# **LECTURE 9: REDUCED SYSTEMS**

### 1. Projection of mappings

#### 1.1. Reduced matrices

The least-squares problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y} - \mathbf{A}\mathbf{x}\| \tag{1}$$

focuses on a simpler representation of a data vector  $y \in \mathbb{R}^m$  as a linear combination of column vectors of  $A \in \mathbb{R}^{m \times n}$ . Consider some phenomenon modeled as a function between vector spaces  $f: X \to Y$ , such that for input parameters  $x \in X$ , the state of the system is y = f(x). For most models f is differentiable, a transcription of the condition that the system should not exhibit jumps in behavior when changing the input parameters. Then by appropriate choice of units and origin, a linearized model

$$y = Ax, A \in \mathbb{R}^{m \times n}$$

is obtained if  $y \in C(A)$ , expressed as (1) if  $y \notin C(A)$ .

A simpler description is often sought, typically based on recognition that the inputs and outputs of the model can themselves be obtained as linear combinations x = Bu, y = Cv, involving a smaller set of parameters  $u \in \mathbb{R}^{q}$ ,  $v \in \mathbb{R}^{p}$ , p < m, q < n. The column spaces of the matrices  $B \in \mathbb{R}^{n \times q}$ ,  $C \in \mathbb{R}^{m \times p}$  are vector subspaces of the original set of inputs and outputs,  $C(B) \leq \mathbb{R}^{n}$ ,  $C(C) \leq \mathbb{R}^{m}$ . The sets of column vectors of B, C each form a *reduced basis* for the system inputs and outputs if they are chosed to be of full rank. The reduced bases are assumed to have been orthonormalized through the Gram-Schmidt procedure such that  $B^{T}B = I_{q}$ , and  $C^{T}C = I_{p}$ . Expressing the model inputs and outputs in terms of the reduced basis leads to

$$C v = A B u \Rightarrow v = C^T A B u \Rightarrow v = R u.$$

The matrix  $\mathbf{R} = \mathbf{C}^T \mathbf{A} \mathbf{B} \in \mathbb{R}^{p \times q}$  is called the *reduced system matrix* and is associated with a mapping  $\mathbf{g}: U \to V$ , that is a restriction to the U, V vector subspaces of the mapping  $\mathbf{f}$ . When  $\mathbf{f}$  is an endomorphism,  $\mathbf{f}: X \to X, m = n$ , the same reduced basis is used for both inputs and outputs,  $\mathbf{x} = \mathbf{B}\mathbf{u}, \mathbf{y} = \mathbf{B}\mathbf{v}$ , and the reduced system is

$$\boldsymbol{v} = \boldsymbol{R} \boldsymbol{u}, \boldsymbol{R} = \boldsymbol{B}^T \boldsymbol{A} \boldsymbol{B}$$

Since **B** is assumed to be orthogonal, the projector onto C(B) is  $P_B = BB^T$ . Applying the projector on the initial model

$$P_B y = P_B A x$$

leads to  $BB^T y = BB^T Ax$ , and since  $v = B^T y$  the relation  $Bv = BB^T ABu$  is obtained, and conveniently grouped as

$$\boldsymbol{B}\boldsymbol{v} = \boldsymbol{B}\left(\boldsymbol{B}^T \boldsymbol{A} \boldsymbol{B}\right) \boldsymbol{u} \Rightarrow \boldsymbol{B}\boldsymbol{v} = \boldsymbol{B}(\boldsymbol{R} \boldsymbol{u}),$$

again leading to the reduced model v = Bu. The above calculation highlights that the reduced model is a projection of the full model y = Ax on C(B).

#### 1.2. Dynamical system model reduction

An often encountered situation is the reduction of large-dimensional dynamical system

$$M \ddot{x} + D \dot{x} + Kx = f, M, D, K \in \mathbb{R}^{m \times m}, x, f \colon \mathbb{R}_{+} \to \mathbb{R}^{m},$$

$$\dot{x} = \frac{\mathrm{d}x}{\mathrm{d}t}, \ddot{x} = \frac{\mathrm{d}\dot{x}}{\mathrm{d}t},$$
(2)

a generalization to multiple degrees of freedom of the damped oscillator equation

$$m\ddot{x} + d\dot{x} + kx = f$$

In (2),  $\mathbf{x}(t)$  are the time-dependent coordinates of the system, f(t) the forces acting on the system, and M, D, K are the mass, drag, stiffness matrices, respectively.

When  $m \gg 1$ , a reduced description is sought by linear combination of  $n \ll m$  basis vectors

$$x \cong \tilde{x} = B y \Rightarrow M B \ddot{y} + D B \dot{y} + K B y = f$$

Choose  $B \in \mathbb{R}^{m \times n}$  to have orthonormal columns, and project (2) onto C(B) by multiplication with the projector  $P = BB^T$ 

$$BB^{T}MB \ddot{y} + BB^{T}DB \dot{y} + BB^{T}KB y = BB^{T}f \Rightarrow$$

 $\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{M}\boldsymbol{B}\,\ddot{\boldsymbol{y}} + \boldsymbol{B}^T\boldsymbol{D}\boldsymbol{B}\,\dot{\boldsymbol{y}} + \boldsymbol{B}^T\boldsymbol{K}\boldsymbol{B}\,\boldsymbol{y} - \boldsymbol{B}^T\,\boldsymbol{f}) = \boldsymbol{0} \Leftrightarrow \boldsymbol{B}\boldsymbol{z} = \boldsymbol{0}.$ 

Since  $N(\boldsymbol{B}) = \{\boldsymbol{0}\}$ , deduce  $\boldsymbol{z} = \boldsymbol{0}$ , hence

$$\boldsymbol{B}^{T}\boldsymbol{M}\boldsymbol{B}\,\boldsymbol{y} + \boldsymbol{B}^{T}\boldsymbol{D}\boldsymbol{B}\,\boldsymbol{y} + \boldsymbol{B}^{T}\boldsymbol{K}\boldsymbol{B}\,\boldsymbol{y} = \boldsymbol{B}^{T}\,\boldsymbol{f}.$$

Introduce notations

$$\tilde{M} = B^T M B, \tilde{D} = B^T D B, \tilde{K} = B^T K B$$

for the reduced mass, drag, stiffness matrices, with  $\tilde{M}, \tilde{D}, \tilde{K} \in \mathbb{R}^{n \times n}$  of smaller size. The reduced coordinates and forces are

$$\tilde{f} = \boldsymbol{B}^T \boldsymbol{f}, \ \boldsymbol{y}, \tilde{f} \in \mathbb{R}^n$$

The resulting reduced dynamical system is

$$\tilde{M}$$
  $\ddot{y}$  +  $\tilde{D}$   $y$  +  $\tilde{K}$   $y$  =  $f$ .

## 2. Reduced bases

One elemenet is missing from the description of model reduction above: how is **B** determined? Domain-specific knowledge can often dictate an appropriate basis (e.g., Fourier basis fo periodic phenomena). An alternative approach is to extract an appropriate basis from observations of a phenomenon, known as *data-driven modeling*.

### 2.1. Correlation matrices

**Correlation coefficient.** Consider two functions  $x_1, x_2: \mathbb{R} \to \mathbb{R}$ , that represent data streams in time of inputs  $x_1(t)$  and outputs  $x_2(t)$  of some system. A basic question arising in modeling and data science is whether the inputs and outputs are themselves in a functional relationship. This usually is a consequence of incomplete knowledge of the system, such that while  $x_1, x_2$  might be assumed to be the most relevant input, output quantities, this is not yet fully established. A typical approach is to then carry out repeated measurements leading to a data set  $D = \{(x_1(t_i), x_2(t_i)) | i = 1, ..., N\}$ , thus defining a relation. Let  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$  denote vectors containing the input and output values. The *mean values*  $\mu_1, \mu_2$  of the input and output are estimated by the statistics

$$\mu_1 \cong \bar{x}_1 = \frac{1}{N} \sum_{i=1}^N x_1(t_i) = E[x_1], \ \mu_2 \cong \bar{x}_2 = \frac{1}{N} \sum_{i=1}^N x_2(t_i) = E[x_2],$$

where E is the expectation seen to be a linear mapping,  $E: \mathbb{R}^N \to \mathbb{R}$  whose associated matrix is

$$E = \frac{1}{N} [1 \ 1 \ \dots \ 1],$$

and the means are also obtained by matrix vector multiplication (linear combination),

$$\bar{x}_1 = \boldsymbol{E} \, \boldsymbol{x}_1, \, \bar{x}_2 = \boldsymbol{E} \, \boldsymbol{x}_2.$$

Deviation from the mean is measured by the *standard deviation* defined for  $x_1, x_2$  by

$$\sigma_1 = \sqrt{E[(x_1 - \mu_1)^2]}, \sigma_2 = \sqrt{E[(x_2 - \mu_2)^2]}.$$

Note that the standard deviations are no longer linear mappings of the data.

Assume that the origin is chosen such that  $\bar{x}_1 = \bar{x}_2 = 0$ . One tool to estalish whether the relation *D* is also a function is to compute the *correlation coefficient* 

$$\rho(x_1, x_2) = \frac{E[x_1 x_2]}{\sigma_1 \sigma_2} = \frac{E[x_1 x_2]}{\sqrt{E[x_1^2] E[x_2^2]}}$$

that can be expressed in terms of a scalar product and 2-norm as

$$\rho(x_1, x_2) = \frac{\boldsymbol{x}_1^T \boldsymbol{x}_2}{\|\boldsymbol{x}_1\| \|\boldsymbol{x}_2\|}.$$

Squaring each side of the norm property  $||x_1 + x_2|| \le ||x_1|| + ||x_2||$ , leads to

$$(\boldsymbol{x}_{1} + \boldsymbol{x}_{2})^{T} (\boldsymbol{x}_{1} + \boldsymbol{x}_{2}) \leq \boldsymbol{x}_{1}^{T} \boldsymbol{x}_{1} + \boldsymbol{x}_{2}^{T} \boldsymbol{x}_{2} + 2 \| \boldsymbol{x}_{1} \| \| \boldsymbol{x}_{2} \| \Rightarrow \boldsymbol{x}_{1}^{T} \boldsymbol{x}_{2} \leq \| \boldsymbol{x}_{1} \| \| \boldsymbol{x}_{2} \|$$

known as the Cauchy-Schwarz inequality, which implies  $-1 \le \rho(x_1, x_2) \le 1$ . Depending on the value of  $\rho$ , the variables  $x_1(t), x_2(t)$  are said to be:

- 1. *uncorrelated*, if  $\rho = 0$ ;
- 2. *correlated*, if  $\rho = 1$ ;
- 3. *anti-correlated*, if  $\rho = -1$ .

The numerator of the correlation coefficient is known as the covariance of  $x_1, x_2$ 

$$\operatorname{cov}(x_1, x_2) = E[x_1 x_2].$$

The correlation coefficient can be interpreted as a normalization of the covariance, and the relation

$$\operatorname{cov}(x_1, x_2) = \mathbf{x}_1^T \mathbf{x}_2 = \rho(x_1, x_2) \|\mathbf{x}_1\| \|\mathbf{x}_2\|,$$

is the two-variable version of a more general relationship encountered when the system inputs and outputs become vectors.

**Patterns in data.** Consider now a related problem, whether the input and output parameters  $x \in \mathbb{R}^n$ ,  $y \in \mathbb{R}^m$  thought to characterize a system are actually well chosen, or whether they are redundant in the sense that a more insightful description is furnished by  $u \in \mathbb{R}^q$ ,  $v \in \mathbb{R}^p$  with fewer components p < m, q < n. Applying the same ideas as in the correlation coefficient, a sequence of N measurements is made leading to data sets

$$\boldsymbol{X} = [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \dots \ \boldsymbol{x}_n] \in \mathbb{R}^{N \times n}, \boldsymbol{Y} = [\boldsymbol{y}_1 \ \boldsymbol{y}_2 \ \dots \ \boldsymbol{y}_n] \in \mathbb{R}^{N \times m}$$

Again, by appropriate choice of the origin the means of the above measurements is assumed to be zero

$$E[\mathbf{x}] = \mathbf{0}, E[\mathbf{y}] = \mathbf{0}.$$

Covariance matrices can be constructed by

$$\boldsymbol{C}_{\boldsymbol{X}} = \boldsymbol{X}^{T} \boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_{1}^{T} \\ \boldsymbol{x}_{2}^{T} \\ \vdots \\ \boldsymbol{x}_{n}^{T} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{1} \ \boldsymbol{x}_{2} \ \cdots \ \boldsymbol{x}_{n} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_{1}^{T} \boldsymbol{x}_{1} \ \boldsymbol{x}_{1}^{T} \boldsymbol{x}_{2} \ \cdots \ \boldsymbol{x}_{1}^{T} \boldsymbol{x}_{n} \\ \boldsymbol{x}_{2}^{T} \boldsymbol{x}_{1} \ \boldsymbol{x}_{2}^{T} \boldsymbol{x}_{2} \ \cdots \ \boldsymbol{x}_{2}^{T} \boldsymbol{x}_{n} \\ \vdots \ \vdots \ \ddots \ \vdots \\ \boldsymbol{x}_{n}^{T} \boldsymbol{x}_{1} \ \boldsymbol{x}_{n}^{T} \boldsymbol{x}_{2} \ \cdots \ \boldsymbol{x}_{n}^{T} \boldsymbol{x}_{n} \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Consider now the SVDs of  $C_X = N\Lambda N^T$ ,  $X = U\Sigma S^T$ , and from

$$\boldsymbol{C}_{\boldsymbol{X}} = \boldsymbol{X}^T \boldsymbol{X} = (\boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{S}^T)^T \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{S}^T = \boldsymbol{S} \boldsymbol{\Sigma}^T \boldsymbol{U}^T \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{S}^T = \boldsymbol{S} \boldsymbol{\Sigma}^T \boldsymbol{\Sigma} \boldsymbol{S}^T = \boldsymbol{N} \boldsymbol{\Lambda} \boldsymbol{N}^T,$$

identify N = S, and  $\Lambda = \Sigma^T \Sigma$ .

Recall that the SVD returns an order set of singular values  $\sigma_1 \ge \sigma_2 \ge \cdots \ge$ , and associated singular vectors. In many applications the singular values decrease quickly, often exponentially fast. Taking the first *q* singular modes then gives a basis set suitable for mode reduction

$$\boldsymbol{x} = \boldsymbol{S}_q \, \boldsymbol{u} = \left[ \begin{array}{ccc} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \dots & \boldsymbol{s}_q \end{array} \right] \boldsymbol{u}.$$

# 3. Stochastic systems - Karhunen-Loève theorem

The data reduction inherent in SVD representations is a generic feature of natural phenomena. A paradigm for physical systems is the evolution of correlated behavior against a backdrop of thermal enery, typically represented as a form of noise.

One mathematical technique to model such systems is the definition of a stochastic process  $\{X_t\}_{a \le t \le b}$ , where for each fixed  $t, X_t$  is a random variable, i.e., a measurable function  $X: \Omega \to E$  from a set of possible outcomes  $\Omega$  to a measurable space E. The set  $\Omega$  is the sample space of a probability triple  $(\Omega, \mathcal{F}, P)$ , where for  $\forall S \subseteq E$ 

$$P(X \in S) = P(\{\omega \in \Omega\} | X(\omega) \in S|).$$

A measurable space is a set coupled with procedure to determine measurable subsets, known as a  $\sigma$ -algebra.

THEOREM. Let  $X_t$  be a zero-mean ( $\mathbb{E}[X_t] = 0$ ), square-integrable stochastic process defined over probability space  $(\Omega, \mathcal{F}, P)$  indexed by  $t \in \mathbb{R}$ ,  $a \leq t \leq b$ . Then  $X_t$  admits a representation

$$X_{t} = \sum_{k=1}^{\infty} Z_{k} e_{k}(t),$$
$$Z_{k} = \int_{a}^{b} X_{t} e_{k}(t) dt, \mathbb{E}[Z_{k}] = 0, \mathbb{E}[Z_{i}, Z_{j}] = \delta_{ij} \sigma_{j}.$$

with