Overview

- SVD projection onto dominant subspace (in L_2-norm)
- Iterative projection onto embedded dominant subspaces
- Comparison to DNN composed of linear layers
- Interspersing of nonlinear transformations

SVD projection onto dominant subspaces

- Consider N measurements of a system characterized by m parameters, $N \gg m$

$$\boldsymbol{U} = [\ \boldsymbol{u}_1 \ \ \boldsymbol{u}_2 \ \ \ldots \ \ \boldsymbol{u}_N\ ] \in \mathbb{R}^{m \times N}$$

- Assume centered data, correlation matrix eigenvectors = left singular vectors

$$\boldsymbol{C} = \boldsymbol{U}\boldsymbol{U}^T, \boldsymbol{U} = \boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{V}^T \Rightarrow \boldsymbol{C} = \boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{Q}^T$$

- From full $\boldsymbol{Q} = [\ \boldsymbol{q}_1 \ \ \boldsymbol{q}_2 \ \ \ldots \ \ \boldsymbol{q}_m\ ] \in \mathbb{R}^{m \times m}$, drop kl negligible vectors

$$\boldsymbol{Q}_k = [\ \boldsymbol{q}_1 \ \ \boldsymbol{q}_2 \ \ \ldots \ \ \boldsymbol{q}_{m-kl}\ ], \boldsymbol{Q}_k^T\boldsymbol{Q}_k = \boldsymbol{I}_{m-kl}, \boldsymbol{P}_k = \boldsymbol{Q}_k\boldsymbol{Q}_k^T \in \mathbb{R}^{m \times m}$$

- Coarse grained representation: repeated dimension reduction in steps of l
  - Project onto subspace $C(\boldsymbol{Q}_k)$: $\boldsymbol{z}_k = \boldsymbol{P}_k\boldsymbol{u} \in \mathbb{R}^m$
  - Reduced coordinates $\boldsymbol{y}_k \in \mathbb{R}^{m-kl}$ of $\boldsymbol{z}_k$: $\boldsymbol{z}_k = \boldsymbol{Q}_k\boldsymbol{y}_k \Rightarrow \boldsymbol{y}_k = \boldsymbol{Q}_k^T\boldsymbol{z}_k$

Iterative projection onto embedded subspaces

- $\boldsymbol{y}_k = \boldsymbol{Q}_k^T \boldsymbol{z}_k \in \mathbb{R}^{m-kl}$, $\boldsymbol{y}_0 \equiv \boldsymbol{u} \in \mathbb{R}^m$,but note

$$\boldsymbol{y}_k = \boldsymbol{Q}_k^T \boldsymbol{z}_k = \boldsymbol{Q}_k^T \boldsymbol{P}_k \boldsymbol{u} = \boldsymbol{Q}_k^T \boldsymbol{Q}_k \boldsymbol{Q}_k^T \boldsymbol{u} = \boldsymbol{Q}_k^T \boldsymbol{u}$$

states that reduced coordinates are obtained by taking successively fewer components of the transformed data $\boldsymbol{Q}^T \boldsymbol{u}$

$$\boldsymbol{y}_k = \boldsymbol{Q}_k^T \boldsymbol{u} = \begin{pmatrix} \boldsymbol{q}_1^T \boldsymbol{u} \\ \vdots \\ \boldsymbol{q}_{m-kl}^T \boldsymbol{u} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_{m-kl} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{q}_1^T \boldsymbol{u} \\ \vdots \\ \boldsymbol{q}_m^T \boldsymbol{u} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_{m-kl} & \boldsymbol{0} \end{pmatrix} \boldsymbol{Q}^T \boldsymbol{u}$$

- K steps (fully linear, SVD model): $\boldsymbol{y}_K = \boldsymbol{Q}_K^T \boldsymbol{u}$

$$\boldsymbol{y}_K = \begin{pmatrix} \boldsymbol{I}_{m-Kl} & \boldsymbol{0}_l \end{pmatrix} \ldots \begin{pmatrix} \boldsymbol{I}_{m-2l} & \boldsymbol{0}_l \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_{m-l} & \boldsymbol{0}_l \end{pmatrix} \boldsymbol{Q}^T \boldsymbol{u}$$

- Above composition corresponds to a DNN formed by K unbiased linear layers

$$\boldsymbol{r}(\boldsymbol{u}) = (\boldsymbol{W}_K \circ \ldots \circ \boldsymbol{W}_1 \circ \boldsymbol{W}_0)(\boldsymbol{u}), \boldsymbol{W}_k \in \mathbb{R}^{(m-kl) \times (m-(k-1)l)}$$

- The L_2 optimal representation is linear combination of SVD modes

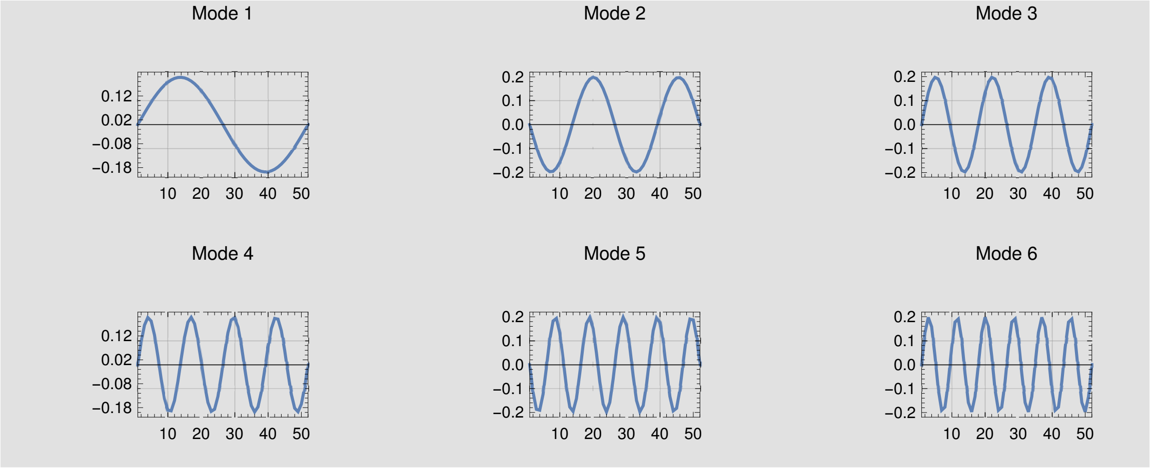

Figure 1. First 6 SVD modes are approximations of Sturm-Liouville eigenmodes to $\mathcal{O}(\epsilon)$

- Goal: find dominant eigenmodes of $\boldsymbol{C} = \boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{Q}^T = \boldsymbol{Q}\boldsymbol{A}\boldsymbol{Q}^T$
- $\mathcal{D} = \{(\boldsymbol{u}_i, \boldsymbol{Q}_K^T\boldsymbol{u}_i), i = 1, \ldots, N\}$
- $\boldsymbol{C}$ s.p.d., easily constructed from available data, easily updated from new data

Algorithm Krylov-Lanczos

$\beta_0 = 0, \boldsymbol{p}_0 = \boldsymbol{0}, \boldsymbol{b} = \boldsymbol{u}_1, \boldsymbol{p}_1 = \boldsymbol{b} / \|\boldsymbol{b}\|$
for $n = 1, 2, \ldots$
  $\boldsymbol{v} = \boldsymbol{C}\boldsymbol{p}_n$, $\alpha_n = \boldsymbol{p}_n^T\boldsymbol{v}$, $\boldsymbol{v} = \boldsymbol{v} - \beta_{n-1}\boldsymbol{p}_{n-1} - \alpha_n\boldsymbol{p}_n$
  $\beta_n = \|\boldsymbol{v}\|$, $\boldsymbol{p}_{n+1} = \boldsymbol{v} / \beta_n$

- Above generates $\boldsymbol{T}_n = \mathrm{diag}(\ \boldsymbol{\beta}\ \ \boldsymbol{\alpha}\ \ \boldsymbol{\beta}\ ) \in \mathbb{R}^{n\times n}$ a symmetric Hessenberg matrix