ \sigma \]

index

Wedin’s Theorem: \[ \min D_i = \sigma \max \hat{D}_z \leq \|N\| \]

\[ \text{Gap } \tau \geq \sigma - \|N\| \]

\[ \sin \Theta(U, \hat{U}) \leq \frac{4 \|N\|}{\sigma - \|N\|} \]

How to prove \( \|N\| \) is small?

1. random matrix

for \( N \in \mathbb{R}^{m \times n} \), if all entries are iid Gaussian \( \pm 1 \)

\[ \|N\| \leq O(\sqrt{n}) \quad (2\sqrt{n}) \]
Wigner's Semicircle Law

\[ \text{Histogram of Singular Values} \]

Limitations: distribution of eigenvalues only known for a few distributions.