Optimal low-dimensional approximations of random vector functions *

J. vom Scheidt H.-J. Starkloff R. Wunderlich

Abstract

The paper considers approximations of first- and second-order moments of random functions with values in a high-dimensional Euclidean space using projections onto suitable low-dimensional linear submanifolds. To quantify the goodness of the approximation a criterion based on the mean squared Euclidean distance is introduced. In case of wide-sense stationary random functions optimal low-dimensional linear submanifolds are given in terms of the mean vector and eigenvectors of the variance matrix.

1 Introduction

Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space, where $\mathcal{A}$ denotes the σ-algebra of subsets of Ω on which is defined a probability measure $\mathbf{P}$ and $\mathcal{D} \subset \mathbb{R}$ be an interval. Further, let $\mathbf{z}(t, \omega)$ be a random function defined on $\mathcal{D} \times \Omega$ with values in a n-dimensional Euclidean space $\mathbf{Z} = \mathbb{C}^n$, $n \geq 1$, where $\mathbb{C}$ denotes the space of complex numbers. It is assumed, that there exist moments of $\mathbf{z}(t, \omega)$ up to second order. The mean of $\mathbf{z}(t, \omega)$ is denoted by

$$\mathbf{m}_{\mathbf{z}}(t) = \mathbf{E}\left\{\mathbf{z}(t)\right\}$$

and the central second-order moment, i.e. the correlation function by

$$\mathbf{K}_{\mathbf{zz}}(t, s) = \mathbf{E}\left\{(\mathbf{z}(t) - \mathbf{m}_{\mathbf{z}}(t))(\mathbf{z}(s) - \mathbf{m}_{\mathbf{z}}(s))^*\right\}$$

where $\mathbf{E}\left\{.\right\}$ denotes expectation with respect to $\mathbf{P}$ and $(.)^*$ the adjoint matrix.

The paper considers approximations of the first- and second-order moments of $\mathbf{z}$ by the corresponding moments of a projection of $\mathbf{z}$ onto a linear submanifold of $\mathbf{Z}$. This problem arises in case of a high-dimensional space $\mathbf{Z}$, where the computation of the moments of $\mathbf{z}$ can be very difficult or impossible because of the high numerical effort (see [4, 5]).

*This work was supported by Deutsche Forschungsgemeinschaft

If the dynamics and the randomness of $\mathbf{z}(t,\omega)$ can be expressed - up to a certain degree of accuracy - by its projection onto a low-dimensional linear submanifold practicable approximations of the moments can be found. This approximation idea can be applied to the computation of moments of the response of discrete vibration systems with a random excitation (see [4]).

Section 2 provides some notations to investigate projections of $\mathbf{z}(t,\omega)$ onto a linear submanifold of $\mathbf{Z}$. To quantify the "degree of accuracy" of the approximation of first- and second-order moments of $\mathbf{z}(t,\omega)$ by the corresponding moments of the projection of $\mathbf{z}$ onto a submanifold we introduce some approximation criterions in Section 3. First, the pointwise squared Euclidean distance between $\mathbf{z}(t,\omega)$ and its projection is used. By averaging with respect to t and ω a non-negative quantity – the approximation risk – is obtained. The approximation risk can be related with the mean and the trace of the variance matrix of $\mathbf{z}$ in the case of orthogonal projections. Further, bounds for the error of the approximation of the characteristics $\mathbf{m}_{\mathbf{z}}(t)$ and $\mathbf{K}_{\mathbf{zz}}(t,s)$ in terms of the approximation risk can be given.

In order to find optimal low-dimensional linear submanifolds Section 4 considers the minimization of the approximation risk with respect to the defining elements of the submanifold. As the main result we find, that in case of a wide-sense stationary function $\mathbf{z}$ the eigenvectors of the variance matrix $\mathbf{K}_{\mathbf{zz}}(0)$ define optimal submanifolds and the mean $\mathbf{m}_{\mathbf{z}}$ leads to the corresponding translation from the origin.

Applying these optimal projections to the approximation of moments of $\mathbf{z}$ the mean and the variance matrix of $\mathbf{z}$ have to be computed. This partial information on the characteristics can be computed much easier and more effective than the complete characteristics.

Similar problems of approximation of (not necessarily vector valued) random functions by projections onto finite dimensional (function) spaces are studied e.g. in [1] and [3].

2 Projections

In this section we consider an affin linear transformation of the coordinate system of $\mathbf{Z}$ and projections onto m-dimensional linear submanifolds of $\mathbf{Z}$, $1 \leq m \leq n$. Afterwards these transformations and projections will be applied to the function $\mathbf{z}(t,\omega)$ pointwise for all t and ω to derive an approximation of $\mathbf{z}$. For abbrevation the dependence of $\mathbf{z}$ on t and ω is omitted and we write $\mathbf{z}$ instead $\mathbf{z}(t,\omega)$.

Let $(\mathbf{e}_k)_{1\leq k\leq n}$ be the canonical basis of $\mathbf{Z}$, i.e. $\mathbf{z} = (z_1,\ldots,z_n)^\tau \in \mathbf{Z}$ can be represented by

$$\mathbf{z} = \sum_{k=1}^{n} z_k \mathbf{e}_k$$

with coordinates $z_k \in \mathbb{C}$, $k = 1,\ldots,n$, with respect to $(\mathbf{e}_k)_{1\leq k\leq n}$. Denote the scalar product and the norm of $\mathbf{Z}$ by

$$(\mathbf{x},\mathbf{y}) = \sum_{k=1}^{n} x_k \overline{y_k} \qquad \text{and} \qquad \|\mathbf{x}\| = (\mathbf{x},\mathbf{x})^{\frac{1}{2}} = \left(\sum_{k=1}^{n} |x_k|^2\right)^{\frac{1}{2}},$$

for $\mathbf{x},\mathbf{y} \in \mathbf{Z}$.

Let $(\mathbf{v}_k)_{1\le k\le n}$ be a set of linear independent elements of $\mathbf{Z}$ with $\|\mathbf{v}_k\| = 1$ and $\mathbf{v}_0 \in \mathbf{Z}$. Denote by $x_1,\ldots,x_n \in \mathbb{C}$ the new coordinates of $\mathbf{z} \in \mathbf{Z}$ with respect to the basis $(\mathbf{v}_k)_{1\le k\le n}$ and the translation element $\mathbf{v}_0$ (see Figure 1), i.e.

$$\mathbf{z} = \sum_{k=1}^{n} x_k \mathbf{v}_k + \mathbf{v}_0. \tag{1}$$

Defining the vector $\mathbf{x} = (x_1,\ldots,x_n)^\tau$ and the non-singular matrix $\mathbf{V} = (\mathbf{v}_1,\ldots,\mathbf{v}_n)$ ($\mathbf{v}_1,\ldots,\mathbf{v}_n$ are linear independent) the transformation (1) and its inverse transformation can be written as

$$\mathbf{z} = \mathbf{V}\mathbf{x} + \mathbf{v}_0 \qquad \text{and} \qquad \mathbf{x} = \mathbf{V}^{-1}(\mathbf{z} - \mathbf{v}_0).$$

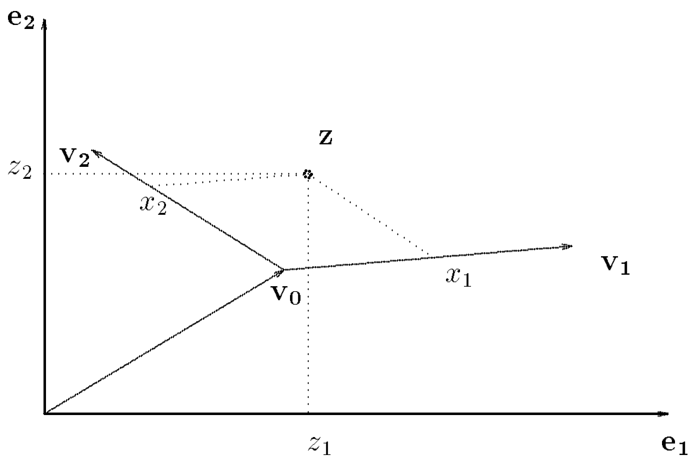

Figure 1: Transformation of coordinate system for $n = 2$

Now we consider projections of elements $\mathbf{z} \in \mathbf{Z}$ onto m-dimensional linear submanifolds $\mathbf{S}_m$ defined by

$$\begin{aligned} \mathbf{S}_m &:= \mathbf{v}_0 + \operatorname{span}(\mathbf{v}_1,\ldots,\mathbf{v}_m), \quad 1 \le m \le n, \\ \text{and} \quad \mathbf{S}_0 &:= \mathbf{v}_0. \end{aligned}$$

Let $(\mathcal{P}_m)_{0\le m\le n}$ be the sequence of projectors $\mathcal{P}_m : \mathbf{Z} \to \mathbf{S}_m$ defined by

$$\begin{aligned} \mathcal{P}_m\mathbf{z} &:= \mathbf{v}_0 + \sum_{k=1}^{m} x_k\mathbf{v}_k, \quad 1 \le m \le n, \\ \text{and} \quad \mathcal{P}_0\mathbf{z} &:= \mathbf{v}_0. \end{aligned} \tag{2}$$

Defining real and symmetric matrices $\mathbf{P}_m$ by

$$\begin{aligned} \mathbf{P}_m &:= \begin{pmatrix} \mathbf{I}_m & \mathbf{O}_{m,n-m} \\ \mathbf{O}_{n-m,m} & \mathbf{O}_{n-m,n-m} \end{pmatrix} \quad 1 \le m \le n, \\ \text{and} \quad \mathbf{P}_0 &:= \mathbf{O}_{n,n}, \end{aligned}$$

where $\mathbf{I}_m$ denotes the $m \times m$ identity matrix and $\mathbf{O}_{..}$ denote null matrices of the corresponding dimensions. Then relation (2) can be written as

$$\begin{aligned} \mathcal{P}_m\mathbf{z} &= \mathbf{v}_0 + \mathbf{V}\mathbf{P}_m\mathbf{x} \\ &= \mathbf{v}_0 + \mathbf{V}\mathbf{P}_m\mathbf{V}^{-1}(\mathbf{z} - \mathbf{v}_0). \end{aligned} \tag{3}$$

For the sake of shorter notation complementary projection operators $\mathcal{Q}_m$ are introduced by

$$\begin{aligned} \mathcal{Q}_m\mathbf{z} &:= (\mathcal{I} - \mathcal{P}_m)\mathbf{z} = \mathbf{z} - \mathcal{P}_m\mathbf{z} \\ &= \mathbf{V}\mathbf{Q}_m\mathbf{x} = \mathbf{V}\mathbf{Q}_m\mathbf{V}^{-1}(\mathbf{z} - \mathbf{v}_0), \end{aligned} \tag{4}$$

where the real and symmetric matrix $\mathbf{Q}_m$ is defined by $\mathbf{Q}_m := \mathbf{I}_n - \mathbf{P}_m$, i.e.

$$\begin{aligned} \mathbf{Q}_m &= \begin{pmatrix} \mathbf{O}_{m,m} & \mathbf{O}_{m,n-m} \\ \mathbf{O}_{n-m,m} & \mathbf{I}_{n-m} \end{pmatrix} \quad 1 \le m \le n, \\ \text{and} \quad \mathbf{Q}_0 &= \mathbf{I}_n, \end{aligned} \tag{5}$$

and $\mathcal{I}$ denotes the identity operator $\mathcal{I} : \mathbf{Z} \to \mathbf{Z}$ with $\mathcal{I}\mathbf{z} = \mathbf{z}$.

3 Approximation criterions

In this section the projection operator $\mathcal{P}_m$ is applied to $\mathbf{z}(t, \omega)$ pointwise for all t and ω and $\mathcal{P}_m\mathbf{z}$ is used as an approximation of $\mathbf{z}$. In order to quantify and to assess the accuracy of this approximation several criterions are studied. We start with the definition of a non-negative loss function, which measures pointwise the distance between $\mathbf{z}(t, \omega)$ and its approximation $\mathcal{P}_m\mathbf{z}(t, \omega)$. Averaging with respect to t and ω leads to a non-negative quantity – the approximation risk – which characterizes the "goodness" of the projector $\mathcal{P}_m$. If the squared Euclidean distance is used as loss function it will be proved that orthogonal projections are optimal with respect to the approximation risk.

Finally, we deal with the approximation of first- and second-order moments of $\mathbf{z}$ by the corresponding moments of $\mathcal{P}_m\mathbf{z}$ and estimate the accuracy of these approximations in terms of the approximation risk.

Definition 1
A $(\mathcal{B}(\mathbf{Z}) \otimes \mathcal{B}(\mathbf{Z}), \mathcal{B}(\mathbb{R}_+))$-measurable non-negative function $l(.,.) : \mathbf{Z} \times \mathbf{Z} \to \mathbb{R}_+$ is called loss function *and*

$$\beta = \beta(\mathcal{P}_m, t, \omega) := l(\mathbf{z}(t,\omega), \mathcal{P}_m\mathbf{z}(t,\omega))$$

is called approximation error *of the approximation $\mathcal{P}_m\mathbf{z}$ of $\mathbf{z}$ (with respect to $l(.,.)$ and $\mathcal{P}_m$). Here $\mathcal{B}(.)$ denotes the corresponding σ-algebra of Borel-sets.*

The approximation error $\beta = \beta(\mathcal{P}_m, t, \omega)$ with $\beta(\mathcal{P}_m, ., .) : \mathcal{D} \times \Omega \to \mathbb{R}_+$, is a non-negative scalar random function. By averaging with respect to t and ω a non-random and non-time-dependent quantity is obtained which will be used as objective in the the optimization afterwards.

Definition 2
Let $\psi(.) : \mathbb{R} \to \mathbb{R}_+$ be an integrable non-negative weighting function with $\int\limits_{\mathcal{D}} \psi(t)dt = 1$. Then, the quantity

$$\gamma = \gamma(\mathcal{P}_m) := \int\limits_{\mathcal{D}} \mathbf{E}\left\{\beta(\mathcal{P}_m, t)\right\} \psi(t)dt$$

is called approximation risk *of $\mathcal{P}_m$ with respect to the loss function l and the weighting function ψ.*

Denote by $\mathcal{M}$ the set of all sequences $(\mathcal{P}_m)_{0\le m\le n}$ of projection operators defined in (2). Each sequence can be identified by n linear independent basis elements $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbf{Z}$, $\|\mathbf{v}_k\| = 1, k = 1, \ldots, n$, and a translation element $\mathbf{v}_0 \in \mathbf{Z}$.

Definition 3
A sequence $(\mathcal{P}^*_m)_{0\le m\le n} \in \mathcal{M}$ is said to be γ-optimal *with respect to the set $\mathcal{M}$ if*

$$\gamma(\mathcal{P}^*_m) \le \gamma(\mathcal{P}_m), \qquad m = 0, \ldots, n,$$

holds for all sequences $(\mathcal{P}_m)_{0\le m\le n} \in \mathcal{M}$.

Now some possible choices of the loss function $l(.,.)$ are listed.

i) $l_1(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|^2$ with $\beta(\mathcal{P}_m, t, \omega) = \|\mathbf{z}(t,\omega) - \mathcal{P}_m\mathbf{z}(t,\omega)\|^2 = \|\mathcal{Q}_m\mathbf{z}(t,\omega)\|^2$

This function is an appropriate choice if the approximation error can be controlled by the squared Euclidean distance between $\mathbf{z}(t,\omega)$ and $\mathcal{P}_m\mathbf{z}(t,\omega)$.

ii) $l_2(\mathbf{x},\mathbf{y}) := \|\mathbf{W}(\mathbf{x}-\mathbf{y})\|^2$ with $\beta(\mathcal{P}_m,t,\omega) = \|\mathbf{W}\mathcal{Q}_m\mathbf{z}(t,\omega)\|^2$

where $\mathbf{W}$ denotes a $n\times n$ diagonal matrix with diagonal elements $w_k \geq 0$, $k=1,\ldots,n$. The goodness of approximation of several components of $\mathbf{z}$ can be weighted by this function. Less important components z_k can be neglected by choosing of weights $w_k = 0$.

iii) $l_3(\mathbf{x},\mathbf{y}) := \|\mathbf{C}(\mathbf{x}-\mathbf{y})\|^2_{\mathbf{Z}'}$ with $\beta(\mathcal{P}_m,t,\omega) = \|\mathbf{C}\mathcal{Q}_m\mathbf{z}(t,\omega)\|^2_{\mathbf{Z}'}$

where $\mathbf{C}$ denotes a rectangular $n'\times n$-matrix with $1\leq n'\leq n$, and $\mathbf{Z}'$ is a n'-dimensional Euclidean space with the norm $\|.\|^2_{\mathbf{Z}'}$. This function can be applied if an output equation

$$\mathbf{y} = \mathbf{C}\mathbf{z}$$

is given and the goodness of the approximation $\tilde{\mathbf{y}} = \mathbf{C}\mathcal{P}_m\mathbf{z}$ of the "output" has to be assessed .

Further loss functions and the corresponding approximation errors and risks can be found in [3] p. 93 ff.

In mechanical applications the random dynamics of multibody systems is investigated using a vector function $\mathbf{z}(t,\omega)$ which contains positions and velocities of the bodies of the system. In this case not only the approximation goodness with respect to the moments of $\mathbf{z}(t,\omega)$ but also of its derivative $\dot{\mathbf{z}}(t,\omega)$ (which contains the accelerations) has to be assessed. This can be done with a modified approximation error and risk

$$\begin{aligned}\breve{\beta}(\mathcal{P}_m,t,\omega) &:= l(\dot{\mathbf{z}}(t,\omega),\mathcal{P}_m\dot{\mathbf{z}}(t,\omega))\\ \breve{\gamma}(\mathcal{P}_m) &:= \int_{\mathcal{D}} \mathbf{E}\left\{\breve{\beta}(\mathcal{P}_m,t,\omega)\right\}\psi(t)dt.\end{aligned}$$

For this paper the loss function $l = l_1$ with the approximation error

$$\beta(\mathcal{P}_m,t,\omega) = \|\mathcal{Q}_m\mathbf{z}(t,\omega)\|^2$$

is considered, only. Then the corresponding approximation risk can be written as

$$\gamma(\mathcal{P}_m) \;=\; \int_{\mathcal{D}} \mathbf{E}\left\{\|\mathcal{Q}_m\mathbf{z}(t)\|^2\right\}\psi(t)dt. \tag{6}$$

Now we deal with the case of orthogonal projectors. Denote by $\mathcal{N}$ the set of all sequences $(\mathcal{P}_m)_{0\leq m\leq n}$ of projection operators with respect to a translation element $\mathbf{v}_0\in\mathbf{Z}$ and orthonormal elements $\mathbf{v}_1,\ldots,\mathbf{v}_n\in\mathbf{Z}$, i.e. $(\mathbf{v}_k,\mathbf{v}_l)=\delta_{kl}$, for $1\leq k,l\leq n$. Then $\mathcal{N}$ is a subset of $\mathcal{M}$. Subsequently, we consider γ-optimal sequences $(\mathcal{P}^*_m)_{0\leq m\leq n}$ with repect to $\mathcal{N}$ and compare these sequences with the corresponding γ-optimal sequences for the larger set $\mathcal{M}$.

Lemma 1
If a sequence $(\mathcal{P}^*_m)_{0\le m\le n} \in \mathcal{N}$ is γ-optimal with respect to the set $\mathcal{N} \subset \mathcal{M}$ then it is also γ-optimal with respect to the set $\mathcal{M}$.

Proof: If $(\mathcal{P}^*_m)_{0\le m\le n} \in \mathcal{N}$ is γ-optimal with respect to the set $\mathcal{N}$ then

$$\gamma(\mathcal{P}^*_m) \le \gamma(\mathcal{P}_m), \quad m = 0, \dots, n$$

holds $\forall\, (\mathcal{P}_m)_{0\le m\le n} \in \mathcal{N} \subset \mathcal{M}$. We have to prove that the relations

$$\gamma(\mathcal{P}^*_m) \le \gamma(\mathcal{P}_m), \quad m = 0, \dots, n \tag{7}$$

are fulfilled $\forall\, (\mathcal{P}_m)_{0\le m\le n} \in \mathcal{M} \setminus \mathcal{N}$.

Now it is assumed that relations (7) are not fulfilled, i.e. there exists a sequence $(\mathcal{R}_m)_{0\le m\le n} \in \mathcal{M} \setminus \mathcal{N}$ and an index r, $1 \le r \le n$, with

$$\gamma(\mathcal{P}^*_r) > \gamma(\mathcal{R}_r). \tag{8}$$

The sequence $(\mathcal{R}_m)_{0\le m\le n}$ is assumed to be defined by the translation element $\mathbf{v}_0$ and the linear independent elements $\mathbf{v}_1, \dots, \mathbf{v}_n$. A set of pairwise orthonormal elements $\mathbf{w}_1, \dots, \mathbf{w}_n$ with respect to $\mathbf{v}_1, \dots, \mathbf{v}_n$ can be found (e.g. by means of Schmidt-orthogonalization) which fulfill the conditions

$$\begin{aligned} \mathrm{span}(\mathbf{v}_1, \dots, \mathbf{v}_m) &= \mathrm{span}(\mathbf{w}_1, \dots, \mathbf{w}_m), \qquad m = 1, \dots, n, \\ (\mathbf{w}_k, \mathbf{w}_l) &= \delta_{kl}, \quad 1 \le k, l \le n, \\ (\mathbf{w}_k, \mathbf{v}_l) &= 0, \quad 1 \le l < k \le n. \end{aligned} \tag{9}$$

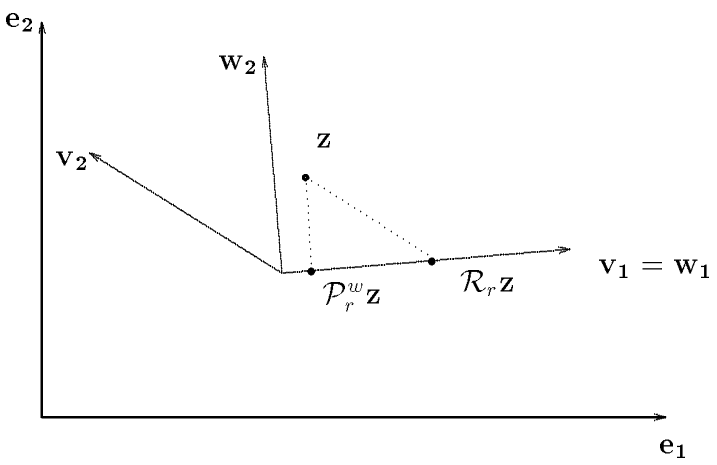

Figure 2: Orthogonal and non-orthogonal projections of $\mathbf{z}$ for $n = 2,\ r = 1$

Denote by $(\mathcal{P}_m^w)_{0\le m\le n}$ the sequence of projectors defined by $\mathbf{v}_0, \mathbf{w}_1, \dots, \mathbf{w}_n$, where $(\mathcal{P}_m^w)_{0\le m\le n} \in \mathcal{N}$. Now we consider the case $m = r$ and the identity

$$\mathbf{z} - \mathcal{R}_r\mathbf{z} = (\mathbf{z} - \mathcal{P}_r^w\mathbf{z}) + (\mathcal{P}_r^w\mathbf{z} - \mathcal{R}_r\mathbf{z}), \qquad \forall\, t \in [0,T] \text{ and } \forall\, \omega \in \Omega.$$

Let the coordinates of $\mathbf{z}$ with respect to the orthonormal basis $(\mathbf{w}_k)_{0\le k\le n}$ and the translation element $\mathbf{v}_0$ be denoted by $y_1, \dots, y_n$. By means of notations in (1) and (2) it follows

$$\begin{aligned}
\mathbf{z} &= \mathbf{v}_0 + \sum_{k=1}^{n} x_k\mathbf{v}_k = \mathbf{v}_0 + \sum_{k=1}^{n} y_k\mathbf{w}_k,\\
\mathcal{R}_r\mathbf{z} &= \mathbf{v}_0 + \sum_{k=1}^{r} x_k\mathbf{v}_k,\\
\mathcal{P}_r^w\mathbf{z} &= \mathbf{v}_0 + \sum_{k=1}^{r} y_k\mathbf{w}_k.
\end{aligned}$$

Applying relations (9) we derive

$$\begin{aligned}
(\mathbf{z} - \mathcal{P}_r^w\mathbf{z}, \mathcal{P}_r^w\mathbf{z} - \mathcal{R}_r\mathbf{z}) &= \left(\sum_{k=r+1}^{n} y_k\mathbf{w}_k, \sum_{l=1}^{r}(y_l\mathbf{w}_l - x_l\mathbf{v}_l)\right)\\
&= \sum_{k=r+1}^{n}\sum_{l=1}^{r}\left(y_k y_l\,(\mathbf{w}_k,\mathbf{w}_l) - y_k x_l\,(\mathbf{w}_k,\mathbf{v}_l)\right)\\
&= 0.
\end{aligned}$$

Hence, $\mathbf{z} - \mathcal{P}_r^w\mathbf{z}$ and $\mathcal{P}_r^w\mathbf{z} - \mathcal{R}_r\mathbf{z}$ are orthogonal and it follows

$$\|\mathbf{z} - \mathcal{R}_r\mathbf{z}\|^2 = \|\mathbf{z} - \mathcal{P}_r^w\mathbf{z}\|^2 + \|\mathcal{P}_r^w\mathbf{z} - \mathcal{R}_r\mathbf{z}\|^2 \ge \|\mathbf{z} - \mathcal{P}_r^w\mathbf{z}\|^2.$$

Since

$$\begin{aligned}
\beta(\mathcal{R}_r, t, \omega) &= \|\mathbf{z}(t,\omega) - \mathcal{R}_r\mathbf{z}(t,\omega)\|^2\\
\beta(\mathcal{P}_r^w, t, \omega) &= \|\mathbf{z}(t,\omega) - \mathcal{P}_r^w\mathbf{z}(t,\omega)\|^2
\end{aligned}$$

the relation $\beta(\mathcal{R}_r, t, \omega) \ge \beta(\mathcal{P}_r^w, t, \omega)$ pointwise $\forall\, t \in [0,T]$ and $\forall\, \omega \in \Omega$ is obvious and it holds

$$\gamma(\mathcal{R}_r) = \int_{\mathcal{D}} \mathbf{E}\left\{\beta(\mathcal{R}_r, t)\right\}\psi(t)dt \ge \int_{\mathcal{D}} \mathbf{E}\left\{\beta(\mathcal{P}_r^w, t)\right\}\psi(t)dt = \gamma(\mathcal{P}_r^w).$$

Since $(\mathcal{P}_m^w)_{0\le m\le n} \in \mathcal{N}$ we have $\gamma(\mathcal{R}_r) \ge \gamma(\mathcal{P}_r^w) \ge \gamma(\mathcal{P}_r^*)$ in contradiction to the assumption (8) and the proof is complete. ∎

By means of Lemma 1 the search of γ-optimal projections can be restricted to the case of orthonormal elements $\mathbf{v}_1, \dots, \mathbf{v}_n$. In this case the transformation matrix $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ is an unitary matrix, i.e. $\mathbf{V}^{-1} = \mathbf{V}^*$.

Another result which needs not to be true in the non-orthogonal case is given in the following lemma. This result leads to a more convenient measure for the goodness of approximation

Lemma 2
Let $(\mathcal{P}_m)_{0 \le m \le n} \in \mathcal{N}$ be a sequence of orthogonal projectors defined by a translation element $\mathbf{v}_0 \in \mathbf{Z}$ and orthonormal elements $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbf{Z}$. Then for the approximation error $\beta(\mathcal{P}_m, t, \omega)$ with respect to the loss function $l_1(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2$, $\mathbf{x}, \mathbf{y} \in \mathbf{Z}$, it holds for $0 < m_1 < m_2 < n$ and $\forall\, t \in [0, T]$, $\forall\, \omega \in \Omega$

$$0 = \beta(\mathcal{P}_n, t, \omega) \le \beta(\mathcal{P}_{m_2}, t, \omega) \le \beta(\mathcal{P}_{m_1}, t, \omega) \le \beta(\mathcal{P}_0, t, \omega),$$

i.e. $\beta(\mathcal{P}_m, t, \omega)$ is monotonically decreasing with respect to m.

Proof: In case of $0 \le m < n$ we obtain

$$\begin{aligned} \beta(\mathcal{P}_m, t, \omega) &= \|\mathcal{Q}_m \mathbf{z}(t, \omega)\|^2 = \left\| \sum_{k=m+1}^{n} x_k(t, \omega) \mathbf{v}_k \right\|^2 \\ &= \sum_{k,l=m+1}^{n} x_k(t, \omega) \overline{x_l}(t, \omega) \left(\mathbf{v}_k, \mathbf{v}_l\right) \\ &= \sum_{k=m+1}^{n} |x_k(t, \omega)|^2, \end{aligned} \tag{10}$$

by means of Eq. (4), where x_k, $k = 1, \dots, n$, denote coordinates of $\mathbf{z}$ with respect to the orthonormal elements $\mathbf{v}_1, \dots, \mathbf{v}_n$ and the translation element $\mathbf{v}_0$, i.e. $\mathbf{z} = \mathbf{v}_0 + \sum_{k=1}^{n} x_k \mathbf{v}_k$. Then using $m_1 < m_2$ it follows

$$\begin{aligned} 0 &= \|\mathcal{Q}_n \mathbf{z}(t, \omega)\|^2 = \beta(\mathcal{P}_n, t, \omega) \\ &\le \sum_{k=m_2+1}^{n} |x_k(t, \omega)|^2 = \|\mathcal{Q}_{m_2} \mathbf{z}(t, \omega)\|^2 = \beta(\mathcal{P}_{m_2}, t, \omega) \\ &\le \sum_{k=m_1+1}^{n} |x_k(t, \omega)|^2 = \|\mathcal{Q}_{m_1} \mathbf{z}(t, \omega)\|^2 = \beta(\mathcal{P}_{m_1}, t, \omega) \\ &\le \sum_{k=1}^{n} |x_k(t, \omega)|^2 = \|\mathbf{z}(t, \omega) - \mathbf{v}_0\|^2 = \beta(\mathcal{P}_0, t, \omega). \end{aligned}$$

■

Remark 1 The assumption of orthogonal elements $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in lemma 2 is essential. In case of non-orthogonal elements the inequalities

$$\beta(\mathcal{P}_{m_2}, t, \omega) \leq \beta(\mathcal{P}_{m_1}, t, \omega) \leq \beta(\mathcal{P}_0, t, \omega),$$

for $0 < m_1 < m_2 < n$ need not to be correct.

Definition 4
Let γ be the approximation risk of $\mathcal{P}_m$ with respect to the loss function $l_1(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2$ and a weighting function ψ.

i) For a sequence $(\mathcal{P}_m)_{0 \leq m \leq n} \in \mathcal{N}$ of orthogonal projection operators with $\gamma(\mathcal{P}_0) > 0$

$$\rho = \rho(\mathcal{P}_m) = 1 - \frac{\gamma(\mathcal{P}_m)}{\gamma(\mathcal{P}_0)}, \quad m = 0, \ldots, n,$$

is called <u>approximation efficiency</u> *of $\mathcal{P}_m$.*

ii) A sequence $(\mathcal{P}^*_m)_{0 \leq m \leq n} \in \mathcal{N}$ is said to be <u>ρ-optimal</u> *with respect to $\mathcal{N}$ if*

$$\rho(\mathcal{P}^*_m) \geq \rho(\mathcal{P}_m), \quad m = 0, \ldots, n$$

holds for all sequences $(\mathcal{P}_m)_{0 \leq m \leq n} \in \mathcal{N}$.

Remark 2 The condition $\gamma(\mathcal{P}_0) > 0$ is not restrictive, since

$$0 = \gamma(\mathcal{P}_0) = \int_{\mathcal{D}} \mathbf{E}\left\{\beta(\mathcal{P}_0, t)\right\} \psi(t) dt = \int_{\mathcal{D}} \mathbf{E}\left\{\|\mathbf{z}(t) - \mathbf{v}_0\|^2\right\} \psi(t) dt$$

only holds in the degenerate case $\mathbf{z}(t, \omega) \equiv \mathbf{v}_0 \quad \forall\, t \in \{s \in \mathcal{D} : \psi(s) \neq 0\}$ a.s. .

Corollary 1

i) Applying Lemma 2 a sequence $(\mathcal{P}_m)_{0 \leq m \leq n} \in \mathcal{N}$ fulfills the inequality

$$0 = \rho(\mathcal{P}_0) \leq \rho(\mathcal{P}_{m_1}) \leq \rho(\mathcal{P}_{m_2}) \leq \rho(\mathcal{P}_n) = 1$$

for $0 < m_1 < m_2 < n$.

ii) Assuming a fixed translation element $\mathbf{v}_0$ a sequence $(\mathcal{P}^*_m)_{0 \leq m \leq n} \in \mathcal{N}$ is ρ-optimal with respect to $\mathcal{N}$ if and only if $(\mathcal{P}^*_m)_{0 \leq m \leq n}$ is γ-optimal with respect to $\mathcal{N}$.

Now several representations of the approximation risk and efficiency are derived using sequences $(\mathcal{P}_m)_{0\leq m\leq n} \in \mathcal{N}$ defined by orthonormal elements $\mathbf{v}_1,\ldots,\mathbf{v}_n$. Using representation (10) of the approximation error we obtain for $(\mathcal{P}_m)_{0\leq m\leq n} \in \mathcal{N}$

$$\begin{aligned}
\beta(\mathcal{P}_m,t,\omega) = \|\mathcal{Q}_m\mathbf{z}(t,\omega)\|^2 &= \begin{cases} \sum\limits_{k=m+1}^{n} |x_k(t,\omega)|^2 & 0\leq m<n \\ 0 & m=n \end{cases} \\
&= \operatorname{tr}\left[\mathbf{Q}_m\mathbf{x}(t,\omega)\mathbf{x}^*(t,\omega)\mathbf{Q}_m\right],
\end{aligned}$$

where $\operatorname{tr}[.]$ denotes the trace of a matrix and $\mathbf{Q}_m$ is the complementary projection matrix defined in (5). Then the approximation risk can be written as

$$\begin{aligned}
\gamma(\mathcal{P}_m) &= \int\limits_{\mathcal{D}} \mathbf{E}\left\{\beta(\mathcal{P}_m,t)\right\}\psi(t)dt = \int\limits_{\mathcal{D}} \mathbf{E}\left\{\operatorname{tr}\left[\mathbf{Q}_m\mathbf{x}(t)\mathbf{x}^*(t)\mathbf{Q}_m\right]\right\}\psi(t)dt \\
&= \int\limits_{\mathcal{D}} \operatorname{tr}\left[\mathbf{Q}_m\mathbf{E}\left\{\mathbf{x}(t)\mathbf{x}^*(t)\right\}\mathbf{Q}_m\right]\psi(t)dt \\
&= \int\limits_{\mathcal{D}} \operatorname{tr}\left[\mathbf{Q}_m\left(\mathbf{K}_{\mathbf{xx}}(t,t)+\mathbf{m}_{\mathbf{x}}(t)\mathbf{m}_{\mathbf{x}}^*(t)\right)\mathbf{Q}_m\right]\psi(t)dt \\
&= \gamma_1(\mathcal{P}_m)+\gamma_2(\mathcal{P}_m),
\end{aligned}$$

where $\gamma_1(\mathcal{P}_m)$ and $\gamma_2(\mathcal{P}_m)$ are defined by

$$\begin{aligned}
\gamma_1(\mathcal{P}_m) &:= \int\limits_{\mathcal{D}} \operatorname{tr}\left[\mathbf{Q}_m\mathbf{m}_{\mathbf{x}}(t)\mathbf{m}_{\mathbf{x}}^*(t)\mathbf{Q}_m\right]\psi(t)dt = \int\limits_{\mathcal{D}} \left\|\mathbf{Q}_m\mathbf{m}_{\mathbf{x}}(t)\right\|^2\psi(t)dt \\
\gamma_2(\mathcal{P}_m) &:= \int\limits_{\mathcal{D}} \operatorname{tr}\left[\mathbf{Q}_m\mathbf{K}_{\mathbf{xx}}(t,t)\mathbf{Q}_m\right]\psi(t)dt.
\end{aligned} \tag{11}$$

Representations of $\gamma_1(\mathcal{P}_m)$ and $\gamma_2(\mathcal{P}_m)$ in terms of the moments of $\mathbf{z}(t,\omega)$ can be derived using the relations

$$\mathbf{z}(t,\omega) = \mathbf{v}_0+\mathbf{V}\mathbf{x}(t,\omega) \quad \text{and} \quad \mathbf{x}(t,\omega) = \mathbf{V}^*(\mathbf{z}(t,\omega)-\mathbf{v}_0)$$

with an unitary matrix $\mathbf{V}$. They imply the following identities for the mean and the correlation function

$$\begin{aligned}
\mathbf{m}_{\mathbf{x}}(t) &= \mathbf{V}^*(\mathbf{m}_{\mathbf{z}}(t)-\mathbf{v}_0) \\
\mathbf{K}_{\mathbf{xx}}(t,s) &= \mathbf{V}^*\mathbf{K}_{\mathbf{zz}}(t,s)\mathbf{V}
\end{aligned} \tag{12}$$

and it follows

$$\begin{aligned}
\gamma_1(\mathcal{P}_m) &= \int\limits_{\mathcal{D}} \left\|\mathbf{Q}_m\mathbf{V}^*(\mathbf{m}_{\mathbf{z}}(t)-\mathbf{v}_0)\right\|^2\psi(t)dt \\
\gamma_2(\mathcal{P}_m) &= \int\limits_{\mathcal{D}} \operatorname{tr}\left[\mathbf{Q}_m\mathbf{V}^*\mathbf{K}_{\mathbf{zz}}(t,t)\mathbf{V}\mathbf{Q}_m\right]\psi(t)dt.
\end{aligned} \tag{13}$$

For square matrices $\mathbf{A}$ and $\mathbf{C}$ where $\mathbf{C}^{-1}$ exists it holds the identity

$$\mathrm{tr}\left[\mathbf{C}\mathbf{A}\mathbf{C}^{-1}\right] = \mathrm{tr}\left[\mathbf{A}\right]. \tag{14}$$

Substituting $\mathbf{C} = \mathbf{V},\ \mathbf{C}^{-1} = \mathbf{V}^*$ and applying Eq. (4)

$$\mathcal{Q}_m \mathbf{z}(t,\omega) = \mathbf{V}\mathbf{Q}_m\mathbf{V}^*(\mathbf{z}(t,\omega) - \mathbf{v}_0) = \mathbf{V}\mathbf{Q}_m\mathbf{x}(t,\omega)$$

we derive by means of Eqs. (11) and (13) a representation in terms of the complementary projection $\mathcal{Q}_m\mathbf{z}(t,\omega)$

$$\begin{aligned} \gamma_1(\mathcal{P}_m) &= \int\limits_{\mathcal{D}} \left\|\mathbf{m}_{\mathcal{Q}_m\mathbf{z}}(t)\right\|^2 \psi(t)dt \\ \gamma_2(\mathcal{P}_m) &= \int\limits_{\mathcal{D}} \mathrm{tr}\left[\mathbf{K}_{\mathcal{Q}_m\mathbf{z}\mathcal{Q}_m\mathbf{z}}(t,t)\right]\psi(t)dt. \end{aligned} \tag{15}$$

Eqs. (15) show that $\gamma_1(\mathcal{P}_m)$ is based on the mean of the complementary projection $\mathcal{Q}_m\mathbf{z}$ and $\gamma_2(\mathcal{P}_m)$ is based on the correlation function of $\mathcal{Q}_m\mathbf{z}$, only.

Similar representations can be found for the approximation efficiency $\rho(\mathcal{P}_m)$ which is introduced in Definition 4.

In the following considerations the goodness of the approximation of first- and second-order moments of $\mathbf{z}(t,\omega)$ by the corresponding moments of the projection $\mathcal{P}_m\mathbf{z}(t,\omega)$ is studied. Again, let us start with the definition of a non-negative loss function which gives a pointwise measure for the approximation error. In order to derive a non-negative real number which characterizes the goodness of the projection with respect to the approximation of the moments this pointwise measure is averaged with respect to t.

Definition 5
The non-negative functions

$$\begin{aligned} l^{\mathbf{m}}(.,.): &\quad \mathbf{Z}\times\mathbf{Z} \to \mathbb{R}_+ \\ \textit{and}\quad l^{\mathbf{K}}(.,.): &\quad \mathbf{Z}^2\times\mathbf{Z}^2 \to \mathbb{R}_+ \end{aligned}$$

are called <u>loss functions</u> *with respect to the statistical characteristics* $\mathbf{m}(.)$ *and* $\mathbf{K}(.,.)$, *respectively.*

To a given approximation $\tilde{\mathbf{z}}$ *of* $\mathbf{z}$ *the quantities*

$$\begin{aligned} \delta^{\mathbf{m}}(\mathbf{z},\tilde{\mathbf{z}}) &:= \int\limits_{\mathcal{D}} l^{\mathbf{m}}(\mathbf{m}_{\mathbf{z}}(t),\mathbf{m}_{\tilde{\mathbf{z}}}(t))\psi(t)dt \\ \delta^{\mathbf{K}}(\mathbf{z},\tilde{\mathbf{z}}) &:= \int\limits_{\mathcal{D}}\int\limits_{\mathcal{D}} l^{\mathbf{K}}(\mathbf{K}_{\mathbf{z}\mathbf{z}}(t,s),\mathbf{K}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}(t,s))\psi(t)\psi(s)dtds \end{aligned}$$

are called <u>approximation error</u> *of the approximation* $\tilde{\mathbf{z}}$ *of* $\mathbf{z}$ *with respect to* $\mathbf{m}(.)$ *and* $\mathbf{K}(.,.)$, *the loss function* $l^{\mathbf{m}}$ *and* $l^{\mathbf{K}}$ *and the weighting function* ψ, *respectively.*

For the approximation of $\mathbf{z}$ by its projection, i.e. $\tilde{\mathbf{z}} = \mathcal{P}_m\mathbf{z}$, the next lemma gives representations and estimates of the introduced measures of the approximation error in terms of the approximation risk $\gamma(\mathcal{P}_m)$. These estimates motivate the choice of γ as objective for the optimization in Section 4.

Lemma 3
Assuming a sequence $(\mathcal{P}_m)_{0\leq m\leq n} \in \mathcal{N}$ of orthogonal projection operators it holds

i) for the approximation error with respect to $\mathbf{m}(.)$ and the loss function $l^{\mathbf{m}}(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|^2$, $\mathbf{x},\mathbf{y}\in\mathbf{Z}$,

$$\delta^{\mathbf{m}}(\mathbf{z},\mathcal{P}_m\mathbf{z}) = \gamma_1(\mathcal{P}_m), \qquad m=0,\dots,n,$$

ii) for the approximation error with respect to $\mathbf{K}(.,.)$ and the loss function $l^{\mathbf{K}}(\mathbf{X},\mathbf{Y}) = |\mathrm{tr}\,[\mathbf{X}] - \mathrm{tr}\,[\mathbf{Y}]\,|$, $\quad \mathbf{X},\mathbf{Y}\in\mathbf{Z}^2$,

$$\delta^{\mathbf{K}}(\mathbf{z},\mathcal{P}_m\mathbf{z}) \leq \gamma_2(\mathcal{P}_m), \qquad m=0,\dots,n,$$

iii) for the approximation error with respect to $\mathbf{K}(.,.)$ and the loss function $l^{\mathbf{K}}(\mathbf{X},\mathbf{Y}) = \|\mathbf{X}-\mathbf{Y}\|^2_{\mathbf{Z}^2} = \sum_{i,j=1}^{n}|\mathbf{X}_{ij}-\mathbf{Y}_{ij}|^2$, $\quad \mathbf{X},\mathbf{Y}\in\mathbf{Z}^2$,

$$\delta^{\mathbf{K}}(\mathbf{z},\mathcal{P}_m\mathbf{z}) \;\leq\; 3\gamma_2(\mathcal{P}_m)(2\gamma_2(\mathcal{P}_0)-\gamma_2(\mathcal{P}_m)), \qquad m=0,\dots,n.$$

The quantities $\gamma_1(\mathcal{P}_m)$ and $\gamma_2(\mathcal{P}_m)$ are given by Eqs. (11).

Proof:

i) By means of the chosen loss function and Eqs. (4) and (15) it follows from the definition of $\delta^{\mathbf{m}}$

$$\delta^{\mathbf{m}}(\mathbf{z},\mathcal{P}_m\mathbf{z}) = \int_{\mathcal{D}} \|\mathbf{m}_{\mathbf{z}}(t) - \mathbf{m}_{\mathcal{P}_m\mathbf{z}}(t)\|^2\,\psi(t)dt = \int_{\mathcal{D}} \|\mathbf{m}_{\mathcal{Q}_m\mathbf{z}}(t)\|^2\,\psi(t)dt = \gamma_1(\mathcal{P}_m).$$

ii) In case of the chosen loss function the approximation error $\delta^{\mathbf{K}}$ can be written as

$$\delta^{\mathbf{K}}(\mathbf{z},\mathcal{P}_m\mathbf{z}) = \int_{\mathcal{D}}\int_{\mathcal{D}} |\mathrm{tr}\,[\mathbf{K}_{\mathbf{zz}}(t,s)] - \mathrm{tr}\,[\mathbf{K}_{\mathcal{P}_m\mathbf{z}\mathcal{P}_m\mathbf{z}}(t,s)]|\,\psi(t)\psi(s)dtds.$$

Using $\mathbf{z}(t,\omega) = \mathbf{v}_0 + \mathbf{V}\mathbf{x}(t,\omega)$ and $\mathcal{P}_m\mathbf{z}(t,\omega) = \mathbf{v}_0 + \mathbf{V}\mathbf{P}_m\mathbf{x}(t,\omega)$ (see Eq. (3)) we obtain

$$\mathbf{K}_{\mathbf{zz}}(t,s) = \mathbf{V}\mathbf{K}_{\mathbf{xx}}(t,s)\mathbf{V}^* \quad \text{and} \quad \mathbf{K}_{\mathcal{P}_m\mathbf{z}\mathcal{P}_m\mathbf{z}}(t,s) = \mathbf{V}\mathbf{P}_m\mathbf{K}_{\mathbf{xx}}(t,s)\mathbf{P}_m\mathbf{V}^*$$

and by means of Eq. (14) and Cauchy-Schwarz inequality

$$
\begin{aligned}
&|\mathrm{tr}\,[\mathbf{K}_{\mathbf{zz}}(t,s)] - \mathrm{tr}\,[\mathbf{K}_{\mathcal{P}_m\mathbf{z}\mathcal{P}_m\mathbf{z}}(t,s)]| = |\mathrm{tr}\,[\mathbf{K}_{\mathbf{xx}}(t,s)] - \mathrm{tr}\,[\mathbf{P}_m\mathbf{K}_{\mathbf{xx}}(t,s)\mathbf{P}_m]| \\
&= \quad |\mathrm{tr}\,[\mathbf{Q}_m\mathbf{K}_{\mathbf{xx}}(t,s)\mathbf{Q}_m]| = \left|\sum_{k=m+1}^{n} R_{x_kx_k}(t,s)\right| \qquad (16)\\
&\leq \quad \sum_{k=m+1}^{n} |R_{x_kx_k}(t,s)| \leq \sum_{k=m+1}^{n} \sqrt{R_{x_kx_k}(t,t)R_{x_kx_k}(s,s)}.
\end{aligned}
$$

Then it follows

$$
\begin{aligned}
\delta^{\mathbf{K}}(\mathbf{z},\mathcal{P}_m\mathbf{z}) &\leq \int_{\mathcal{D}}\int_{\mathcal{D}} \sum_{k=m+1}^{n} \sqrt{R_{x_kx_k}(t,t)R_{x_kx_k}(s,s)}\psi(t)\psi(s)dtds \\
&= \sum_{k=m+1}^{n} \left(\int_{\mathcal{D}} \sqrt{R_{x_kx_k}(t,t)}\psi(t)dt\right)^2 \\
&\leq \sum_{k=m+1}^{n} \int_{\mathcal{D}} R_{x_kx_k}(t,t)\psi(t)dt \cdot \int_{\mathcal{D}} \psi(t)dt \\
&= \int_{\mathcal{D}} \mathrm{tr}\,[\mathbf{Q}_m\mathbf{K}_{\mathbf{xx}}(t,t)\mathbf{Q}_m]\,\psi(t)dt = \gamma_2(\mathcal{P}_m)
\end{aligned}
$$

(see (11)).

iii) In the third case the definition of $\delta^{\mathbf{K}}$ leads to

$$
\delta^{\mathbf{K}}(\mathbf{z},\mathcal{P}_m\mathbf{z}) = \int_{\mathcal{D}}\int_{\mathcal{D}} \sum_{i,j=1}^{n} \left|R_{z_iz_j}(t,s) - R_{[\mathcal{P}_m\mathbf{z}]_i[\mathcal{P}_m\mathbf{z}]_j}(t,s)\right|^2 \psi(t)\psi(s)dtds,
$$

where $[.]_i$ denotes the i-th component of a vector.
Because of $\mathbf{z} = \mathcal{P}_m\mathbf{z} + \mathcal{Q}_m\mathbf{z}$ it follows

$$
\mathbf{K}_{\mathbf{zz}}(t,s) = \mathbf{K}_{\mathcal{P}_m\mathbf{z}\mathcal{P}_m\mathbf{z}}(t,s) + \mathbf{K}_{\mathcal{P}_m\mathbf{z}\mathcal{Q}_m\mathbf{z}}(t,s) + \mathbf{K}_{\mathcal{Q}_m\mathbf{z}\mathcal{P}_m\mathbf{z}}(t,s) + \mathbf{K}_{\mathcal{Q}_m\mathbf{z}\mathcal{Q}_m\mathbf{z}}(t,s).
$$

and using again Cauchy-Schwarz inequality it holds

$$
\begin{aligned}
&\left|R_{z_iz_j}(t,s) - R_{[\mathcal{P}_m\mathbf{z}]_i[\mathcal{P}_m\mathbf{z}]_j}(t,s)\right|^2 \\
&= \quad \left|R_{[\mathcal{P}_m\mathbf{z}]_i[\mathcal{Q}_m\mathbf{z}]_j}(t,s) + R_{[\mathcal{Q}_m\mathbf{z}]_i[\mathcal{P}_m\mathbf{z}]_j}(t,s) + R_{[\mathcal{Q}_m\mathbf{z}]_i[\mathcal{Q}_m\mathbf{z}]_j}(t,s)\right|^2 \\
&\leq \quad 3\left(\left|R_{[\mathcal{P}_m\mathbf{z}]_i[\mathcal{Q}_m\mathbf{z}]_j}(t,s)\right|^2 + \left|R_{[\mathcal{Q}_m\mathbf{z}]_i[\mathcal{P}_m\mathbf{z}]_j}(t,s)\right|^2 + \left|R_{[\mathcal{Q}_m\mathbf{z}]_i[\mathcal{Q}_m\mathbf{z}]_j}(t,s)\right|^2\right) \\
&\leq \quad 3\left(R_{[\mathcal{P}_m\mathbf{z}]_i[\mathcal{P}_m\mathbf{z}]_i}(t,t)\; R_{[\mathcal{Q}_m\mathbf{z}]_j[\mathcal{Q}_m\mathbf{z}]_j}(s,s)\right. \\
&\qquad +R_{[\mathcal{Q}_m\mathbf{z}]_i[\mathcal{Q}_m\mathbf{z}]_i}(t,t)\; R_{[\mathcal{P}_m\mathbf{z}]_j[\mathcal{P}_m\mathbf{z}]_j}(s,s) \\
&\qquad \left.+\, R_{[\mathcal{Q}_m\mathbf{z}]_i[\mathcal{Q}_m\mathbf{z}]_i}(t,t)\; R_{[\mathcal{Q}_m\mathbf{z}]_j[\mathcal{Q}_m\mathbf{z}]_j}(s,s)\right).
\end{aligned}
$$

The substitution of this inequality into the representation of $\delta^{\mathbf{K}}(\mathbf{z},\mathcal{P}_m\mathbf{z})$ leads to

$$\begin{aligned}
\delta^{\mathbf{K}}(\mathbf{z},\mathcal{P}_m\mathbf{z}) \le 3\int\limits_{\mathcal{D}}\int\limits_{\mathcal{D}} & \left[\left(\sum_{i=1}^{n} R_{[\mathcal{P}_m\mathbf{z}]_i[\mathcal{P}_m\mathbf{z}]_i}(t,t)\right)\left(\sum_{j=1}^{n} R_{[\mathcal{Q}_m\mathbf{z}]_j[\mathcal{Q}_m\mathbf{z}]_j}(s,s)\right)\right.\\
&+\left(\sum_{i=1}^{n} R_{[\mathcal{Q}_m\mathbf{z}]_i[\mathcal{Q}_m\mathbf{z}]_i}(t,t)\right)\left(\sum_{j=1}^{n} R_{[\mathcal{P}_m\mathbf{z}]_j[\mathcal{P}_m\mathbf{z}]_j}(s,s)\right)\\
&\left.+\left(\sum_{i=1}^{n} R_{[\mathcal{Q}_m\mathbf{z}]_i[\mathcal{Q}_m\mathbf{z}]_i}(t,t)\right)\left(\sum_{j=1}^{n} R_{[\mathcal{Q}_m\mathbf{z}]_j[\mathcal{Q}_m\mathbf{z}]_j}(s,s)\right)\right]\psi(t)\psi(s)dtds\\
= \quad 3 & \left[\int\limits_{\mathcal{D}}\mathrm{tr}\left[\mathbf{K}_{\mathcal{P}_m\mathbf{z}\mathcal{P}_m\mathbf{z}}(t,t)\right]\psi(t)dt\cdot\int\limits_{\mathcal{D}}\mathrm{tr}\left[\mathbf{K}_{\mathcal{Q}_m\mathbf{z}\mathcal{Q}_m\mathbf{z}}(s,s)\right]\psi(s)ds\right.\\
&+\int\limits_{\mathcal{D}}\mathrm{tr}\left[\mathbf{K}_{\mathcal{Q}_m\mathbf{z}\mathcal{Q}_m\mathbf{z}}(t,t)\right]\psi(t)dt\cdot\int\limits_{\mathcal{D}}\mathrm{tr}\left[\mathbf{K}_{\mathcal{P}_m\mathbf{z}\mathcal{P}_m\mathbf{z}}(s,s)\right]\psi(s)ds\\
&\left.+\int\limits_{\mathcal{D}}\mathrm{tr}\left[\mathbf{K}_{\mathcal{Q}_m\mathbf{z}\mathcal{Q}_m\mathbf{z}}(t,t)\right]\psi(t)dt\cdot\int\limits_{\mathcal{D}}\mathrm{tr}\left[\mathbf{K}_{\mathcal{Q}_m\mathbf{z}\mathcal{Q}_m\mathbf{z}}(s,s)\right]\psi(s)ds\right].
\end{aligned}$$

From Eq. (15), $\int\limits_{\mathcal{D}}\mathrm{tr}\left[\mathbf{K}_{\mathcal{Q}_m\mathbf{z}\mathcal{Q}_m\mathbf{z}}(t,t)\right]\psi(t)dt = \gamma_2(\mathcal{P}_m)$, and (16) we can deduce

$$\begin{aligned}
\int\limits_{\mathcal{D}}\mathrm{tr}\left[\mathbf{K}_{\mathcal{P}_m\mathbf{z}\mathcal{P}_m\mathbf{z}}(t,t)\right]\psi(t)dt &= \int\limits_{\mathcal{D}}(\mathrm{tr}\left[\mathbf{K}_{\mathbf{z}\mathbf{z}}(t,t)\right]-\mathrm{tr}\left[\mathbf{K}_{\mathcal{Q}_m\mathbf{z}\mathcal{Q}_m\mathbf{z}}(t,t)\right])\psi(t)dt\\
&= \gamma_2(\mathcal{P}_0)-\gamma_2(\mathcal{P}_m),
\end{aligned}$$

hence

$$\begin{aligned}
\delta^{\mathbf{K}}(\mathbf{z},\mathcal{P}_m\mathbf{z}) &\le 3\gamma_2(\mathcal{P}_m)\left(2(\gamma_2(\mathcal{P}_0)-\gamma_2(\mathcal{P}_m))+\gamma_2(\mathcal{P}_m)\right)\\
&= 3\gamma_2(\mathcal{P}_m)\left(2\gamma_2(\mathcal{P}_0)-\gamma_2(\mathcal{P}_m)\right).
\end{aligned}$$

■

Finally, the case of wide-sense stationary vector functions is considered, i.e. it holds

$$\mathbf{m}_{\mathbf{z}}(t)\equiv\mathbf{m}_{\mathbf{z}} \qquad \text{and} \qquad \mathbf{K}_{\mathbf{z}\mathbf{z}}(t,s) =: \mathbf{R}_{\mathbf{z}\mathbf{z}}(s-t).$$

Evaluating the integrals in Eqs. (13) it follows

$$\gamma(\mathcal{P}_m)=\gamma_1(\mathcal{P}_m)+\gamma_2(\mathcal{P}_m)=\|\mathbf{Q}_m\mathbf{V}^*(\mathbf{m}_{\mathbf{z}}-\mathbf{v}_0)\|^2+\mathrm{tr}\left[\mathbf{Q}_m\mathbf{V}^*\mathbf{R}_{\mathbf{z}\mathbf{z}}(0)\mathbf{V}\mathbf{Q}_m\right].$$

Then, the approximation error with respect to $\mathbf{m}(.)$ can be written as

$$\delta^{\mathbf{m}}(\mathbf{z},\mathcal{P}_m\mathbf{z}) \quad = \quad l^{\mathbf{m}}(\mathbf{m}_{\mathbf{z}},\mathbf{m}_{\mathcal{P}_m\mathbf{z}}).$$

Furthermore, for the approximation error with respect to $\mathbf{K}(.,.)$ hold the following relations

$$\begin{aligned}
\delta^{\mathbf{K}}(\mathbf{z},\tilde{\mathbf{z}}) &= \int_{\mathcal{D}}\int_{\mathcal{D}} l^{\mathbf{K}}(\mathbf{K}_{\mathbf{zz}}(t,s),\mathbf{K}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}(t,s))\psi(t)\psi(s)dtds \\
&= \int_{\mathcal{D}}\int_{\{\tau\in\mathbb{R}:t+\tau\in\mathcal{D}\}} l^{\mathbf{K}}(\mathbf{K}_{\mathbf{zz}}(t,t+\tau),\mathbf{K}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}(t,t+\tau))\psi(t)\psi(t+\tau)dtd\tau \\
&= \int_{\{\tau\in\mathbb{R}:\tau=s-t,s,t\in\mathcal{D}\}}\int_{\{t\in\mathcal{D}:t+\tau\in\mathcal{D}\}} l^{\mathbf{K}}(\mathbf{K}_{\mathbf{zz}}(t,t+\tau),\mathbf{K}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}(t,t+\tau))\psi(t)\psi(t+\tau)dtd\tau \\
&= \int_{\mathcal{T}} l^{\mathbf{K}}(\mathbf{R}_{\mathbf{zz}}(\tau),\mathbf{R}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}(\tau))\breve{\psi}(\tau)d\tau
\end{aligned}$$

where the notations

$$\mathcal{T} := \{\tau\in\mathbb{R} : \tau = s-t, s,t\in\mathcal{D}\} \quad \text{and} \quad \breve{\psi}(\tau) := \int_{\{t\in\mathcal{D}:t+\tau\in\mathcal{D}\}} \psi(t)\psi(t+\tau)dt,\ \tau\in\mathcal{T}$$

are used. Obviously, it holds $\int_{\mathcal{T}}\breve{\psi}(\tau)d\tau = 1$. These relations motivate the introduction of a slightly modified approximation error with repect to $\mathbf{R}(.)$ using an arbitrary weighting function $\theta(.) : \mathcal{T}\to\mathbb{R}_+$ with $\int_{\mathcal{T}}\theta(\tau)d\tau = 1$.

Definition 6
To a given wide-sense stationary approximation $\tilde{\mathbf{z}}$ of wide-sense stationary vector function $\mathbf{z}$ the quantity

$$\delta^{\mathbf{R}}(\mathbf{z},\tilde{\mathbf{z}}) \;:=\; \int_{\mathcal{T}} l^{\mathbf{K}}(\mathbf{R}_{\mathbf{zz}}(\tau),\mathbf{R}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}(\tau))\,\theta(\tau)d\tau$$

is called approximation error *of the approximation $\tilde{\mathbf{z}}$ of $\mathbf{z}$ with respect to $\mathbf{R}(\tau)$, the loss function $l^{\mathbf{K}}$ and the weighting function θ.*

The next Lemma provides estimates for $\delta^{\mathbf{R}}$ in terms of the approximation risk which are the corresponding counterparts of the estimates for $\delta^{\mathbf{K}}$ given in Lemma 3 for the general case.

Lemma 4
Assuming a wide-sense stationary vector function $\mathbf{z}$ and a sequence of orthogonal projection operators $(\mathcal{P}_m)_{0\le m\le n}\in\mathcal{N}$ then it holds for the approximation error with respect to $\mathbf{R}_{\mathbf{zz}}(\tau)$ and

i) a loss function
$l^{\mathbf{K}}(\mathbf{X},\mathbf{Y}) = |\mathrm{tr}\,[\mathbf{X}] - \mathrm{tr}\,[\mathbf{Y}]\,|, \quad \mathbf{X},\mathbf{Y}\in\mathbf{Z}^2$

$$\delta^{\mathbf{R}}(\mathbf{z},\mathcal{P}_m\mathbf{z}) \le \gamma_2(\mathcal{P}_m) = \mathrm{tr}\,[\mathbf{Q}_m\mathbf{V}^*\mathbf{R}_{\mathbf{zz}}(0)\mathbf{V}\mathbf{Q}_m]\,, \qquad m = 0,\ldots,n,$$

ii) a loss function

$$l^{\mathbf{K}}(\mathbf{X},\mathbf{Y}) = \|\mathbf{X}-\mathbf{Y}\|^2_{\mathbf{Z}^2} = \sum_{i,j=1}^{n} |\mathbf{X}_{ij}-\mathbf{Y}_{ij}|^2, \quad \mathbf{X},\mathbf{Y}\in\mathbf{Z}^2,$$

$$\delta^{\mathbf{R}}(\mathbf{z},\mathcal{P}_m\mathbf{z}) \leq 3\gamma_2(\mathcal{P}_m)(2\gamma_2(\mathcal{P}_0)-\gamma_2(\mathcal{P}_m)), \qquad m=0,\dots,n.$$

Proof: The proof can be obtained using the techniques of the proofs of Lemma 3 (ii) and (iii).

Remark 3 While the approximation error $\delta^{\mathbf{R}}(\mathbf{z},\mathcal{P}_m\mathbf{z})$ depends on the weighting function θ the upper bounds for $\delta^{\mathbf{R}}(\mathbf{z},\mathcal{P}_m\mathbf{z})$ derived in the lemma above do not depend on the selection of θ.

In a similar way the goodness of the approximation of the spectral density

$$\mathbf{S}_{\mathbf{zz}}(\alpha) := \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\alpha\tau}\mathbf{R}_{\mathbf{zz}}(\tau)d\tau, \qquad \alpha\in\mathbb{R},$$

can be treated.

Definition 7
The non-negative function

$$l^{\mathbf{S}}(.,.): \qquad \mathbf{Z}^2\times\mathbf{Z}^2\to\mathbb{R}_+$$

is called <u>loss function</u> *with respect to the spectral density* $\mathbf{S}(\alpha)$*.*

For a wide-sense stationary approximation $\tilde{\mathbf{z}}$ *of the wide-sense stationary vector function* $\mathbf{z}$ *the quantity*

$$\delta^{\mathbf{S}}(\mathbf{z},\tilde{\mathbf{z}}) \;:=\; \int_{-\infty}^{\infty} l^{\mathbf{S}}(\mathbf{S}_{\mathbf{zz}}(\alpha),\mathbf{S}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}(\alpha))\,d\alpha$$

is called <u>approximation error</u> *of the approximation* $\tilde{\mathbf{z}}$ *of* $\mathbf{z}$ *with respect to* $\mathbf{S}(.)$ *and the loss function* $l^{\mathbf{S}}$*.*

Lemma 5
Assuming a wide-sense stationary vector function $\mathbf{z}$*, a sequence* $(\mathcal{P}_m)_{0\leq m\leq n}\in\mathcal{N}$ *of orthogonal projection operators and a loss function*

$$l^{\mathbf{S}}(\mathbf{X},\mathbf{Y}) = |\mathrm{tr}\,[\mathbf{X}]-\mathrm{tr}\,[\mathbf{Y}]\,|, \quad \mathbf{X},\mathbf{Y}\in\mathbf{Z}^2$$

then it holds for the approximation error with respect to $\mathbf{S}(.)$

$$\delta^{\mathbf{S}}(\mathbf{z},\tilde{\mathbf{z}}) = \gamma_2(\mathcal{P}_m) = \mathrm{tr}\,[\mathbf{Q}_m\mathbf{V}^*\mathbf{R}_{\mathbf{zz}}(0)\mathbf{V}\mathbf{Q}_m]\,, \qquad m=0,\dots,n.$$

Proof: Using the identities

$$\mathbf{R}_{\mathbf{zz}}(\tau) = \mathbf{V}\mathbf{R}_{\mathbf{xx}}(\tau)\mathbf{V}^* \qquad \text{and} \qquad \mathbf{R}_{\mathcal{P}_m\mathbf{z}\mathcal{P}_m\mathbf{z}}(\tau) = \mathbf{V}\mathbf{P}_m\mathbf{R}_{\mathbf{xx}}(\tau)\mathbf{P}_m\mathbf{V}^*$$

it follows

$$\mathbf{S}_{\mathbf{zz}}(\alpha) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-i\alpha\tau}\mathbf{V}\mathbf{R}_{\mathbf{xx}}(\tau)\mathbf{V}^* d\tau = \mathbf{V}\mathbf{S}_{\mathbf{xx}}(\alpha)\mathbf{V}^*, \qquad \alpha \in \mathbb{R},$$

and analogously $\mathbf{S}_{\mathcal{P}_m\mathbf{z}\mathcal{P}_m\mathbf{z}}(\alpha) = \mathbf{V}\mathbf{P}_m\mathbf{S}_{\mathbf{xx}}(\alpha)\mathbf{P}_m\mathbf{V}^*$. Applying Eq. (14) the approximation error $\delta^{\mathbf{S}}$ can be written as

$$\begin{aligned}
\delta^{\mathbf{S}}(\mathbf{z}, \mathcal{P}_m\mathbf{z}) &= \int\limits_{-\infty}^{\infty} \left|\mathrm{tr}\left[\mathbf{S}_{\mathbf{zz}}(\alpha)\right] - \mathrm{tr}\left[\mathbf{S}_{\mathcal{P}_m\mathbf{z}\mathcal{P}_m\mathbf{z}}(\alpha)\right]\right| d\alpha \\
&= \int\limits_{-\infty}^{\infty} \left|\mathrm{tr}\left[\mathbf{S}_{\mathbf{xx}}(\alpha)\right] - \mathrm{tr}\left[\mathbf{P}_m\mathbf{S}_{\mathbf{xx}}(\alpha)\mathbf{P}_m\right]\right| d\alpha = \int\limits_{-\infty}^{\infty} \sum_{k=m+1}^{n} S_{x_k x_k}(\alpha)\, d\alpha \\
&= \sum_{k=m+1}^{n} R_{x_k x_k}(0) = \mathrm{tr}\left[\mathbf{Q}_m\mathbf{R}_{\mathbf{xx}}(0)\mathbf{Q}_m\right] = \mathrm{tr}\left[\mathbf{Q}_m\mathbf{V}^*\mathbf{R}_{\mathbf{zz}}(0)\mathbf{V}\mathbf{Q}_m\right] \\
&= \gamma_2(\mathcal{P}_m).
\end{aligned}$$

■

4 Optimal projection

This section considers the problem of finding γ-optimal sequences $(\mathcal{P}_m^*)_{0\le m\le n}$ with respect to the set $\mathcal{N}$ of orthogonal projectors, which are also γ-optimal with respect to the set $\mathcal{M}$ of all sequences of projectors (see Lemma 1). The optimization problem will be solved for the case of wide-sense stationary vector functions $\mathbf{z}(t,\omega)$.

First we note that the relations

$$\gamma_1(\mathcal{P}_m) \ge 0 \quad \text{and} \quad \gamma_1(\mathcal{P}_m) = 0 \quad \text{for} \quad \mathbf{m}_{\mathbf{z}} = \mathbf{v}_0 \tag{17}$$

hold for $(\mathcal{P}_m)_{0\le m\le n} \in \mathcal{M}$. To find a lower bound for $\gamma_2(\mathcal{P}_m)$ the eigenvalue decomposition of the variance matrix $\mathbf{R}_{\mathbf{zz}}(0)$ is considered. $\mathbf{R}_{\mathbf{zz}}(0)$ is assumed to be non-singular, then it is a hermitian ($\mathbf{R}_{\mathbf{zz}}^*(0) = \mathbf{R}_{\mathbf{zz}}(0)$) and (strictly) positive definite matrix. As a standard result of algebraic eigenvalue theory it is known that $\mathbf{R}_{\mathbf{zz}}(0)$ possesses n real and (strictly) positive eigenvalues $\sigma_1^2 \ge \ldots \ge \sigma_n^2 > 0$ and that n pairwise orthonormal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbf{Z}$ exist which form an unitary matrix $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ (i.e. $\mathbf{U}^{-1} = \mathbf{U}^*$). The decomposition

$$\mathbf{R}_{\mathbf{zz}}(0) = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^*$$

with $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ can be derived.

Lemma 6
Let $\mathbf{z}(t,\omega)$ be a wide-sense stationary vector function with a non-singular variance matrix $\mathbf{R}_{\mathbf{zz}}(0)$ which possesses the eigenvalues $\sigma_1^2 \geq \ldots \geq \sigma_n^2 > 0$. Then it holds for all sequences $(\mathcal{P}_m)_{0\leq m\leq n} \in \mathcal{N}$ (of orthogonal projection operators) the inequality

$$\gamma_2(\mathcal{P}_m) \geq \sum_{k=m+1}^{n} \sigma_k^2.$$

In case of $\mathbf{V} = \mathbf{U}$ the equality is fulfilled where $\mathbf{V}$ denotes the matrix belonging to $(\mathcal{P}_m)_{0\leq m\leq n}$ (see Eq. (3)).

Proof: First the assertion concerning the equality for $\mathbf{V} = \mathbf{U}$ will be proved. Applying the decomposition $\mathbf{R}_{\mathbf{zz}}(0) = \mathbf{U\Sigma U}^*$, the relation $\mathbf{U}^*\mathbf{U} = \mathbf{I}_n$ and the representation (5) of the matrix $\mathbf{Q}_m$ it can be found that

$$\begin{aligned}
\gamma_2(\mathcal{P}_m) &= \operatorname{tr}\left[\mathbf{Q}_m\mathbf{V}^*\mathbf{R}_{\mathbf{zz}}(0)\mathbf{V}\mathbf{Q}_m\right] \\
&= \operatorname{tr}\left[\mathbf{Q}_m\mathbf{U}^*(\mathbf{U}\Sigma\mathbf{U}^*)\mathbf{U}\mathbf{Q}_m\right] = \operatorname{tr}\left[\mathbf{Q}_m\Sigma\mathbf{Q}_m\right] \\
&= \operatorname{tr}\left[\begin{pmatrix} \mathbf{O}_{m,m} & \mathbf{O}_{m,n-m} \\ \mathbf{O}_{n-m,m} & \Sigma^{\mathbf{c}}_{n-m} \end{pmatrix}\right] = \sum_{k=m+1}^{n} \sigma_k^2
\end{aligned}$$

where $\Sigma^{\mathbf{c}}_{n-m}$ is a $n-m \times n-m$ submatrix of the matrix Σ, which is defined by $\Sigma^{\mathbf{c}}_{n-m} := \operatorname{diag}(\sigma_{m+1}^2, \ldots, \sigma_n^2)$.

Now we deal with the general assertion. Since in case of $(\mathcal{P}_m)_{0\leq m\leq n} \in \mathcal{N}$ the matrix $\mathbf{V}$ is unitary it follows that the matrix $\mathbf{B} := \mathbf{V}^*\mathbf{R}_{\mathbf{zz}}(0)\mathbf{V}$ is the result of a similarity transformation of the variance matrix $\mathbf{R}_{\mathbf{zz}}(0)$. Therefore $\mathbf{B}$ and $\mathbf{R}_{\mathbf{zz}}(0)$ possess the same positive eigenvalues $\sigma_1^2, \ldots, \sigma_n^2$. Further it holds

$$\mathbf{P}_m\mathbf{B}\mathbf{P}_m = \begin{pmatrix} \mathbf{B}_m & \mathbf{O}_{m,n-m} \\ \mathbf{O}_{n-m,m} & \mathbf{O}_{n-m,n-m} \end{pmatrix}$$

where $\mathbf{B}_m$ denotes the $m \times m$ principal submatrix of $\mathbf{B}$ possessing the eigenvalues $\mu_1 \geq \ldots \geq \mu_m$. The eigenvalues of a principal submatrix of the hermitian matrix $\mathbf{B}$ fulfill the inequalities

$$\mu_k \leq \sigma_k^2, \quad k = 1, \ldots, m \quad \text{(see [2], p. 186)}$$

and it follows

$$\operatorname{tr}\left[\mathbf{B}_m\right] = \sum_{k=1}^{m} \mu_k \leq \sum_{k=1}^{m} \sigma_k^2.$$

This inequality and the identity $\mathbf{Q}_m = \mathbf{I}_n - \mathbf{P}_m$ lead to

$$\begin{aligned}
\gamma_2(\mathcal{P}_m) &= \operatorname{tr}\left[\mathbf{Q}_m\mathbf{V}^*\mathbf{R}_{\mathbf{zz}}(0)\mathbf{V}\mathbf{Q}_m\right] = \operatorname{tr}\left[(\mathbf{I}_n - \mathbf{P}_m)\mathbf{B}(\mathbf{I}_n - \mathbf{P}_m)\right] \\
&= \operatorname{tr}\left[\mathbf{B}\right] - \operatorname{tr}\left[\mathbf{P}_m\mathbf{B}\right] - \operatorname{tr}\left[\mathbf{B}\mathbf{P}_m\right] + \operatorname{tr}\left[\mathbf{P}_m\mathbf{B}\mathbf{P}_m\right] \\
&= \sum_{k=1}^{n} \sigma_k^2 - \sum_{k=1}^{m} \mu_k - \sum_{k=1}^{m} \mu_k + \sum_{k=1}^{m} \mu_k \\
&\geq \sum_{k=1}^{n} \sigma_k^2 - \sum_{k=1}^{m} \sigma_k^2 = \sum_{k=m+1}^{n} \sigma_k^2.
\end{aligned}$$

The proof is complete. ■

Applying the preceding results for $\gamma_1(\mathcal{P}_m)$ and $\gamma_2(\mathcal{P}_m)$ we are now able to find γ-optimal projections.

Theorem 1
Let $\mathbf{z}(t,\omega)$ be a wide-sense stationary vector function with

- *mean $\mathbf{m_z} = \mathbf{E}\{\mathbf{z}(t)\}$*
- *and a non-singular variance matrix $\mathbf{R_{zz}}(0) = \mathbf{E}\{(\mathbf{z}(t)-\mathbf{m_z})(\mathbf{z}(t)-\mathbf{m_z})^*\}$ with eigenvalues $\sigma_1^2 \geq \ldots \geq \sigma_n^2 > 0$ and corresponding eigenvectors $\mathbf{u}_1,\ldots,\mathbf{u}_n$.*

A sequence $\left(\tilde{\mathcal{P}}_m\right)_{0\leq m\leq n}$ of projection operators defined by

$$\mathbf{v}_0 = \mathbf{m_z} \quad \textit{and} \quad \mathbf{v}_k = \mathbf{u}_k, \quad k = 1,\ldots,n,$$

is γ-optimal with respect to $\mathcal{M}$ and it holds

$$\gamma(\tilde{\mathcal{P}}_m) = \sum_{k=m+1}^{n} \sigma_k^2.$$

Proof: Using the orthonormality of $\mathbf{u}_1,\ldots,\mathbf{u}_n$ it follows $\left(\tilde{\mathcal{P}}_m\right)_{0\leq m\leq n} \in \mathcal{N} \subset \mathcal{M}$. First the γ-optimality of $\left(\tilde{\mathcal{P}}_m\right)_{0\leq m\leq n}$ with respect to the subset $\mathcal{N} \subset \mathcal{M}$ will be proved. We write

$$\begin{aligned}
\gamma(\mathcal{P}_m) &= \gamma(m,\mathbf{v}_0,\mathbf{V}) \\
\gamma_j(\mathcal{P}_m) &= \gamma_j(m,\mathbf{v}_0,\mathbf{V}),\ j=1,2,
\end{aligned}$$

in order to represent the dependence on $\mathbf{v}_0$ and $\mathbf{V} = (\mathbf{v}_1,\ldots,\mathbf{v}_n)$ of the sequence $(\mathcal{P}_m)_{0\leq m\leq n} \in \mathcal{N}$. Now it has to be proved that the inequality

$$\gamma(\tilde{\mathcal{P}}_m) = \gamma(m,\mathbf{m_z},\mathbf{U}) \leq \gamma(m,\mathbf{v}_0,\mathbf{V})$$

is true for $m = 0,\ldots,n$, $\forall \mathbf{v}_0 \in \mathbf{Z}$, $\forall\, \mathbf{V} = (\mathbf{v}_1,\ldots,\mathbf{v}_n), \mathbf{v}_k \in \mathbf{Z}$, $k = 1,\ldots,n$, with $(\mathbf{v}_k,\mathbf{v}_l) = \delta_{kl}$, $1 \leq k,l \leq n$ and the eigenvectors $\mathbf{U} = (\mathbf{u}_1,\ldots,\mathbf{u}_n)$ of $\mathbf{R_{zz}}(0)$.

We have seen that $\gamma_2(m,\mathbf{v}_0,\mathbf{V})$ does not depend on $\mathbf{v}_0$, i.e. $\gamma_2(m,\mathbf{v}_0,\mathbf{V}) = \gamma_2(m,\mathbf{V})$.

By means of relation (17) and Lemma 6 it can be deduced

$$\begin{aligned}
\gamma(m,\mathbf{v}_0,\mathbf{V}) &= \gamma_1(m,\mathbf{v}_0,\mathbf{V}) + \gamma_2(m,\mathbf{V}) \\
&\geq \gamma_2(m,\mathbf{V}) \geq \gamma_2(m,\mathbf{U}) \\
&= \gamma_2(m,\mathbf{U}) + \underbrace{\gamma_1(m,\mathbf{m_z},\mathbf{U})}_{=0} = \gamma(m,\mathbf{m_z},\mathbf{U}).
\end{aligned}$$

Then $\left(\tilde{\mathcal{P}}_m\right)_{0\leq m\leq n}$ is γ-optimal with respect to $\mathcal{N} \subset \mathcal{M}$. Applying Lemma 1 the γ-optimality with respect to the set $\mathcal{M}$ can be obtained.

Finally, the equation

$$\gamma(m, \mathbf{m_z}, \mathbf{U}) = \gamma(\tilde{\mathcal{P}}_m) = \gamma_2(\tilde{\mathcal{P}}_m) = \sum_{k=m+1}^{n} \sigma_k^2$$

follows from Lemma 6 and $\gamma_1(\tilde{\mathcal{P}}_m) = 0$. ∎

Corollary 1 ii) leads to the following assertion.

Corollary 2 Assuming the same conditions as in Theorem 1 the sequence $\left(\tilde{\mathcal{P}}_m\right)_{0\leq m\leq n}$ is ρ-optimal with respect to $\mathcal{N}$ and $\rho(\tilde{\mathcal{P}}_m)$ is given by

$$\rho(\tilde{\mathcal{P}}_m) = 1 - \frac{\gamma(\tilde{\mathcal{P}}_m)}{\gamma(\tilde{\mathcal{P}}_0)} = 1 - \frac{\sum\limits_{k=m+1}^{n} \sigma_k^2}{\sum\limits_{k=1}^{n} \sigma_k^2} = \frac{\sum\limits_{k=1}^{m} \sigma_k^2}{\sum\limits_{k=1}^{n} \sigma_k^2}.$$

Remark 4 In case of $\mathbf{Z} = \mathbb{R}^n$, we have real mean $\mathbf{m_z}$ and real variance matrix $\mathbf{R_{zz}}(0)$, which has real eigenvectors. Hence, the transformation matrix $\mathbf{V} = \mathbf{U}$ and the resulting γ- or ρ-optimal projection $\mathcal{P}_m\mathbf{z}$ is real, too.

Remark 5 The projection used in [5] which is based on the modal decomposition is not γ-optimal, since the transformation matrix used there is the eigenvector matrix of the system matrix of a linear vibration system, which is non-unitary in general.

In contrast to the γ-optimal projection the projection considered in [5] does not demand the knowledge of the variance matrix $\mathbf{R_{zz}}(0)$ (as well as its eigenvalues and eigenvectors) of the original system.

Remark 6 The assertion of Theorem 1 corresponds to a result of principal component analysis of multivariate statistics. There, a random vector $\mathbf{y}$ is considered which has values in a n-dimensional Euclidean space and a given mean $\mathbf{m_y}$ and variance matrix $\mathbf{R_{yy}}$ (e.g. empirical). A m-dimensional linear submanifold ($m < n$) which contains the "maximum of the variance" is found to be spaned by the eigenvectors belonging to the largest eigenvalues of $\mathbf{R_{yy}}$ and shifted by the mean $\mathbf{m_y}$.

References

[1] A.V. Gnedin. On mean-square epsilon dimension. *Math. Jap.*, 37(4):623–627, 1992.

[2] B.N. Parlett. *The symmetric eigenvalue problem.* Prentice Hall, Englewood Cliffs, 1980.

[3] M. Richter. *Approximation of Gaussian random elements and statistics.* Teubner, Leipzig, 1992.

[4] J. vom Scheidt, H.-J. Starkloff, and R. Wunderlich. Low-dimensional approximations of linear vibration systems. to appear.

[5] R. Wunderlich, J. Gruner, and J. vom Scheidt. Model reduction of random vibration systems. In M. Brøns, M.P. Bendsøe, and M.P. Sørensen, editors, *Progress in Industrial Mathematics at ECMI 96*, pages 382 – 389, Stuttgart, 1997. Teubner.