Lecture 31

Proper Orthogonal Decomposition – Introduction

In this lecture we introduce the method of proper orthogonal decomposition (POD) in the Euclidean space $\mathbb{R}^m$ and study the close connection to the singular value decomposition of rectangular matrices; see [23]. We also refer to the monograph [19].

31.1 POD and SVD

Let $Y = [y_1, \dots, y_n]$ be a real-valued $m \times n$ matrix of rank $d \le \min\{m, n\}$ with columns $y_j \in \mathbb{R}^m$, $1 \le j \le n$. Consequently,

$$\bar{y} = \frac{1}{n} \sum_{j=1}^{n} y_j \tag{31.1}$$

can be viewed as the column-averaged mean of the matrix Y.

Theorem 31.1.1 (SVD)**.** *There exist uniquely determined real numbers $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_d > 0$ and orthogonal matrices $U \in \mathbb{R}^{m\times m}$ with columns $\{u_i\}_{i=1}^m$ and $V \in \mathbb{R}^{n\times n}$ with columns $\{v_i\}_{i=1}^n$ such that*

$$U^\top Y V = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} =: \Sigma \in \mathbb{R}^{m\times n}, \tag{31.2}$$

where $D = \mathrm{diag}\ (\sigma_1, \dots, \sigma_d) \in \mathbb{R}^{d\times d}$ and the zeros in (31.2) denote matrices of appropriate dimensions. Moreover the vectors $\{u_i\}_{i=1}^d$ and $\{v_i\}_{i=1}^d$ satisfy

$$Y v_i = \sigma_i u_i \quad \textit{and} \quad Y^\top u_i = \sigma_i v_i \quad \textit{for } i = 1, \dots, d. \tag{31.3}$$

Proof. We follow the arguments given in [9, pp. 144-145]. For $Y = 0$ the claim is clear. Suppose that $Y \ne 0$ holds. Then,

$$\sigma_1 = \|Y\|_2 = \max_{\|v\|_{\mathbb{R}^n}=1} \|Yv\|_2 > 0.$$

Let $v \in \mathbb{R}^n$ be vector with $\|v\|_2 = 1$, where the maximum is attained. We set $u = Yv/\sigma_1 \in \mathbb{R}^m$. It follows that $\|u\|_2 = \|Yv\|_2/\sigma_1 = 1$. We extend u and v to orthonormal bases $\{u, \tilde{u}_2, \dots, \tilde{u}_m\}$ and $\{v, \tilde{v}_2, \dots, \tilde{v}_n\}$ in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. Next we define the two orthogonal matrices $U_1 = [u, \tilde{u}_2, \dots, \tilde{u}_m] \in \mathbb{R}^{m\times m}$ and $V_1 = [v, \tilde{v}_2, \dots, \tilde{v}_m] \in \mathbb{R}^{n\times n}$. Since $\langle \tilde{u}, Yv\rangle_2 = \sigma_1 \langle \tilde{u}_i, u\rangle_2 = \sigma_1 \tilde{u}_i^\top u = 0$ holds for $i = 2, \dots, m$, we find that

$$Y_1 = U_1^\top Y V_1 = \begin{pmatrix} \sigma_1 & w^\top \\ 0 & \tilde{Y} \end{pmatrix} \in \mathbb{R}^{m\times n}$$

with $w \in \mathbb{R}^{n-1}$ and $\tilde{Y} \in \mathbb{R}^{(m-1)\times(n-1)}$. We observe that

$$\left\| Y_1 \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \right\|_2 = \left\| \begin{pmatrix} \sigma_1^2 + w^\top w \\ \tilde{Y} w \end{pmatrix} \right\|_2 \ge \sigma_1^2 + \|w\|^2_{\mathbb{R}^{n-1}} = \left\| \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \right\|_2^2.$$

Moreover, $\|Y\|_2 = \|Y_1\|_2$ holds. Therefore, we have

$$\sigma_1 = \|Y_1\|_2 \geq \frac{\left\| Y_1 \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \right\|_2}{\left\| \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \right\|_2} \geq \sqrt{\sigma_1^2 + \|w\|^2_{\mathbb{R}^{n-1}}}.$$

Consequently, $w = 0$ and

$$U_1^\top Y V_1 = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \tilde{Y} \end{pmatrix} \in \mathbb{R}^{m\times n}.$$

Thus, the claim has been proved for $m = 1$ or $n = 1$. For the case $m, n > 1$ we apply an induction argument. For that purpose we assume that $U_2^\top \tilde{Y} V_2 = \Sigma_2$ with two orthogonal matrices $U_2 \in \mathbb{R}^{(m-1)\times(m-1)}$, $V_2 \in \mathbb{R}^{(n-1)\times(n-1)}$ and with a matrix $\Sigma_2 \in \mathbb{R}^{(m-1)\times(n-1)}$ of the same structure as the marix Σ in (31.2). Then, we find

$$\sigma_2 := \|\tilde{Y}\|_2 \leq \|Y_1\|_2 = \|U_1^\top Y V_1\|_2 = \|Y\|_2 = \sigma_1.$$

Setting

$$U = U_1 \begin{pmatrix} 1 & 0 \\ 0 & U_2 \end{pmatrix} \in \mathbb{R}^{m\times m} \quad \text{and} \quad V = V_1 \begin{pmatrix} 1 & 0 \\ 0 & V_2 \end{pmatrix} \in \mathbb{R}^{n\times n}$$

we get the decomposition

$$U^\top Y V = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}$$

which yields the claim by using the hypothesis of the induction. □

It follows directly from (31.3) that $\{u_i\}_{i=1}^m \subset \mathbb{R}^m$ and $\{v_i\}_{i=1}^n \subset \mathbb{R}^n$ are eigenvectors of $YY^\top$ and $Y^\top Y$, respectively, with eigenvalues $\lambda_i = \sigma_i^2 > 0$, $i = 1, \ldots, d$. The vectors $\{u_i\}_{i=d+1}^m$ and $\{v_i\}_{i=d+1}^n$ (if $d < m$ respectively $d < n$) are eigenvectors of $YY^\top$ and $Y^\top Y$ with eigenvalue 0.

From (31.2) we deduce that

$$Y = U\Sigma V^\top.$$

We infer (31.3) from the columnwise evaluation of (31.2). The follows It follows that Y can also be expressed as

$$Y = U^d D(V^d)^\top, \tag{31.4}$$

where $U^d \in \mathbb{R}^{m\times d}$ and $V^d \in \mathbb{R}^{n\times d}$ are given by

$$U^d_{ij} = U_{ij} \text{ for } 1 \leq i \leq m,\ 1 \leq j \leq d \quad \text{and} \quad V^d_{ij} = V_{ij} \text{ for } 1 \leq i \leq n,\ 1 \leq j \leq d.$$

Setting $B^d = D(V^d)^\top \in \mathbb{R}^{d\times n}$ we can write (31.4) in the form

$$Y = U^d B^d \quad \text{with } B^d = D(V^d)^\top \in \mathbb{R}^{d\times n}.$$

Thus, the column space of Y can be represented in terms of the d linearly independent columns of U^d. The coefficients in the expansion for the columns y_j, $j = 1, \ldots, n$, in the basis $\{u_i\}_{i=1}^d$ are given by the jth-column of B^d. Since U is orthogonal, we find that

$$\begin{aligned} y_j &= \sum_{i=1}^d B^d_{ij} U^d_{\cdot,i} = \sum_{i=1}^d \big(D(V^d)^\top\big)_{ij} u_i = \sum_{i=1}^d \big(\underbrace{(U^d)^T U^d}_{=I^d \in \mathbb{R}^{d\times d}} D(V^d)^\top\big)_{ij} u_i \\ &\overset{(31.4)}{=} \sum_{i=1}^d \big((U^d)^\top Y\big)_{ij} u_i = \sum_{i=1}^d \Big(\underbrace{\sum_{k=1}^m U^d_{ki} Y_{kj}}_{=u_i^\top y_j}\Big) u_i = \sum_{i=1}^d \langle u_i, y_j \rangle_2\, u_i, \end{aligned}$$

where $\langle \cdot\,, \cdot \rangle_{\mathbb{R}^m}$ denotes the canonical inner product in $\mathbb{R}^m$. Thus,

$$y_j = \sum_{i=1}^d \langle y_j, u_i \rangle_2\, u_i \quad \text{for } j = 1, \ldots, n \tag{31.5}$$

31.2 The POD method

Let us now interprete SVD in terms of POD. One of the central issues of POD is the reduction of data expressing their *essential information* by means of a few basis vectors. The problem of approximating all spatial coordinate vectors y_j of Y simultaneously by a single, normalized vector as well as possible can be expressed as

$$\max_{u\in\mathbb{R}^m}\sum_{j=1}^n\big|\langle y_j,u\rangle_2\big|^2 \quad \text{subject to (s.t.)} \quad \|u\|_2^2=1, \tag{$\mathbf{P}^1$}$$

where $\|u\|_{\mathbb{R}^m}=\sqrt{\langle u,u\rangle_{\mathbb{R}^m}}$ for $u\in\mathbb{R}^m$.

Note that ($\mathbf{P}^1$) is a constrained optimization problem that can be solved by considering first-order necessary optimality conditions; cf. [8, Theorem 13.3.3]. We introduce the function $e:\mathbb{R}^m\to\mathbb{R}$ by $e(u)=1-\|u\|_2^2$ for $u\in\mathbb{R}^m$. Then, the equality constraint in ($\mathbf{P}^1$) can be expressed as $e(u)=0$. Notice that $\nabla e(u)=-2u$ is linear independent if $u\neq 0$ holds. In particular, a solution to ($\mathbf{P}^1$) satisfies $u\neq 0$. Thus, any solution to ($\mathbf{P}^1$) is a *regular point*. Let $\mathcal{L}:\mathbb{R}^m\times\mathbb{R}\to\mathbb{R}$ be the Lagrange functional associated with ($\mathbf{P}^1$), i.e.,

$$\mathcal{L}(u,\lambda)=\sum_{j=1}^n\big|\langle y_j,u\rangle_2\big|^2+\lambda\big(1-\|u\|_2^2\big)\quad\text{for } (u,\lambda)\in\mathbb{R}^m\times\mathbb{R}.$$

Suppose that $u\in\mathbb{R}^m$ is a solution to ($\mathbf{P}^1$). Since u is regular, there exists a Lagrange multiplier satisfying the first-order necessary optimality condition

$$\nabla\mathcal{L}(u,\lambda)\overset{!}{=}0\quad\text{in } \mathbb{R}^m\times\mathbb{R}.$$

We compute the gradient of $\mathcal{L}$ with respect to u:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial u_i}(u,\lambda)&=\frac{\partial}{\partial u_i}\left(\sum_{j=1}^n\Big|\sum_{k=1}^m Y_{kj}u_k\Big|^2+\lambda\Big(1-\sum_{k=1}^m u_k^2\Big)\right)=2\sum_{j=1}^n\Big(\sum_{k=1}^m Y_{kj}u_k\Big)Y_{ij}-2\lambda u_i\\&=2\sum_{k=1}^m\Big(\underbrace{\sum_{j=1}^n Y_{ij}Y_{jk}^\top}_{=(YY^\top)_{ik}}u_k\Big)-2\lambda u_i.\end{aligned}$$

Thus,

$$\nabla_u\mathcal{L}(u,\lambda)=2\big(YY^\top u-\lambda u\big)\overset{!}{=}0\quad\text{in } \mathbb{R}^m. \tag{31.6}$$

Equation (31.6) yields the eigenvalue problem

$$YY^\top u=\lambda u\quad\text{in } \mathbb{R}^m. \tag{31.7a}$$

Notice that $YY^T\in\mathbb{R}^{m\times m}$ is a symmetric matrix satisfying

$$u^\top(YY^\top)u=(Y^\top u)^\top Y^\top u=\|Y^\top u\|_2^2\geq 0\quad\text{for all } u\in\mathbb{R}^m.$$

Thus, $YY^\top$ is positive semi-definite. It follows that YY^T possesses m non-negative eigenvalues $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_m\geq 0$ and the corresponding eigenvectors can be chosen such that they are pairwise orthonormal.

From $\frac{\partial\mathcal{L}}{\partial\lambda}(u,\lambda)\overset{!}{=}0$ in $\mathbb{R}$ we infer the constraint

$$\|u\|_2=1. \tag{31.7b}$$

Due to SVD the vector u_1 solves (31.7) and

$$\begin{aligned}\sum_{j=1}^n\big|\langle y_j,u_1\rangle_2\big|^2&=\sum_{j=1}^n\langle y_j,u_1\rangle_2\langle y_j,u_1\rangle_2=\sum_{j=1}^n\big\langle\langle y_j,u_1\rangle_2\,y_j,u_1\big\rangle_2=\Big\langle\sum_{j=1}^n\langle y_j,u_1\rangle_2\,y_j,u_1\Big\rangle_2\\&=\Big\langle\sum_{j=1}^n\Big(\sum_{k=1}^m Y_{kj}(u_1)_k\Big)y_j,u_1\Big\rangle_2=\Big\langle\sum_{k=1}^m\Big(\sum_{j=1}^n Y_{\cdot,j}Y_{jk}^\top(u_1)_k\Big),u_1\Big\rangle_2=\big\langle YY^\top u_1,u_1\big\rangle_2\\&=\lambda_1\,\big\langle u_1,u_1\big\rangle_2=\lambda_1\,\|u_1\|_2^2=\lambda_1.\end{aligned}$$

We next prove that u_1 solves ($\mathbf{P}^1$). Suppose that $\tilde{u} \in \mathbb{R}^m$ is an arbitrary vector with $\|\tilde{u}\|_{\mathbb{R}^m} = 1$. Since $\{u_i\}_{i=1}^m$ is an orthonormal basis in $\mathbb{R}^m$, we have

$$\tilde{u} = \sum_{i=1}^m \langle \tilde{u}, u_i \rangle_2 \, u_i.$$

Thus,

$$\begin{aligned}
\sum_{j=1}^n \left| \langle y_j, \tilde{u} \rangle_2 \right|^2 &= \sum_{j=1}^n \left| \Big\langle y_j, \sum_{i=1}^m \langle \tilde{u}, u_i \rangle_2 \, u_i \Big\rangle_2 \right|^2 = \sum_{j=1}^n \sum_{i=1}^m \sum_{k=1}^m \big(\langle y_j, \langle \tilde{u}, u_i \rangle_2 \, u_i \rangle_2 \langle y_j, \langle \tilde{u}, u_k \rangle_2 \, u_k \rangle_2 \big) \\
&= \sum_{j=1}^n \sum_{i=1}^m \sum_{k=1}^m \big(\langle y_j, u_i \rangle_2 \langle y_j, u_k \rangle_2 \langle \tilde{u}, u_i \rangle_2 \langle \tilde{u}, u_k \rangle_2 \big) = \sum_{i=1}^m \sum_{k=1}^m \Bigg(\Big\langle \underbrace{\sum_{j=1}^n \langle y_j, u_i \rangle_2 \, y_j}_{=\lambda_i u_i}, u_k \Big\rangle_2 \langle \tilde{u}, u_i \rangle_2 \langle \tilde{u}, u_k \rangle_2 \Bigg) \\
&= \sum_{i=1}^m \sum_{k=1}^m \Big(\underbrace{\langle \lambda_i u_i, u_k \rangle_2}_{=\lambda_i \delta_{ik}} \langle \tilde{u}, u_i \rangle_2 \langle \tilde{u}, u_k \rangle_2 \Big) = \sum_{i=1}^m \lambda_i \left| \langle \tilde{u}, u_i \rangle_2 \right|^2 \\
&\leq \lambda_1 \sum_{i=1}^m \left| \langle \tilde{u}, u_i \rangle_2 \right|^2 = \lambda_1 \, \|\tilde{u}\|_2^2 = \lambda_1 = \sum_{j=1}^n \left| \langle y_j, u_1 \rangle_2 \right|^2.
\end{aligned}$$

Consequently, u_1 solves ($\mathbf{P}^1$) and argmax ($\mathbf{P}^1$) $= \sigma_1^2 = \lambda_1$.

If we look for a second vector, orthogonal to u_1 that again describes the data set $\{y_i\}_{i=1}^n$ as well as possible then we need to solve

$$\max_{u \in \mathbb{R}^m} \sum_{j=1}^n \left| \langle y_j, u \rangle_2 \right|^2 \quad \text{s.t.} \quad \|u\|_2 = 1 \text{ and } \langle u, u_1 \rangle_2 = 0. \tag{$\mathbf{P}^2$}$$

SVD implies that u_2 is a solution to ($\mathbf{P}^2$) and argmax ($\mathbf{P}^2$) $= \sigma_2^2 = \lambda_2$. In fact, u_2 solves the first-order necessary optimality conditions (31.7) and for

$$\tilde{u} = \sum_{i=2}^m \langle \tilde{u}, u_i \rangle_2 \, u_i \in \operatorname{span}\{u_1\}^\perp$$

we have

$$\sum_{j=1}^n \left| \langle y_j, \tilde{u} \rangle_2 \right|^2 \leq \lambda_2 = \sum_{j=1}^n \left| \langle y_j, u_2 \rangle_2 \right|^2.$$

Clearly this procedure can be continued by finite induction. We summarize our results in the following theorem.

Theorem 31.2.1. *Let $Y = [y_1, \ldots, y_n] \in \mathbb{R}^{m \times n}$ be a given matrix with rank $d \leq \min\{m, n\}$. Further, let $Y = U \Sigma V^\top$ be the singular value decomposition of Y, where $U = [u_1, \ldots, u_m] \in \mathbb{R}^{m \times m}$, $V = [v_1, \ldots, v_n] \in \mathbb{R}^{n \times n}$ are orthogonal matrices and the matrix $\Sigma \in \mathbb{R}^{m \times n}$ has the form as (31.2). Then, for any $\ell \in \{1, \ldots, d\}$ the solution to*

$$\max_{\tilde{u}_1, \ldots, \tilde{u}_\ell \in \mathbb{R}^m} \sum_{i=1}^\ell \sum_{j=1}^n \left| \langle y_j, \tilde{u}_i \rangle_2 \right|^2 \quad \textit{s.t.} \quad \langle \tilde{u}_i, \tilde{u}_j \rangle_2 = \delta_{ij} \textit{ for } 1 \leq i, j \leq \ell \tag{$\mathbf{P}^\ell$}$$

is given by the singular vectors $\{u_i\}_{i=1}^\ell$, i.e., by the first ℓ columns of U. Moreover,

$$\operatorname{argmax}\,(\mathbf{P}^\ell) = \sum_{i=1}^\ell \sigma_i^2 = \sum_{i=1}^\ell \lambda_i. \tag{31.8}$$

Proof. Since ($\mathbf{P}^\ell$) is an equality constrained optimization problem, we introduce the Lagrangian

$$\mathcal{L} : \underbrace{\mathbb{R}^m \times \ldots \times \mathbb{R}^m}_{\ell\text{-times}} \times \mathbb{R}^{\ell \times \ell}$$

by

$$\mathcal{L}(\psi_1, \ldots, \psi_\ell, \Lambda) = \sum_{i=1}^\ell \sum_{j=1}^n \left| \langle y_j, \psi_i \rangle_2 \right|^2 + \sum_{i,j=1}^\ell \lambda_{ij} \big(\delta_{ij} - \langle \psi_i, \psi_j \rangle_2 \big)$$

for $\psi_1,\dots,\psi_\ell \in \mathbb{R}^m$ and $\Lambda = ((\lambda_{ij})) \in \mathbb{R}^{\ell\times\ell}$. First-order necessary optimality conditions for ($\mathbf{P}^\ell$) are given by

$$\frac{\partial \mathcal{L}}{\partial \psi_k}(\psi_1,\dots,\psi_\ell,\Lambda)\delta\psi_k = 0 \quad \text{for all } \delta\psi_k \in \mathbb{R}^m \text{ and } k \in \{1,\dots,\ell\}. \tag{31.9}$$

From

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \psi_k}(\psi_1,\dots,\psi_\ell,\Lambda)\delta\psi_k &= 2\sum_{i=1}^{\ell}\sum_{j=1}^{n}\langle y_j,\psi_i\rangle_2\langle y_j,\delta\psi_k\rangle_2\delta_{ik} - \sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\lambda_{ij}\langle\psi_i,\delta\psi_k\rangle_2\delta_{jk}\\ &\quad - \sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\lambda_{ij}\langle\delta\psi_k,\psi_j\rangle_2\delta_{ki}\\ &= 2\sum_{j=1}^{n}\langle y_j,\psi_k\rangle_2\langle y_j,\delta\psi_k\rangle_2 - \sum_{i=1}^{\ell}\left(\lambda_{ik}+\lambda_{ki}\right)\langle\psi_i,\delta\psi_k\rangle_2\\ &= \left\langle 2\sum_{j=1}^{n}\langle y_j,\psi_k\rangle_2\, y_j - \sum_{i=1}^{\ell}\left(\lambda_{ik}+\lambda_{ki}\right)\psi_i,\delta\psi_k\right\rangle_2\end{aligned}$$

and (31.9) we infer that

$$\sum_{j=1}^{n}\langle y_j,\psi_k\rangle_{\mathbb{R}^m}\, y_j = \frac{1}{2}\sum_{i=1}^{\ell}\left(\lambda_{ik}+\lambda_{ki}\right)\psi_i \quad \text{in } \mathbb{R}^m \text{ and for all } k\in\{1,\dots,\ell\}. \tag{31.10}$$

Note that

$$YY^\top\psi = \sum_{j=1}^{n}\langle y_j,\psi\rangle_{\mathbb{R}^m}\, y_j \quad \text{for } \psi\in\mathbb{R}^m.$$

Thus, condition (31.10) can be expressed as

$$YY^\top\psi_k = \frac{1}{2}\sum_{i=1}^{\ell}\left(\lambda_{ik}+\lambda_{ki}\right)\psi_i \quad \text{in } \mathbb{R}^m \text{ and for all } k\in\{1,\dots,\ell\}. \tag{31.11}$$

Now we proceed by induction. For $\ell = 1$ we have $k = 1$. It follows from (31.11) that

$$YY^\top\psi_1 = \lambda_1\psi_1 \quad \text{in } \mathbb{R}^m \tag{31.12}$$

with $\lambda_1 = \lambda_{11}$. Next we suppose that for $\ell \ge 1$ the first-order optimality conditions are given by

$$YY^\top\psi_k = \lambda_k\psi_k \quad \text{in } \mathbb{R}^m \text{ and for all } k\in\{1,\dots,\ell\}. \tag{31.13}$$

We want to show that the first-order necessary optimality conditions for a POD basis $\{\psi_i\}_{i=1}^{\ell+1}$ of rank $\ell+1$ are given by

$$YY^\top\psi_k = \lambda_k\psi_k \quad \text{in } \mathbb{R}^m \text{ and for all } k\in\{1,\dots,\ell+1\}. \tag{31.14}$$

By assumption we have (31.13). Thus, we only have to prove that

$$YY^\top\psi_{\ell+1} = \lambda_{\ell+1}\psi_{\ell+1} \quad \text{in } \mathbb{R}^m. \tag{31.15}$$

Due to (31.11) we have

$$YY^\top\psi_{\ell+1} = \frac{1}{2}\sum_{i=1}^{\ell+1}\left(\lambda_{i,\ell+1}+\lambda_{\ell+1,i}\right)\psi_i \quad \text{in } \mathbb{R}^m. \tag{31.16}$$

Since $\{\psi_i\}_{i=1}^{\ell+1}$ is a POD basis we have $\langle\psi_{\ell+1},\psi_j\rangle_2 = 0$ for $1\le j\le\ell$. Using (31.13) and the symmetry of $YY^\top$ we have for any $j\in\{1,\dots,\ell\}$

$$\begin{aligned}0 = \lambda_j\,\langle\psi_{\ell+1},\psi_j\rangle_2 &= \langle\psi_{\ell+1},YY^\top\psi_j\rangle_2 = \langle YY^\top\psi_{\ell+1},\psi_j\rangle_2\\ &= \frac{1}{2}\sum_{i=1}^{\ell+1}\left(\lambda_{i,\ell+1}+\lambda_{\ell+1,i}\right)\langle\psi_i,\psi_j\rangle_2 = \frac{1}{2}\left(\lambda_{j,\ell+1}+\lambda_{\ell+1,j}\right).\end{aligned}$$

This gives

$$\lambda_{\ell+1,i} = -\lambda_{i,\ell+1} \quad \text{for any } i \in \{1,\dots,\ell\}. \tag{31.17}$$

Inserting (31.17) into (31.16) we obtain

$$\begin{aligned} YY^\top \psi_{\ell+1} &= \frac{1}{2}\sum_{i=1}^{\ell} (\lambda_{i,\ell+1} + \lambda_{\ell+1,i})\,\psi_i + \lambda_{\ell+1,\ell+1}\,\psi_{\ell+1} \\ &= \frac{1}{2}\sum_{i=1}^{\ell} (\lambda_{i,\ell+1} - \lambda_{i,\ell+1})\,\psi_i + \lambda_{\ell+1,\ell+1}\,\psi_{\ell+1} = \lambda_{\ell+1,\ell+1}\,\psi_{\ell+1}. \end{aligned}$$

Setting $\lambda_{\ell+1} = \lambda_{\ell+1,\ell+1}$ we obtain (31.15).
Summarizing, the necessary optimaity conditions for ($\mathbf{P}^\ell$) are given by the symmetric $m \times m$ eigenvalue problem

$$YY^\top u_i = \lambda_i u_i \quad \text{for } i = 1,\dots,\ell. \tag{31.18}$$

It follows from SVD that $\{u_i\}_{i=1}^{\ell}$ solves (31.18). The proof that $\{u_i\}_{i=1}^{\ell}$ is a solution to ($\mathbf{P}^\ell$) and that argmax ($\mathbf{P}^\ell$) $= \sum_{i=1}^{\ell} \sigma_i^2$ holds is analogous to the proof for ($\mathbf{P}^1$); see Exercise 1.2). □

Motivated by the previous theorem we give the next definition.

Definition 31.2.2. For $\ell \in \{1,\dots,d\}$ the vectors $\{u_i\}_{i=1}^{\ell}$ are called *POD basis of rank* ℓ.

Lecture 32

Proper Orthogonal Decomposition – Properties and Applications

After introducing the POD method in the previous lecture we discuss now properties of the POD basis and applications to dynamical systems.

32.1 Optimality of the POD basis

The following result states that for every $\ell \leq d$ the approximation of the columns of Y by the first ℓ singular vectors $\{u_i\}_{i=1}^{\ell}$ is optimal in the mean among all rank ℓ approximations to the columns of Y.

Corollary 32.1.1 (Optimality of the POD basis)**.** *Let all hypotheses of Theorem 31.2.1 be satisfied. Suppose that $\hat{U}^d \in \mathbb{R}^{m\times d}$ denotes a matrix with pairwise orthonormal vectors $\hat{u}_i$ and that the expansion of the columns of Y in the basis $\{\hat{u}_i\}_{i=1}^{d}$ be given by*

$$Y = \hat{U}^d C^d, \quad \text{where } C^d_{ij} = \langle \hat{u}_i, y_j\rangle_2 \text{ for } 1 \leq i \leq d, 1 \leq j \leq n.$$

Then for every $\ell \in \{1,\dots,d\}$ we have

$$\|Y - U^\ell B^\ell\|_F \leq \|Y - \hat{U}^\ell C^\ell\|_F \tag{32.1}$$

with

$$B^d_{ij} = \langle u_i, y_j\rangle_2 \text{ for } 1 \leq i \leq d, 1 \leq j \leq n.$$

In (32.1), $\|\cdot\|_F$ denotes the Frobenius norm given by

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |A_{ij}|^2} = \sqrt{\operatorname{trace}\left(A^\top A\right)} \quad \text{for } A \in \mathbb{R}^{m\times n},$$

the matrix U^ℓ denotes the first ℓ columns of U, B^ℓ the first ℓ rows of B and similarly for $\hat{U}^\ell$ and C^ℓ.

Remark 32.1.2. Notice that

$$\|Y - \hat{U}^\ell C^\ell\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\left|Y_{ij} - \sum_{k=1}^{\ell} \hat{U}^\ell_{ik} C_{kj}\right|^2 = \sum_{j=1}^{n}\sum_{i=1}^{m}\left|Y_{ij} - \sum_{k=1}^{\ell}\langle \hat{u}_k, y_j\rangle_{\mathbb{R}^m} \hat{U}^\ell_{ik}\right|^2 = \sum_{j=1}^{n}\left\|y_j - \sum_{k=1}^{\ell}\langle y_j, \hat{u}_k\rangle_{\mathbb{R}^m}\hat{u}_k\right\|_2^2.$$

Analogously,

$$\|Y - U^\ell B^\ell\|_F^2 = \sum_{j=1}^{n}\left\|y_j - \sum_{k=1}^{\ell}\langle y_j, u_k\rangle_2 u_k\right\|_2^2.$$

Thus, (32.1) implies that

$$\sum_{j=1}^{n}\left\|y_j - \sum_{k=1}^{\ell}\langle y_j, u_k\rangle_2 u_k\right\|_2^2 \leq \sum_{j=1}^{n}\left\|y_j - \sum_{k=1}^{\ell}\langle y_j, \hat{u}_k\rangle_2 \hat{u}_k\right\|_2^2$$

for any other set $\{\hat{u}_i\}_{i=1}^{\ell}$ of ℓ pairwise orthonormal vectors. Hence, the POD basis of rank ℓ can also be determined by solving

$$\min_{\tilde{u}_1,\ldots,\tilde{u}_\ell\in\mathbb{R}^m} \sum_{j=1}^{n} \Big\| y_j - \sum_{i=1}^{\ell} \langle y_j, \tilde{u}_i\rangle_2\, \tilde{u}_i \Big\|_2^2 \text{ s.t. } \langle \tilde{u}_i, \tilde{u}_j\rangle_2 = \delta_{ij},\ 1\le i,j\le \ell. \tag{$\hat{\mathbf{P}}^\ell$}$$

◊

Proof of Corollary 32.1.1. Note that

$$\|Y - \hat{U}^\ell C^\ell\|_F^2 = \|\hat{U}^d(C^d - C_0^\ell)\|_F^2 = \|C^d - C_0^\ell\|_F^2 = \sum_{i=\ell+1}^{d}\sum_{j=1}^{n} |C_{ij}^d|^2,$$

where $C_0^\ell \in \mathbb{R}^{d\times n}$ results from $C\in\mathbb{R}^{d\times n}$ by replacing the last $d-\ell$ rows by 0. Similarly,

$$\begin{aligned}\|Y - U^\ell B^\ell\|_F^2 &= \|U^k(B^d - B_0^\ell)\|_F^2 = \|B^d - B_0^\ell\|_F^2 = \sum_{i=\ell+1}^{d}\sum_{j=1}^{n} |B_{ij}^d|^2 = \sum_{i=\ell+1}^{d}\sum_{j=1}^{n} |\langle y_j, u_i\rangle_2|^2 \\ &= \sum_{i=\ell+1}^{d}\sum_{j=1}^{n} \langle \langle y_j, u_i\rangle_2\, y_j, u_i\rangle_2 = \sum_{i=\ell+1}^{d} \langle YY^\top u_i, u_i\rangle_2 = \sum_{i=\ell+1}^{d} \sigma_i^2,\end{aligned} \tag{32.2}$$

By Theorem 31.2.1 the vectors $u_1,\ldots,u_\ell$ solve ($\mathbf{P}^\ell$). From (32.2),

$$\|Y\|_F^2 = \|\hat{U}^d C^d\|_F^2 = \|C^d\|_F^2 = \sum_{i=1}^{d}\sum_{j=1}^{n} |C_{ij}^d|^2$$

and

$$\|Y\|_F^2 = \|U^d B^d\|_F^2 = \|B^d\|_F^2 = \sum_{i=1}^{d}\sum_{j=1}^{n} |B_{ij}^d|^2 = \sum_{i=1}^{d} \sigma_i^2$$

we infer that

$$\begin{aligned}\|Y - U^\ell B^\ell\|_F^2 &= \sum_{i=\ell+1}^{d} \sigma_i^2 = \sum_{i=1}^{d}\sigma_i^2 - \sum_{i=1}^{\ell}\sigma_i^2 = \|Y\|_F^2 - \sum_{i=1}^{\ell}\sum_{j=1}^{n} |\langle y_j, u_i\rangle_2|^2 \\ &\le \|Y\|_F^2 - \sum_{i=1}^{\ell}\sum_{j=1}^{n} |\langle y_j, \hat{u}_i\rangle_2|^2 = \sum_{i=1}^{d}\sum_{j=1}^{n} |C_{ij}^d|^2 - \sum_{i=1}^{\ell}\sum_{j=1}^{n} |C_{ij}^d|^2 = \sum_{i=\ell+1}^{d}\sum_{j=1}^{n} |C_{ij}^d|^2 = \|Y - \hat{U}^\ell C^\ell\|_F^2,\end{aligned}$$

which gives (32.1). □

Remark 32.1.3. It follows from Corollary 32.1.1 that the POD basis of rank ℓ is optimal in the sense of representing in the mean the columns $\{y_j\}_{j=1}^{n}$ of Y as a linear combination by an orthonormal basis of rank ℓ:

$$\sum_{i=1}^{\ell}\sum_{j=1}^{n} |\langle y_j, u_i\rangle_2|^2 = \sum_{i=1}^{\ell}\sigma_i^2 = \sum_{i=1}^{\ell}\lambda_i \ge \sum_{i=1}^{\ell}\sum_{j=1}^{n} |\langle y_j, \hat{u}_i\rangle_2|^2$$

for any other set of orthonormal vectors $\{\hat{u}_i\}_{i=1}^{\ell}$. ◊

The next corollary states that the POD coefficients are uncorrelated.

Corollary 32.1.4 (Uncorrelated POD coefficients)**.** *Let all hypotheses of Theorem 31.2.1 hold. Then.*

$$\sum_{j=1}^{n} \langle y_j, u_i\rangle_2 \langle y_j, u_k\rangle_2 = \sum_{j=1}^{n} B_{ij}^\ell B_{kj}^\ell = \sigma_i^2\delta_{ik} \quad \textit{for } 1\le i,k\le \ell.$$

Proof. The claim follows from (31.18) and $\langle u_i, u_k\rangle_{\mathbb{R}^m} = \delta_{ik}$ for $1\le i,k\le\ell$. In fact, we have

$$\sum_{j=1}^{n} \langle y_j, u_i\rangle_2 \langle y_j, u_k\rangle_2 = \Big\langle \underbrace{\sum_{j=1}^{n} \langle y_j, u_i\rangle_2\, y_j}_{=YY^\top u_i}, u_k \Big\rangle_2 = \langle \sigma_i^2 u_i, u_k\rangle_2 = \sigma_i^2\delta_{ik}.$$

□

Next we turn to the practical computation of a POD-basis of rank ℓ. If $n < m$ then one can determine the POD basis of rank ℓ as follows: Compute the eigenvectors $v_1, \ldots, v_\ell \in \mathbb{R}^n$ by solving the symmetric $n \times n$ eigenvalue problem

$$Y^\top Y v_i = \lambda_i v_i \quad \text{for } i = 1, \ldots, \ell \tag{32.3}$$

and set, by (31.3),

$$u_i = \frac{1}{\sqrt{\lambda_i}} Y v_i \quad \text{for } i = 1, \ldots, \ell.$$

For historical reasons [29] this method of determing the POD-basis is sometimes called the *method of snapshots*. On the other hand, if $m < n$ holds, we can obtain the POD basis by solving the $m \times m$ eigenvalue problem (31.18).

For the application of POD to concrete problems the choice of ℓ is certainly of central importance for applying POD. It appears that no general a-priori rules are available. Rather the choice of ℓ is based on heuristic considerations combined with observing the ratio of the modeled to the total energy contained in the system Y, which is expressed by

$$\mathcal{E}(\ell) = \frac{\sum_{i=1}^{\ell} \lambda_i}{\sum_{i=1}^{d} \lambda_i}.$$

Let us mention that POD is also called *Principal Component Analysis* (PCA) and *Karhunen-Loève Decomposition*.

32.2 Application to dynamical systems

For $T > 0$ we consider the semi-linear initial value problem

$$\dot{y}(t) = Ay(t) + f(t, y(t)) \quad \text{for } t \in (0, T], \tag{32.4a}$$

$$y(0) = y_0, \tag{32.4b}$$

where $y_0 \in \mathbb{R}^m$ is a chosen initial condition, $A \in \mathbb{R}^{m\times m}$ is a given matrix, $f : [0, T] \times \mathbb{R}^m \to \mathbb{R}^m$ is continuous in both arguments and locally Lipschitz-continuous with respect to the second argument. It is well known that there exists a time $T_\circ \in (, T]$ such that (32.4) has a unique (classical) solution $y \in C^1(0, T_\circ; \mathbb{R}^m) \cap C([0, T_\circ]; \mathbb{R}^m)$ given by the implicit integral representation

$$y(t) = e^{tA} y_0 + \int_0^t e^{(t-s)A} f(s, y(s))\,\mathrm{d}s \quad \text{for } t \in [0, T_\circ],$$

with $e^{tA} = \sum_{i=0}^{\infty} t^n A^n/(n!)$ (local existence in time; cf. [11, Satz 16.5]). Here we suppose that we can choose $T_\circ = T$ (global existence in time; cf. [11, Satz 16.1]). Let $0 \le t_1 < t_2 < \ldots < t_n \le T$ be a given time grid in the interval $[0, T]$. For simplicity of the presentation, the time grid is assumed to be equidistant with step-size $\Delta t = T/(n - 1)$, i.e., $t_j = (j - 1)\Delta t$. We suppose that we know the solution to (32.4) at the given time instances t_j, $j \in \{1, \ldots, n\}$. Our goal is to determine a POD basis of rank $\ell \le n$ that desribes the ensemble

$$y_j = y(t_j) = e^{t_j A} y_0 + \int_0^{t_j} e^{(t_j - s)A} f(s, y(s))\,\mathrm{d}s \quad \text{for } j = 1, \ldots, n,$$

as well as possible with respect to the weighted inner product:

$$\min_{\tilde{u}_1, \ldots, \tilde{u}_\ell \in \mathbb{R}^m} \sum_{j=1}^{n} \alpha_j \left\| y_j - \sum_{i=1}^{\ell} \langle y_j, \tilde{u}_i \rangle_2 \, \tilde{u}_i \right\|_2^2 \quad \text{s.t.} \quad \langle \tilde{u}_i, \tilde{u}_j \rangle_2 = \delta_{ij} \text{ for } 1 \le i, j \le \ell, \tag{$\hat{\mathbf{P}}^{n,\ell}$}$$

where the α_j's denote non-negative weights which will be specified later on. Note that for $\alpha_j = 1$ for $j = 1, \ldots, n$ problem ($\hat{\mathbf{P}}^{n,\ell}$) coincides with ($\hat{\mathbf{P}}^{\ell}$). To solve ($\hat{\mathbf{P}}^{n,\ell}$) we apply the techniques used in Chapter 29, i.e., we use the Lagrangian framework. Thus, we introduce the Lagrange functional

$$\mathcal{L} : \underbrace{\mathbb{R}^m \times \ldots \times \mathbb{R}^m}_{\ell-\text{times}} \times \mathbb{R}^{\ell\times\ell} \to \mathbb{R}, \quad \mathcal{L}(u_1, \ldots, u_\ell, \Lambda) = \sum_{j=1}^{n} \alpha_j \left\| y_j - \sum_{i=1}^{\ell} \langle y_j, u_i \rangle_2 \, u_i \right\|_2^2 + \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \Lambda_{ij} \big(1 - \langle u_i, u_j \rangle_2\big)$$

for $u_1, \ldots, u_\ell \in \mathbb{R}^m$ and $\Lambda \in \mathbb{R}^{\ell\times\ell}$ with elements Λ_{ij}, $1 \le i, j \le \ell$. It turns out that the solution to ($\hat{\mathbf{P}}^{n,\ell}$) is given by the first-order necessary optimality condions

$$\nabla_{u_i} \mathcal{L}(u_1, \ldots, u_\ell, \Lambda) \stackrel{!}{=} 0 \quad \text{in } \mathbb{R}^m \text{ for } 1 \le i \le \ell, \tag{32.5a}$$

and

$$\langle u_i, u_j \rangle_2 \stackrel{!}{=} \delta_{ij} \quad \text{for } 1 \leq i, j \leq \ell. \tag{32.5b}$$

From (32.5a) we derive

$$Y D Y^\top u_i = \lambda_i u_i \quad \text{for } i = 1, \ldots, \ell, \tag{32.6}$$

where $D = \operatorname{diag}(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^{n \times n}$. Setting $\bar{Y} = Y D^{1/2} \in \mathbb{R}^{m \times n}$ and using $D^\top = D$ we infer from (32.6) that the solution $\{u_i\}_{i=1}^{\ell}$ to ($\hat{\mathbf{P}}^{n,\ell}$) is given by the symmetric $m \times m$ eigenvalue problem

$$\bar{Y} \bar{Y}^\top \bar{u}_i = \lambda_i \bar{u}_i, \ 1 \leq i \leq \ell \quad \text{and} \quad \langle \bar{u}_i, \bar{u}_j \rangle_2 = \delta_{ij}, \ 1 \leq i, j \leq \ell.$$

Note that

$$\bar{Y}^\top \bar{Y} = D^{1/2} Y^\top Y D^{1/2} \in \mathbb{R}^{n \times n}.$$

Thus, the POD basis of rank ℓ can also be computed by the methods of snapshots as follows: First solve the symmetric $n \times n$ eigenvalue problem

$$\bar{Y}^\top \bar{Y} \bar{v}_i = \lambda_i \bar{v}_i \text{ for } 1 \leq i \leq \ell \quad \text{and} \quad \langle \bar{v}_i, \bar{v}_j \rangle_2 = \delta_{ij} \text{ for } 1 \leq i, j \leq \ell.$$

Then we set (by SVD)

$$\bar{u}_i = \frac{1}{\sqrt{\lambda_i}} \bar{Y} \bar{v}_i = \frac{1}{\sqrt{\lambda_i}} Y D^{1/2} \bar{v}_i \quad \text{for } 1 \leq i \leq \ell.$$

Lecture 33

Proper Orthogonal Decomposition – Continuous Variant

33.1 Continuous variant of the POD method

Of course, the snapshot ensemble $\{y_j\}_{j=1}^n$ for ($\hat{\mathbf{P}}^{n,\ell}$) and therefore the snapshot set $\operatorname{span}\{y_1,\ldots,y_n\}$ depend on the chosen time instances $\{t_j\}_{j=1}^n$. Consequently, the POD basis vectors $\{u_i\}_{i=1}^\ell$ and the corresponding eigenvalues $\{\lambda_i\}_{i=1}^\ell$ depend also on the time instances, i.e.,

$$u_i = u_i^n \quad \text{and} \quad \lambda_i = \lambda_i^n, \quad 1 \le i \le \ell.$$

Moreover, we have not discussed so far what is the motivation to introduce the non-negative weights $\{\alpha_j\}_{j=1}^n$ in ($\hat{\mathbf{P}}^{n,\ell}$). For this reason we proceed by investigating the following two questions:

- How to choose good time instances for the snapshots?
- What are appropriate non-negative weights $\{\alpha_j\}_{j=1}^n$?

To address these two questions we will introduce a *continuous version* of POD. Let $y : [0,T] \to \mathbb{R}^m$ be the unique solution to (32.4). If we are interested to find a POD basis of rank ℓ that describes the whole trajectory $\{y(t)\,|\,t \in [0,T]\} \subset \mathbb{R}^m$ as good as possible we have to consider the following minimization problem

$$\min_{\tilde u_1,\ldots,\tilde u_\ell \in \mathbb{R}^m} \int_0^T \Big\| y(t) - \sum_{i=1}^\ell \langle y(t), \tilde u_i\rangle_2\, \tilde u_i \Big\|_2^2 \,\mathrm{d}t \quad \text{s.t.} \quad \langle \tilde u_i, \tilde u_j\rangle_2 = \delta_{ij},\ 1 \le i,j \le \ell, \tag{$\hat{\mathbf{P}}^\ell$}$$

To solve ($\hat{\mathbf{P}}^\ell$) we use similar arguments as in Chapter 29. For $\ell = 1$ we obtain instead of ($\hat{\mathbf{P}}^\ell$) the minimization problem

$$\min_{\tilde u \in \mathbb{R}^m} \int_0^T \Big\| y(t) - \langle y(t), \tilde u\rangle_2\, \tilde u \Big\|_2^2 \,\mathrm{d}t \quad \text{s.t.} \quad \|\tilde u\|_2^2 = 1, \tag{$\hat{\mathbf{P}}^1$}$$

Suppose that $\{\tilde u_i\}_{i=2}^m$ are chosen in such a way that $\{\tilde u, \tilde u_2, \ldots, \tilde u_m\}$ is an orthonormal basis in $\mathbb{R}^m$ with respect to the Euclidean inner product. Then we have

$$y(t) = \langle y(t), \tilde u\rangle_2\, \tilde u + \sum_{i=2}^m \langle y(t), \tilde u_i\rangle_2\, \tilde u_i \quad \text{for all } t \in [0,T].$$

Thus,

$$\int_0^T \Big\| y(t) - \langle y(t), \tilde u\rangle_2\, \tilde u \Big\|_2^2 \,\mathrm{d}t = \int_0^T \Big\| \sum_{i=2}^m \langle y(t), \tilde u\rangle_2\, \tilde u_i \Big\|_2^2 \,\mathrm{d}t = \sum_{i=2}^m \int_0^T \big| \langle y(t), \tilde u_i\rangle_2 \big|^2 \,\mathrm{d}t$$

we conclude that ($\hat{\mathbf{P}}^1$) is equivalent with the following maximization problem

$$\max_{\tilde u \in \mathbb{R}^m} \int_0^T \big| \langle y(t), \tilde u\rangle_2 \big|^2 \,\mathrm{d}t \quad \text{s.t.} \quad \|\tilde u\|_2^2 = 1. \tag{33.1}$$

The Lagrange functional $\mathcal{L} : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}$ associated with (33.1) is given by

$$\mathcal{L}(u,\lambda) = \int_0^T \big|\langle y(t), u\rangle_2\big|^2\,\mathrm{d}t + \lambda\big(1 - \|u\|_2^2\big) \quad \text{for } (u,\lambda) \in \mathbb{R}^m \times \mathbb{R}.$$

First-order necessary optimality conditions are given by

$$\nabla\mathcal{L}(u,\lambda) \overset{!}{=} 0 \quad \text{in } \mathbb{R}^m \times \mathbb{R}.$$

Therefore, we compute the partial derivative of $\mathcal{L}$ with respect to the ith component u_i of the vector u:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial u_i}(u,\lambda) &= \frac{\partial}{\partial u_i}\left(\int_0^T \Big|\sum_{k=1}^m y_k(t)u_k\Big|^2\,\mathrm{d}t + \lambda\Big(1 - \sum_{k=1}^m u_k^2\Big)\right) = 2\int_0^T \Big(\sum_{k=1}^m y_k(t)u_k\Big)y_i(t)\,\mathrm{d}t - 2\lambda u_1\\ &= 2\bigg(\int_0^T \langle y(t), u\rangle_2\, y(t)\,\mathrm{d}t - \lambda u\bigg)_i\end{aligned}$$

for $i \in \{1,\dots,m\}$. Thus,

$$\nabla_u\mathcal{L}(u,\lambda) = 2\bigg(\int_0^T \langle y(t), u\rangle_2\, y(t)\,\mathrm{d}t - \lambda u\bigg) \overset{!}{=} 0 \quad \text{in } \mathbb{R}^m,$$

which gives

$$\int_0^T \langle y(t), u\rangle_2\, y(t)\,\mathrm{d}t = \lambda u \quad \text{in } \mathbb{R}^m. \tag{33.2}$$

We define the operator $\mathcal{R} : \mathbb{R}^m \to \mathbb{R}^m$ as

$$\mathcal{R}u = \int_0^T \langle y(t), u\rangle_2\, y(t)\,\mathrm{d}t \quad \text{for } u \in \mathbb{R}^m. \tag{33.3}$$

Lemma 33.1.1. The operator $\mathcal{R}$ is linear and bounded (i.e., continuous). Moreover,

1) $\mathcal{R}$ is non-negative:

$$\langle\mathcal{R}u, u\rangle_2 \geq 0 \quad \text{for all } u \in \mathbb{R}^m.$$

2) $\mathcal{R}$ is self-adjoint (or symmetric):

$$\langle\mathcal{R}u, \tilde u\rangle_2 = \langle u, \mathcal{R}\tilde u\rangle_2 \quad \text{for all } u,\, \tilde u \in \mathbb{R}^m.$$

Proof. For arbitrary $u,\, \tilde u \in \mathbb{R}^m$ and $\alpha,\, \tilde\alpha \in \mathbb{R}$ we have

$$\begin{aligned}\mathcal{R}\big(\alpha u + \tilde\alpha\tilde u\big) &= \int_0^T \langle y(t), \alpha u + \tilde\alpha\tilde u\rangle_2\, y(t)\,\mathrm{d}t = \int_0^T \big(\alpha\,\langle y(t), u\rangle_2 + \tilde\alpha\,\langle y(t), \tilde u\rangle_2\big)\, y(t)\,\mathrm{d}t\\ &= \alpha\int_0^T \langle y(t), u\rangle_2\, y(t)\,\mathrm{d}t + \tilde\alpha\int_0^T \langle y(t), \tilde u\rangle_2\, y(t)\,\mathrm{d}t = \alpha\mathcal{R}u + \tilde\alpha\mathcal{R}\tilde u,\end{aligned}$$

so that $\mathcal{R}$ is linear. From the Cauchy-Schwarz inequality (cf. [11, Satz 5.49]) we derive

$$\begin{aligned}\|\mathcal{R}u\|_2 &\leq \int_0^T \big\|\langle y(t), u\rangle_2\, y(t)\big\|_2\,\mathrm{d}t = \int_0^T \big|\langle y(t), u\rangle_2\big|\,\|y(t)\|_2\,\mathrm{d}t\\ &\leq \int_0^T \|y(t)\|_2^2\|u\|_2\,\mathrm{d}t = \bigg(\int_0^T \|y(t)\|_2^2\,\mathrm{d}t\bigg)\|u\|_2 = \|y\|^2_{L^2(0,T;\mathbb{R}^m)}\|u\|_2\end{aligned}$$

for an arbitrary $u \in \mathbb{R}^m$. Since $y \in C([0,T];\mathbb{R}^m) \subset L^2(0,T;\mathbb{R}^m)$ holds, the norm $\|y\|_{L^2(0,T;\mathbb{R}^m)}$ is bounded. Therefore, $\mathcal{R}$ is bounded. Since

$$\langle\mathcal{R}u, u\rangle_2 = \bigg(\int_0^T \langle y(t), u\rangle_2\, y(t)\,\mathrm{d}t\bigg)^\top u = \int_0^T \langle y(t), u\rangle_2\, y(t)^\top u\,\mathrm{d}t = \int_0^T \big|\langle y(t), u\rangle_2\big|^2\,\mathrm{d}t \geq 0$$

for all $u \in \mathbb{R}^m$ holds, $\mathcal{R}$ is non-negative. Finally, we infer from

$$\langle\mathcal{R}u, \tilde u\rangle_2 = \int_0^T \langle y(t), u\rangle_2\langle y(t), \tilde u\rangle_2\,\mathrm{d}t = \bigg\langle \int_0^T \langle y(t), \tilde u\rangle_2 y(t)\,\mathrm{d}t, u\bigg\rangle_2 = \langle\mathcal{R}\tilde u, u\rangle_2 = \langle u, \mathcal{R}\tilde u\rangle_2$$

for all $u,\, \tilde u \in \mathbb{R}^m$ that $\mathcal{R}$ is self-adjoint. □

Utilizing the operator $\mathcal{R}$ we can write (33.2) as the eigenvalue problem

$$\mathcal{R}u = \lambda u \quad \text{in } \mathbb{R}^m.$$

It follows from Lemma 33.1.1 that $\mathcal{R}$ possesses eigenvectors $\{u_i\}_{i=1}^m$ and associated real eigenvalues $\{\lambda_i\}_{i=1}^m$ such that

$$\mathcal{R}u_i = \lambda_i u_i \text{ for } 1 \leq i \leq m \quad \text{and} \quad \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq 0. \tag{33.4}$$

Note that

$$\int_0^T \left|\langle y(t), u_i\rangle_2\right|^2 \mathrm{d}t = \int_0^T \left\langle \langle y(t), u_i\rangle_2 y(t), u_i\right\rangle_2 \mathrm{d}t = \langle \mathcal{R}u_i, u_i\rangle_2 = \lambda_i \|u_i\|_2^2 = \lambda_i$$

for $i \in \{1, \ldots, m\}$ so that u_1 solves ($\hat{\mathbf{P}}^1$).

Proceeding as in Chapter 29 we obtain the following result.

Theorem 33.1.2. *Let $y \in C([0,T];\mathbb{R}^m)$ be the unique solution to (32.4). Then the POD basis of rank ℓ solving the minimization problem ($\hat{\mathbf{P}}^\ell$) is given by the eigenvectors $\{u_i\}_{i=1}^\ell$ of $\mathcal{R}$ corresponding to the ℓ largest eigenvalues $\lambda_1 \geq \ldots \geq \lambda_\ell$.*

Remark 33.1.3 (Methods of snapshots)**.** Let us introduce the linear and bounded operator $\mathcal{Y} : L^2(0,T) \to \mathbb{R}^m$ by

$$\mathcal{Y}v = \int_0^T v(t)y(t)\,\mathrm{d}t \quad \text{for } v \in L^2(0,T).$$

The adjoint $\mathcal{Y}^\star : \mathbb{R}^m \to L^2(0,T)$ satisfying

$$\langle \mathcal{Y}^\star u, v\rangle_{L^2(0,T)} = \langle u, \mathcal{Y}v\rangle_2 \quad \text{for all } (u,v) \in \mathbb{R}^m \times L^2(0,T)$$

is given as

$$(\mathcal{Y}^\star u)(t) = \langle u, y(t)\rangle_2 \quad \text{for } u \in \mathbb{R}^m \text{ and almost all } t \in [0,T].$$

Then we have

$$\mathcal{Y}\mathcal{Y}^\star u = \int_0^T \langle u, y(t)\rangle_2\, y(t)\,\mathrm{d}t = \int_0^T \langle y(t), u\rangle_2\, y(t)\,\mathrm{d}t = \mathcal{R}u$$

for all $u \in \mathbb{R}^m$, i.e., $\mathcal{R} = \mathcal{Y}\mathcal{Y}^*$ holds. Furthermore,

$$(\mathcal{Y}^\star \mathcal{Y}v)(t) = \left\langle \int_0^T v(s)y(s)\,\mathrm{d}s, y(t)\right\rangle_2 = \int_0^T \langle y(s), y(t)\rangle_2\, v(s)\,\mathrm{d}s =: (\mathcal{K}v)(t)$$

for all $v \in L^2(0,T)$ and almost all $t \in [0,T]$. Thus, $\mathcal{K} = \mathcal{Y}^\star\mathcal{Y}$. It can be shown that the operator $\mathcal{K}$ is linear, bounded, non-negative and self-adjoint. Moreover, $\mathcal{K}$ is compact. Therefore, the POD basis can also be computed as follows: Solve

$$\mathcal{K}v_i = \lambda_i v_i \text{ for } 1 \leq i \leq \ell, \quad \lambda_1 \geq \ldots \geq \lambda_\ell > 0, \quad \int_0^T v_i(t)v_j(t)\,\mathrm{d}t = \delta_{ij} \tag{33.5}$$

and set

$$u_i = \frac{1}{\sqrt{\lambda_i}}\mathcal{Y}v_i = \frac{1}{\sqrt{\lambda_i}}\int_0^T v_i(t)y(t)\,\mathrm{d}t \quad \text{for } i = 1, \ldots, \ell.$$

Note that (33.5) is a symmetric eigenvalue problem in the infinite-dimensional function space $L^2(0,T)$. For the functional analytic theory we refer, e.g., to [27]. ♢

33.2 Perturbation theory

Let us turn back to the optimality conditions (32.6). For any $u \in \mathbb{R}^m$ and $i \in \{1, \ldots, m\}$ we derive

$$\left(YDY^\top u\right)_i = \sum_{j=1}^m \sum_{k=1}^m \alpha_j Y_{ij} Y_{kj} u_k = \sum_{j=1}^n \alpha_j Y_{ij}\, \langle y_j, u\rangle_2 = \sum_{j=1}^n \alpha_j\, \langle y_j, u\rangle_2\, (y_j)_i,$$

where $(y_j)_i$ stands for the ith component of the vector $y_j \in \mathbb{R}^m$. Thus,

$$YDY^\top u = \sum_{j=1}^n \alpha_j \langle y_j, u\rangle_2 \, y_j =: \mathcal{R}^n u.$$

Note that the operator $\mathcal{R}^n : \mathbb{R}^m \to \mathbb{R}^m$ is linear and bounded. Moreover,

$$\langle \mathcal{R}^n u, u\rangle_2 = \Big\langle \sum_{j=1}^n \alpha_j \langle y_j, u\rangle_2 \, y_j, u \Big\rangle_2 = \sum_{j=1}^n \alpha_j \left|\langle y_j, u\rangle_2\right|^2 \geq 0$$

holds for all $u \in \mathbb{R}^m$ so that $\mathcal{R}^n$ is non-negative. Further,

$$\langle \mathcal{R}^n u, \tilde u\rangle_2 = \Big\langle \sum_{j=1}^n \alpha_j \langle y_j, u\rangle_2 \, y_j, \tilde u \Big\rangle_2 = \sum_{j=1}^n \alpha_j \langle y_j, u\rangle_2 \langle y_j, \tilde u\rangle_W = \Big\langle \sum_{j=1}^n \alpha_j \langle y_j, \tilde u\rangle_2 \, y_j, u \Big\rangle_2 = \langle \mathcal{R}^n \tilde u, u\rangle_2 = \langle u, \mathcal{R}^n \tilde u\rangle_2$$

for all u, $\tilde u \in \mathbb{R}^m$, i.e., $\mathcal{R}^n$ is self-adjoint. Therefore, $\mathcal{R}^n$ has the same properties as the operator $\mathcal{R}$. Summarizing, we have

$$\mathcal{R}^n u_i^n = \lambda_i^n u_i^n, \qquad \lambda_1^n \geq \ldots \lambda_\ell^n \geq \ldots \lambda_{d(n)}^n > \lambda_{d(n)+1}^n = \ldots = \lambda_m^n = 0, \tag{33.6a}$$

$$\mathcal{R} u_i = \lambda_i u_i, \qquad \lambda_1 \geq \ldots \lambda_\ell \geq \ldots \lambda_d > \lambda_{d+1} = \ldots = \lambda_m = 0. \tag{33.6b}$$

Let us note that

$$\int_0^T \|y(t)\|_2^2 \, \mathrm{d}t = \sum_{i=1}^d \lambda_i = \sum_{i=1}^m \lambda_i. \tag{33.7}$$

In fact,

$$\mathcal{R} u_i = \int_0^T \langle y(t), u_i\rangle_2 \, y(t) \, \mathrm{d}t \quad \text{for every } i \in \{1, \ldots, m\}.$$

Taking the inner product with u_i, using (33.6b) and summing over i we arrive at

$$\sum_{i=1}^d \int_0^T \left|\langle y(t), u_i\rangle_2\right|^2 \mathrm{d}t = \sum_{i=1}^d \langle \mathcal{R} u_i, u_i\rangle_2 = \sum_{i=1}^d \lambda_i = \sum_{i=1}^m \lambda_i.$$

Expanding $y(t) \in \mathbb{R}^m$ in terms of $\{u_i\}_{i=1}^m$ we have

$$y(t) = \sum_{i=1}^m \langle y(t), u_i\rangle_2 \, u_i$$

and hence

$$\int_0^T \|y(t)\|_2^2 \, \mathrm{d}t = \sum_{i=1}^m \int_0^T \left|\langle y(t), u_i\rangle_2\right|^2 \mathrm{d}t = \sum_{i=1}^m \lambda_i,$$

which is (33.7). Analogously, we obtain

$$\sum_{j=1}^n \alpha_j \, \|y(t_j)\|_2^2 = \sum_{i=1}^{d(n)} \lambda_i^n = \sum_{i=1}^m \lambda_i^n \quad \text{for every } n \in \mathbb{N}. \tag{33.8}$$

For convenience we do not indicate the dependence of α_j on n. Let $y \in C([0,T]; \mathbb{R}^m)$ hold. To ensure

$$\sum_{j=1}^n \alpha_j \, \|y(t_j)\|_2^2 \to \int_0^T \|y(t)\|_2^2 \, \mathrm{d}t \quad \text{as } \Delta t \to 0 \tag{33.9}$$

we have to choose the α_j's appropriately. Here we take the trapezoidal weights

$$\alpha_1 = \frac{\Delta t}{2}, \quad \alpha_j = \Delta t \text{ for } 2 \leq j \leq n-1, \quad \alpha_n = \frac{\Delta t}{2}. \tag{33.10}$$

Suppose that we have

$$\lim_{n\to\infty} \|\mathcal{R}^n - \mathcal{R}\|_{\mathscr{L}(\mathbb{R}^m)} = \lim_{n\to\infty} \sup_{\|u\|_2=1} \|\mathcal{R}^n u - \mathcal{R}u\|_2 = 0 \tag{33.11}$$

provided $y \in C^1([0,T];\mathbb{R}^m)$ is satisfied. In (33.11) the linear space $\mathscr{L}(\mathbb{R}^m)$ denotes the Banach space of all linear and bounded operators mapping from $\mathbb{R}^m$ into itself. Combining (33.9) with (33.7) and (33.8) we find

$$\sum_{i=1}^{m} \lambda_i^n \to \sum_{i=1}^{m} \lambda_i \quad \text{as } n \to \infty. \tag{33.12}$$

Now choose and fix

$$\ell \quad \text{such that} \quad \lambda_\ell \neq \lambda_{\ell+1}. \tag{33.13}$$

Then by spectral analysis of compact operators ([21, pp. 212–214]) and (33.11) it follows that

$$\lambda_i^n \to \lambda_i \quad \text{for } 1 \leq i \leq \ell \text{ as } n \to \infty. \tag{33.14}$$

Combining (33.12) and (33.14) there exists $\bar{n} \in \mathbb{N}$ such that

$$\sum_{i=\ell+1}^{m} \lambda_i^n \leq 2 \sum_{i=\ell+1}^{m} \lambda_i \quad \text{for all } n \geq \bar{n}, \tag{33.15}$$

if $\sum_{i=\ell+1}^{m} \lambda_i \neq 0$. Moreover, for ℓ as above, $\bar{n}$ can also be chosen such that

$$\sum_{i=\ell+1}^{d(n)} \left|\langle y_0, u_i^n\rangle_2\right|^2 \leq 2 \sum_{i=\ell+1}^{m} \left|\langle y_0, u_i\rangle_2\right|^2 \quad \text{for all } n \geq \bar{n}, \tag{33.16}$$

provided that $\sum_{i=\ell+1}^{m} |\langle y_0, u_i\rangle_2|^2 \neq 0$ (33.11) hold. Recall that the vector $y_0 \in \mathbb{R}^m$ stands for the initial condition in (32.4b). Then we have

$$\|y_0\|_2^2 = \sum_{i=1}^{m} \left|\langle y_0, u_i\rangle_2\right|^2. \tag{33.17}$$

If $t_1 = 0$ holds, we have $y_0 \in \operatorname{span}\{y_j\}_{j=1}^n$ for every n and

$$\|y_0\|_2^2 = \sum_{i=1}^{d(n)} \left|\langle y_0, u_i^n\rangle_2\right|^2. \tag{33.18}$$

Therefore, for $\ell < d(n)$ by (33.17) and (33.18)

$$\begin{aligned}
\sum_{i=\ell+1}^{d(n)} \left|\langle y_0, u_i^n\rangle_2\right|^2 &= \sum_{i=1}^{d(n)} \left|\langle y_0, u_i^n\rangle_2\right|^2 - \sum_{i=1}^{\ell} \left|\langle y_0, u_i^n\rangle_2\right|^2 + \sum_{i=1}^{\ell} \left|\langle y_0, u_i\rangle_2\right|^2 + \sum_{i=\ell+1}^{m} \left|\langle y_0, u_i\rangle_2\right|^2 - \sum_{i=1}^{m} \left|\langle y_0, u_i\rangle_2\right|^2 \\
&= \sum_{i=1}^{\ell} \left(\left|\langle y_0, u_i\rangle_2\right|^2 - \left|\langle y_0, u_i^n\rangle_2\right|^2\right) + \sum_{i=\ell+1}^{m} \left|\langle y_0, u_i\rangle_2\right|^2.
\end{aligned}$$

As a consequence of (33.11) and (33.13) we have $\lim_{n\to\infty} \|u_i^n - u_i\|_2 = 0$ for $i = 1, \ldots, \ell$ and hence (33.16) follows.

Theorem 33.2.1. *Assume that the function $y \in C^1([0,T];\mathbb{R}^m)$ is the unique solution to (32.4). Let $\{(u_i^n, \lambda_i^n)\}_{i=1}^m$ and $\{(u_i, \lambda_i)\}_{i=1}^m$ be the eigenvector-eigenvalue pairs given by (33.6). Suppose that $\ell \in \{1, \ldots, m\}$ is fixed such that (33.13) and*

$$\sum_{i=\ell+1}^{m} \lambda_i \neq 0, \qquad \sum_{i=\ell+1}^{m} \left|\langle y_0, u_i\rangle_W\right|^2 \neq 0$$

hold. Then we have

$$\lim_{n\to\infty} \|\mathcal{R}^n - \mathcal{R}\|_{\mathscr{L}(\mathbb{R}^m)} = 0. \tag{33.19}$$

This implies

$$\lim_{n\to\infty} \left|\lambda_i^n - \lambda_i\right| = \lim_{n\to\infty} \|u_i^n - u_i\|_2 = 0 \quad \textit{for } 1 \leq i \leq \ell,$$

$$\lim_{n\to\infty} \sum_{i=\ell+1}^{m} \left(\lambda_i^n - \lambda_i\right) = 0 \quad \textit{and} \quad \lim_{n\to\infty} \sum_{i=\ell+1}^{m} \left|\langle y_0, u_i^n\rangle_2\right|^2 = \sum_{i=\ell+1}^{m} \left|\langle y_0, u_i\rangle_2\right|^2.$$

Proof. We only have to verify (33.19). For that purpose we choose an arbitrary $u \in \mathbb{R}^m$ with $\|u\|_W = 1$ and introduce $f_u : [0,T] \to \mathbb{R}^m$ by

$$f_u(t) = \langle y(t), u \rangle_2 \, y(t) \quad \text{for } t \in [0,T].$$

Then, we have $f_u \in C^1([0,T];\mathbb{R}^m)$ with

$$\dot f_u(t) = \langle \dot y(t), u \rangle_2 \, y(t) + \langle y(t), u \rangle_2 \, \dot y(t) \quad \text{for } t \in [0,T]$$

By Taylor expansion there exist $\tau_{j1}(t), \tau_{j2}(t) \in [t_j, t_{j+1}]$ depending on t

$$\begin{aligned}\int_{t_j}^{t_{j+1}} f_u(t)\,\mathrm{d}t &= \frac{1}{2}\int_{t_j}^{t_{j+1}} f_u(t_j) + \dot f_u(\tau_{j1}(t))(t-t_j)\,\mathrm{d}t + \frac{1}{2}\int_{t_j}^{t_{j+1}} f_u(t_{j+1}) + \dot f_u(\tau_{j2}(t))(t-t_{j+1})\,\mathrm{d}t \\ &= \frac{\Delta t}{2}\left(f_u(t_j) + f_u(t_{j+1})\right) + \frac{1}{2}\int_{t_j}^{t_{j+1}} \dot f_u(\tau_{j1}(t))(t-t_j)\,\mathrm{d}t + \frac{1}{2}\int_{t_j}^{t_{j+1}} \dot f_u(\tau_{j2}(t))(t-t_{j+1})\,\mathrm{d}t.\end{aligned}$$

Hence,

$$\begin{aligned}\big\|\mathcal{R}^n u - \mathcal{R}u\big\|_2 &= \left\|\sum_{j=1}^n \alpha_j f_u(t_j) - \int_0^T f_u(t)\,\mathrm{d}t\right\|_2 = \left\|\sum_{j=1}^{n-1}\left(\frac{\Delta t}{2}\left(f_u(t_j)+f_u(t_{j+1})\right) - \int_{t_j}^{t_{j+1}} f_u(t)\,\mathrm{d}t\right)\right\|_2 \\ &\le \frac{1}{2}\sum_{j=1}^{n-1}\int_{t_j}^{t_{j+1}} \big\|\dot f_u(\tau_{j1}(t))\big\|_2 \big|t-t_j\big| + \big\|\dot f_u(\tau_{j2}(t))\big\|_2 \big|t-t_{j+1}\big|\,\mathrm{d}t \\ &\le \frac{1}{2}\max_{t\in[0,T]}\big\|\dot f_u(t)\big\|_2 \sum_{j=1}^{n-1}\left(\frac{(t-t_j)^2}{2} - \frac{(t_{j+1}-t)^2}{2}\bigg|_{t=t_j}^{t=t_{j+1}}\right) \\ &= \frac{\Delta t}{2}\max_{t\in[0,T]}\big\|\dot f_u(t)\big\|_2 \sum_{j=1}^{n-1}\Delta t = \frac{\Delta t\,T}{2}\max_{t\in[0,T]}\big\|\dot f_u(t)\big\|_2 \\ &\le \frac{\Delta t\,T}{2}\max_{t\in[0,T]}\big\|\dot f_u(t)\big\|_2 = \frac{\Delta t\,T}{2}\max_{t\in[0,T]}\big\|\langle \dot y(t), u\rangle_2\, y(t) + \langle y(t), u\rangle_2\, \dot y(t)\big\|_2 \\ &= \Delta t\,T\max_{t\in[0,T]}\|\dot y(t)\|_2\|y(t)\|_2 \le \Delta t\,T\,\|y\|^2_{C^1([0,T];\mathbb{R}^m)}.\end{aligned}$$

Consequently,

$$\|\mathcal{R}^n - \mathcal{R}\|_{\mathscr{L}(\mathbb{R}^m)} = \sup_{\|u\|_2=1}\|\mathcal{R}^n u - \mathcal{R}u\|_2 \le 2\Delta t\,\|y\|^2_{C^1([0,T];\mathbb{R}^m)} \overset{\Delta t\to 0}{\longrightarrow} 0$$

which is (33.19). ☐