

## AN OPTIMAL LINEAR FILTER FOR RANDOM SIGNALS WITH REALISATIONS IN A SEPARABLE HILBERT SPACE

P. G. HOWLETT<sup>1</sup>, C. E. M. PEARCE<sup>2</sup> and A. P. TOROKHTI<sup>3</sup>

(Received 3 August, 1998; revised 3 July, 2000)

### Abstract

Let  $u$  be a random signal with realisations in an infinite-dimensional vector space  $X$  and  $v$  an associated observable random signal with realisations in a finite-dimensional subspace  $Y \subseteq X$ . We seek a pointwise-best estimate of  $u$  using a bounded linear filter on the observed data vector  $v$ . When  $x$  is a finite-dimensional Euclidean space and the covariance matrix for  $v$  is nonsingular, it is known that the best estimate  $\hat{u}$  of  $u$  is given by a standard matrix expression prescribing a linear mean-square filter. For the infinite-dimensional Hilbert space problem we show that the matrix expression must be replaced by an analogous but more general expression using bounded linear operators. The extension procedure depends directly on the theory of the Bochner integral and on the construction of appropriate Hilbert-Schmidt operators. An extended example is given.

### 1. Introduction

A common problem in engineering, applied mathematics and statistics is the estimation of a random signal  $u$  by measuring an associated observable random signal  $v$ . Realisations of  $u$  are often represented as elements in an infinite-dimensional vector space. For a general background to the theory of random signals and noise, we cite the classic text by Davenport and Root [2] and the book by Frank [4]. A more recent work by Haykin [6] is primarily concerned with adaptive filters. There are many other general references. Of particular relevance to this paper is the strong similarity between optimal filtering problems and problems of approximation of linear and nonlinear systems. Representation and approximation are the themes in a sequence of

<sup>1</sup>Centre for Industrial and Applicable Mathematics, University of South Australia.

<sup>2</sup>School of Applied Mathematics, The University of Adelaide, Adelaide, SA 5005, Australia; e-mail: cpearce@maths.adelaide.edu.au.

<sup>3</sup>Centre for Industrial and Applicable Mathematics, University of South Australia and School of Applied Mathematics, The University of Adelaide.

© Australian Mathematical Society 2003, Serial-fee code 1446-1811/03

papers by Sandberg [11–15] and the fundamental theory of constructive approximation for nonlinear systems is discussed by Torokhti and Howlett [17] and by Howlett and Torokhti [7, 8].

Let  $u$  be a random signal with realisations  $u(\omega) = x$  in an infinite-dimensional vector space  $X$  for each outcome  $\omega$  from the set  $\Omega$  of all possible outcomes. We seek an estimate of the vector  $u$  by observing an associated random vector  $v$  where the outcome  $v(\omega) = y$  of the observed data vector is realised as an element of some finite-dimensional subspace  $Y \subseteq X$ . Our goal is to find the best possible estimate of  $u$  using a linear estimator on  $v$ . Previous similar formulations of this problem and the corresponding solutions [9, 10, 16, 18] are justified only for estimation of random vectors with realisations in finite-dimensional vector spaces.

Suppose  $(\Omega, \Sigma, \mu)$  is a probability space and  $u \in L^2(\Omega, \mathbb{R}^m)$ ,  $v \in L^2(\Omega, \mathbb{R}^n)$  are random vectors with realisations  $u(\omega) \in \mathbb{R}^m$ ,  $v(\omega) \in \mathbb{R}^n$  in finite-dimensional Euclidean space and assume that the covariance matrices  $\mathcal{E}[uv^T] \in \mathbb{R}^{m \times n}$ ,  $\mathcal{E}[vv^T] \in \mathbb{R}^{n \times n}$  are known. The symbol  $\mathcal{E}$  denotes the expectation operator. If the matrix  $\mathcal{E}[vv^T]$  is nonsingular, then it is well-known [16] that the best linear mean-square estimate  $\hat{u}$  of the random vector  $u$  from the observed data vector  $v$  is given by a linear estimator of the form

$$\hat{u} = \mathcal{E}[uv^T]\mathcal{E}[vv^T]^{-1}v. \tag{1.1}$$

We emphasise that the formula (1.1) involves finite-dimensional matrices and as such is valid only for random vectors with realisations in finite-dimensional Euclidean space. Each matrix  $F \in \mathbb{R}^{m \times n}$  defines a bounded linear transformation  $\mathcal{M}_F \in \mathcal{L}(L^2(\Omega, \mathbb{R}^n), L^2(\Omega, \mathbb{R}^m))$  via the formula  $[\mathcal{M}_F v](\omega) = Fv(\omega)$  for each  $\omega \in \Omega$ . It is customary to write  $Fv$  rather than  $\mathcal{M}_F v$  so that we can also write  $[Fv](\omega) = Fv(\omega)$  for each  $\omega \in \Omega$ . We note that there are many bounded linear transformations from  $L^2(\Omega, \mathbb{R}^n)$  into  $L^2(\Omega, \mathbb{R}^m)$  that cannot be written in the form  $[\mathcal{M}_F v](\omega) = Fv(\omega)$  for each  $\omega \in \Omega$ .

Some generalisations of formula (1.1) have already been considered. In the case where  $v = u + \xi$  and  $\xi$  is an independent noise term, Kazakos [10] used a truncated singular-value decomposition of the matrix  $\mathcal{E}[uv^T]\mathcal{E}[vv^T]^{-1}$ . More recently, Yamashita and Ogawa [18] have shown that when the inverse matrix  $\mathcal{E}[vv^T]^{-1}$  does not exist the optimal estimate of Kazakos can be replaced by an optimal estimate using a truncated singular-value decomposition of the matrix  $\mathcal{E}[uv^T]\mathcal{E}[vv^T]^\dagger$ , where  $\mathcal{E}[vv^T]^\dagger$  is a generalised inverse matrix. Hua and Liu [9] argue that this truncated form of the matrix formula

$$\hat{u} = \mathcal{E}[uv^T]\mathcal{E}[vv^T]^\dagger v \tag{1.2}$$

is valid when no specific relationship between  $u$  and  $v$  is assumed. They claim that in this case the formula is easily established by finding the matrix  $F_0 \in \mathbb{R}^{m \times n}$  that

minimises

$$Q(F) = \mathcal{E}[\|u - Fv\|^2].$$

The argument proposed by Hua and Liu relies on the identity

$$Q(F) - Q(F_0) = \text{tr}\{(F - F_0)\mathcal{E}[vv^T](F - F_0)\}, \quad (1.3)$$

where  $\text{tr}\{A\}$  denotes the trace of the square matrix  $A$ . Once again the arguments are based on a standard matrix formulation and the random vectors are realised in some finite-dimensional Euclidean space. Similar arguments relating to minimum mean-square errors can be found in a recent paper by Zou and Lu [20], who consider the problem of optimal estimation in systems with linear associative memory.

In this paper we show that an extended form of the optimal estimator (1.2) can be applied to random vectors with realisations in a separable Hilbert space. To be specific, we show that the best linear mean-square estimate  $\hat{u}$  of the random vector  $u$  from the observed data vector  $v$  is given by a linear estimator of the form

$$\hat{u} = \mathcal{E}[J_u J_v^T] \mathcal{E}[J_v J_v^T]^\dagger v + K [I - (\mathcal{E}[J_v J_v^T])^{1/2} (\mathcal{E}[J_v J_v^T]^\dagger)^{1/2}] v, \quad (1.4)$$

where  $K \in \mathcal{L}(Y, X)$  is arbitrary. This extended form of the optimal estimator can be applied to random vectors with realisations in a separable Hilbert space. Although the principles of least squares approximation are well-known, this explicit formula for the optimal estimator appears to be new.

To define the expected value for a random vector with realisations in a separable Hilbert space it is necessary to use a vector-valued Bochner integral. Consequently we must also show that certain special random vectors are strongly  $\Sigma$ -measurable. We use a standard Banach space formulation of the Bochner integral [19] and apply this theory to Hilbert space. To ensure that the trace operator is well-defined the optimal estimator is constructed using Hilbert-Schmidt operators.

In the following section, we address the background mathematical machinery, while Section 3 presents a motivating generic example. Section 4 introduces some bounded linear mappings that are used to establish the main results, which are given in Section 5. We conclude in Section 6 with an analysis of the generic example.

In formulating practical problems it is often necessary to choose between two options, an approximate solution to the exact problem and the exact solution to an approximate problem. We have chosen the former course of action for the following reason. If we construct an optimal filter in infinite-dimensional space using bounded linear operators, then we can estimate the truncation errors in subsequent finite-dimensional approximations. If we truncate the signal before we attempt any estimation, it is difficult to see how we could make sensible judgements about the errors.

## 2. Preliminaries

In this section we outline a theoretical basis for the description of random vectors with realisations in Banach space. We follow the methods of Halmos [5], Dunford and Schwartz [3] and Yosida [19]. Although many of the results are natural extensions of results for real-valued random variables, the extensions require some care.

As in the real context,  $(\Omega, \Sigma, \mu)$  will denote a probability space. Here  $\Omega$  is the set of outcomes,  $\Sigma$  a (completed)  $\sigma$ -field of measurable subsets  $E \subset \Omega$  and  $\mu : \Sigma \mapsto [0, 1]$  an associated probability measure on  $\Sigma$ , so that  $\mu(\Omega) = 1$ . Each element  $\omega \in \Omega$  represents the outcome of an observation of experiment and each  $E \in \Sigma$  a set of outcomes, called an event. We say that the event  $E$  has occurred if  $\omega \in E$ .

Suppose  $E_j \in \Sigma$  ( $j = 1, \dots, n$ ) are mutually disjoint events and  $\Xi \in X$ , a Banach space, for  $j = 1, \dots, n$ . We may define a finitely-valued function  $u : \Omega \mapsto X$  by

$$u(\omega) = \sum_{j=1}^n \chi_j(\omega)\xi_j, \tag{2.1}$$

where the characteristic function  $\chi_j : \Omega \mapsto \{0, 1\}$  of the set  $E_j$  is given by  $\chi_j(\omega) = 0$  or 1 according as  $\omega \in E_j$  or  $\omega \notin E_j$ .

A function  $u : \Omega \mapsto X$  is said to be *strongly  $\Sigma$ -measurable* if there exists a sequence  $(u_n)_{n \geq 1}$  of finitely-valued functions  $u_n : \Omega \mapsto X$  such that  $\|u(\omega) - u_n(\omega)\| \rightarrow 0$  as  $n \rightarrow \infty$  for almost all  $\omega \in \Omega$ . The value  $u(\omega)$  of a strongly  $\Sigma$ -measurable function  $u$  is referred to as a random vector. When  $u$  is finitely-valued, the Bochner  $\mu$ -integral  $\mathcal{I}(u) \in X$  is prescribed by

$$\mathcal{I}\left(\sum_{j=1}^n \chi_j \xi_j\right) = \sum_{j=1}^n \mu(E_j)\xi_j.$$

When  $u$  is strongly  $\Sigma$ -measurable, we say that  $u$  is Bochner  $\mu$ -integrable if there exists a sequence  $(u_n)_{n \geq 1}$  of finitely-valued functions  $u_n : \Omega \mapsto X$  with  $\|u_n(\omega) - u(\omega)\| \rightarrow 0$  for  $\mu$ -almost all  $\omega \in \Omega$  in such a way that

$$\int_{\Omega} \|u_n(\omega) - u(\omega)\| \mu(d\omega) \rightarrow 0$$

as  $n \rightarrow \infty$ . In this case the Bochner  $\mu$ -integral is defined by

$$\int_{\Omega} u(\omega)\mu(d\omega) = \mathcal{I}(u),$$

where  $\mathcal{I}(u) \in X$  is the unique element with  $\|\mathcal{I}(u) - \mathcal{I}(u_n)\| \rightarrow 0$  as  $n \rightarrow \infty$ . In general, for each  $E \in \Sigma$ , we define

$$\int_E u(\omega)\mu(d\omega) = \int_{\Omega} \chi_E(\omega)u(\omega)\mu(d\omega),$$

where  $\chi$  is the characteristic function of the set  $E$ . The following general results can be found in Yosida [19].

**THEOREM 1.** *A strongly  $\Sigma$ -measurable function  $u : \Omega \mapsto X$  is Bochner  $\mu$ -integrable if and only if  $\|u\|$  is  $\mu$ -integrable.*

**COROLLARY 1.** *If  $\|u\|$  is  $\mu$ -integrable, then  $\|\int_{\Omega} u(\omega)\mu(d\omega)\| \leq \int_{\Omega} \|u(\omega)\|\mu(d\omega)$ .*

**COROLLARY 2.** *Let  $X$  and  $Y$  be Banach spaces and  $A \in \mathcal{L}(X, Y)$  a bounded linear map. If  $u : \Omega \mapsto X$  is Bochner  $\mu$ -integrable in  $X$  and if  $v = A[u]$ , then  $v : \Omega \mapsto Y$  is Bochner  $\mu$ -integrable in  $Y$  and  $\int_{\Omega} v(\omega)\mu(d\omega) = A [\int_{\Omega} u(\omega)\mu(d\omega)]$ .*

Suppose  $X$  and  $Y$  are Banach spaces. Let  $u : \Omega \mapsto X$  be a Bochner  $\mu$ -integrable random vector in  $X$ . The expected value of  $u$  is defined by

$$\mathcal{E}[u] = \int_{\Omega} u(\omega)\mu(d\omega)$$

and we note from Corollary 1 that  $\|\mathcal{E}[u]\| \leq \mathcal{E}[\|u\|]$ . When  $A \in \mathcal{L}(X, Y)$  is a bounded linear map, it follows from Corollary 2 that  $\mathcal{E}[A(u)] = A(\mathcal{E}[u])$ .

The theory of random vectors in Hilbert space is an extension of the corresponding theory in Banach space. Let  $X$  be a Hilbert space with scalar product  $\langle \cdot, \cdot \rangle$ . Of particular interest are properties relating to the scalar product, which are used directly in defining special operators for the optimal filter.

Let  $X$  be a Hilbert space with scalar product  $\langle \cdot, \cdot \rangle$  and let  $u : \Omega \mapsto X$  be the finitely-valued random vector given by (2.1). Since  $\|u(\omega)\|^2 = \sum_{j=1}^n \chi_j(\omega)\|\xi_j\|^2$ , it follows that if  $A \in \mathcal{L}(X, X)$  is a bounded linear map, then

$$\begin{aligned} & \left\langle \int_{\Omega} u(\omega)\mu(d\omega), \int_{\Omega} A[u(\omega)]\mu(d\omega) \right\rangle \\ &= \sum_{j=1}^n \sum_{k=1}^n \mu(E_j)\mu(E_k)\langle \xi_j, A[\xi_k] \rangle \\ &= \|A\| \sum_{j=1}^n \sum_{k=1}^n \mu(E_j)\mu(E_k)\|\xi_j\| \cdot \|\xi_k\| \leq \|A\| \sum_{j=1}^n [\mu(E_j)]^2 \|\xi_j\|^2 \\ &\leq \|A\| \sum_{j=1}^n \mu(E_j)\|\xi_j\|^2 = \|A\| \int_{\Omega} \|u(\omega)\|^2 \mu(d\omega), \end{aligned}$$

using the elementary inequalities

$$|\langle \xi_j, A[\xi_k] \rangle| \leq \|A\| \cdot \|\xi_j\| \cdot \|\xi_k\| \quad \text{and} \quad \|\xi_j\| \cdot \|\xi_k\| \leq [\|\xi_j\|^2 + \|\xi_k\|^2] / 2.$$

By taking appropriate limits, we can easily extend the above argument to establish the following general results, which are used to justify the construction of the optimal filter.

**THEOREM 2.** *If  $u : \Omega \mapsto X$  is strongly  $\Sigma$ -measurable and  $\|u\|^2$  is  $\mu$ -integrable, then  $u$  is Bochner  $\mu$ -integrable and for each bounded linear map  $A \in \mathcal{L}(X, X)$  we have*

$$\left\langle \int_{\Omega} u(\omega)\mu(d\omega), \int_{\Omega} A[u(\omega)]\mu(d\omega) \right\rangle \leq \|A\| \int_{\Omega} \|u(\omega)\|^2 \mu(d\omega).$$

**COROLLARY 3.** *If  $u : \Omega \mapsto X$  is strongly  $\Sigma$ -measurable and  $\|u\|^2$  is  $\mu$ -integrable, then*

$$\left\| \int_{\Omega} u(\omega)\mu(d\omega) \right\|^2 \leq \int_{\Omega} \|u(\omega)\|^2 \mu(d\omega).$$

The last two results can be expressed in terms of expected values. Let  $A \in \mathcal{L}(X, X)$  and let  $u : \Omega \mapsto X$  be a random vector. If  $\|u\|^2$  is  $\mu$ -integrable, then

$$\langle \mathcal{E}[u], \mathcal{E}[A(u)] \rangle \leq \|A\| \mathcal{E}[\|u\|^2]$$

and in particular  $\|\mathcal{E}[u]\|^2 \leq \mathcal{E}[\|u\|^2]$ .

We write  $L^2(\Omega, X)$  for the set of all strongly  $\Sigma$ -measurable functions  $u : \Omega \mapsto X$  with  $\mathcal{E}[\|u\|^2] < \infty$ .

To conclude this section, we review some basic structural results for bounded linear maps with finite-dimensional ranges on Hilbert space, which are used directly in our construction of the optimal estimates. We assume that  $X$  is a separable Hilbert space and  $Y \subseteq X$  is a finite-dimensional subspace with dimension  $n$ . The material on Hilbert-Schmidt operators follows Balakrishnan [1].

Let  $A \in \mathcal{L}(X, Y)$  and let  $\mathcal{R}(A) \subseteq Y$  denote the range space of  $A$ . Suppose  $\mathcal{R}(A)$  has dimension  $r \leq n$ . Let  $\mathcal{N}(A) \subseteq X$  denote the null space of  $A$ . The bounded linear map  $A^T \in \mathcal{L}(Y, X)$  is defined uniquely by the equation

$$\langle A^T(y), x \rangle = \langle y, A(x) \rangle$$

for each  $y \in Y$ . We write  $\mathcal{R}(A^T) \subseteq X$  for the range space of  $A^T$  and let  $\mathcal{N}(A^T) \subseteq Y$  denote the null space of  $A^T$ . Since  $\mathcal{R}(A)$  has dimension  $r \leq n$ , it follows that  $\mathcal{R}(A^T)$  also has dimension  $r$ .

Since  $\mathcal{R}(A)$  is finite-dimensional and therefore closed, it follows that  $Y = \mathcal{R}(A) \oplus \mathcal{N}(A^T)$  and that each  $y \in Y$  can be written uniquely in the form  $y = y_{\mathcal{R}} + y_{\mathcal{N}}$ , where  $y_{\mathcal{R}} \in \mathcal{R}(A)$  and  $y_{\mathcal{N}} \in \mathcal{N}(A^T)$  and where  $\langle y_{\mathcal{R}}, y_{\mathcal{N}} \rangle = 0$ . In a similar

fashion  $X = \mathcal{R}(A^T) \oplus \mathcal{N}(A)$  and each  $x \in X$  can be written uniquely in the form  $x = x_{\mathcal{R}} + x_{\mathcal{N}}$ , where  $x_{\mathcal{R}} \in \mathcal{R}(A^T)$ ,  $x_{\mathcal{N}} \in \mathcal{N}(A)$  and  $\langle x_{\mathcal{R}}, x_{\mathcal{N}} \rangle = 0$ .

The generalised inverse  $A^\dagger \in \mathcal{L}(Y, X)$  of  $A$  is defined as follows. Let  $y \in Y$ , put  $y = y_{\mathcal{R}} + y_{\mathcal{N}}$  and choose  $x \in X$  such that  $A(x) = y_{\mathcal{R}}$ . Write  $x = x_{\mathcal{R}} + x_{\mathcal{N}}$  and define

$$A^\dagger(y) = x_{\mathcal{R}}.$$

The bounded linear operators  $A^T A \in \mathcal{L}(X, X)$  and  $AA^T \in \mathcal{L}(Y, Y)$  are positive-definite and self-adjoint.

Since  $A^T A \in \mathcal{L}(X, X)$  and  $AA^T \in \mathcal{L}(Y, Y)$ , we can find orthonormal vectors  $\{e_i\}_{i=1}^r$  forming a basis for  $\mathcal{R}(A^T)$  and  $\{f_i\}_{i=1}^r$  forming a basis for  $\mathcal{R}(A)$  which satisfy

$$A^T A e_i = s_i^2 e_i \quad \text{and} \quad AA^T f_i = s_i^2 f_i$$

for each  $i = 1, 2, \dots, r$ . Here  $s_1 > s_2 > \dots > s_r > 0$  are real numbers and

$$f_i = \frac{1}{s_i} A e_i \quad \text{and} \quad e_i = \frac{1}{s_i} A^T f_i$$

for each  $i = 1, 2, \dots, r$ . Because  $X$  is separable, the orthonormal sets  $\{e_i\}_{i=1}^r$  and  $\{f_i\}_{i=1}^r$  can be extended to form complete orthonormal sets  $\{e_i\}_{i=1}^\infty$  and  $\{f_i\}_{i=1}^\infty$  in  $X$  and the bounded linear operators  $A$  and  $A^T$  are by definition Hilbert-Schmidt operators because

$$\|A\|_{HS}^2 := \sum_{i=1}^\infty \|A e_i\|^2 = \sum_{i=1}^r s_i^2 < \infty$$

and

$$\|A^T\|_{HS}^2 := \sum_{i=1}^\infty \|A^T f_i\|^2 = \sum_{i=1}^r s_i^2 < \infty.$$

We recall ([1, Definition 3.4.2]) the concept of a *nuclear* or *trace class* operator. Suppose  $\mathcal{H}_1, \mathcal{H}_2$  are separable Hilbert spaces with respective orthonormal bases  $\{g_i\}, \{h_i\}$ . A bounded linear operator  $A : \mathcal{H}_1 \mapsto \mathcal{H}_2$  is *nuclear* if  $\sum_{i=1}^\infty |\langle A g_i, h_i \rangle| < \infty$ .

It follows that the operators  $A^T A$  and  $AA^T$  are nuclear operators with finite traces given by

$$\text{tr}(A^T A) := \sum_{i=1}^\infty \langle A^T A e_i, e_i \rangle = \sum_{i=1}^r s_i^2 < \infty$$

and

$$\text{tr}(AA^T) := \sum_{i=1}^\infty \langle AA^T f_i, f_i \rangle = \sum_{i=1}^r s_i^2 < \infty.$$

### 3. A generic example

In this section we show that an elementary random function is equivalent to a random vector with realisations in an infinite-dimensional Hilbert space. This generic example will be used later in the paper to illustrate our proposed optimal linear estimator. It is well-known that a function  $x : [0, 1] \mapsto \mathbb{R}$  for which

$$\int_0^1 [x(t)]^2 dt < \infty$$

can be represented by a Fourier sine series  $x(t) = \sum_{k=1}^\infty x_k \sqrt{2} \sin k\pi t$ , or equivalently by an infinite-dimensional vector  $x = (x_1, x_2, x_3, \dots)^T$ , where  $\sum_{k=1}^\infty |x_k|^2 < \infty$ . In this case we say that the vector  $x \in X = l^2$ .

For the purpose of practical calculations with these functions it is necessary to use a suitable finite-dimensional approximation. Thus we write

$$x \approx (x_1, x_2, \dots, x_n, 0, \dots)^T$$

for some fixed value of  $n$ . We can generate random vectors with realisations in an infinite-dimensional Hilbert space by thinking of each coefficient  $x_k$  in the Fourier sine series as the realisation of a real-valued random variable. Let  $\Omega$  be the set of all possible outcomes and let  $u_k : \Omega \mapsto \mathbb{R}$  be a real-valued random variable. For each outcome  $\omega \in \Omega$  we have  $u_k(\omega) = x_k \in \mathbb{R}$  and we obtain a corresponding realisation

$$u(\omega, t) = \sum_{k=1}^\infty u_k(\omega) \sqrt{2} \sin k\pi t$$

of the random function  $u(t) = u(\cdot, t)$ , or equivalently a realisation

$$u(\omega) = (u_1(\omega), u_2(\omega), u_3(\omega), \dots)^T$$

of the infinite-dimensional random vector  $u = u(\cdot)$ . For the above realisations to be meaningful it is necessary that  $\sum_{k=1}^\infty |u_k(\omega)|^2 < \infty$  for almost all  $\omega \in \Omega$ .

### 4. Some special mappings

In this section we define some special bounded linear mappings that will be used to establish the main results.

For each  $x \in X$ , define a bounded linear map  $J_x \in \mathcal{L}(\mathbb{R}, X)$  by  $J_x(\alpha) = \alpha x$ . The adjoint mapping  $J_x^T \in \mathcal{L}(X, \mathbb{R})$  is given by  $J_x^T(y) = \langle x, y \rangle$ . Now  $J_x^T J_x \in \mathcal{L}(\mathbb{R}, \mathbb{R})$  satisfies

$$J_x^T J_x(\alpha) = J_x^T(\alpha x) = \langle x, \alpha x \rangle = \alpha \|x\|^2$$



and clearly  $\|J_x^T J_x\| = \|x\|^2$ . On the other hand,  $J_x J_x^T \in \mathcal{L}(X, X)$  is prescribed by

$$J_x J_x^T(y) = J_x(\langle x, y \rangle) = \langle x, y \rangle x$$

and hence, once again,  $\|J_x J_x^T\| = \|x\|^2$ .

Relative to the complete orthonormal set  $\{e_i\}$  in  $X$ , we have

$$\sum_{i=1}^{\infty} \langle J_x J_x^T(e_i), e_i \rangle = \sum_{i=1}^{\infty} \langle x, e_i \rangle^2 = \sum_{i=1}^{\infty} x_i^2 = \|x\|^2$$

and in  $\mathbb{R}$  we have  $J_x^T J_x(1) = \|x\|^2$ . Hence  $J_x^T J_x$  and  $J_x J_x^T$  are both nuclear operators with finite trace given by  $\text{tr}(J_x^T J_x) = \text{tr}(J_x J_x^T) = \|x\|^2$ . If  $A \in \mathcal{L}(X, Y)$  and  $B \in \mathcal{L}(Y, X)$  then  $J_{Ax} = A J_x$  and  $J_{By} = B J_y$  for all  $x \in X$  and all  $y \in Y$ .

Suppose that  $Y$  is a finite-dimensional subspace of the Hilbert space  $X$ . Let  $u : \Omega \mapsto X$  and  $v : \Omega \mapsto Y$  be random vectors with  $\mathcal{E}[\|u\|^2] < \infty$  and  $\mathcal{E}[\|v\|^2] < \infty$ .

LEMMA 1. *Suppose  $q \in Y$  is a fixed vector. Then the functions  $J_u J_v^T q : \Omega \mapsto X$  and  $J_v J_v^T q : \Omega \mapsto Y$  defined by*

$$[J_u J_v^T q](\omega) = \langle v(\omega), q \rangle u(\omega) \quad \text{and} \quad [J_v J_v^T q](\omega) = \langle v(\omega), q \rangle v(\omega)$$

for each  $\omega \in \Omega$  are strongly  $\Sigma$ -measurable with

$$\mathcal{E}[\|J_u J_v^T q\|] < \infty \quad \text{and} \quad \mathcal{E}[\|J_v J_v^T q\|] < \infty.$$

REMARK 1. The following proof makes extensive use of the material discussed in Section 2. In order to define the expectations  $\mathcal{E}[J_u J_v^T q]$  and  $\mathcal{E}[J_v J_v^T q]$ , it is necessary to use a Bochner integral. It is therefore also necessary to establish that the functions  $J_u J_v^T q$  and  $J_v J_v^T q$  are strongly  $\Sigma$ -measurable.

PROOF. Let  $\{u_n\}$  and  $\{v_n\}$  be sequences of finitely-valued random vectors with  $\|u_n(\omega) - u(\omega)\| \rightarrow 0$  and  $\|v_n(\omega) - v(\omega)\| \rightarrow 0$  as  $n \rightarrow \infty$  for almost all  $\omega \in \Omega$ . Then  $\{\langle v_n(\omega), q \rangle u_n(\omega)\}$  is a sequence of finitely-valued random vectors with

$$\begin{aligned} & \| \langle v_n(\omega), q \rangle u_n(\omega) - \langle v(\omega), q \rangle u(\omega) \| \\ & \leq \| \langle v_n(\omega) - v(\omega), q \rangle u_n(\omega) \| + \| \langle v(\omega), q \rangle [u_n(\omega) - u(\omega)] \| \\ & \leq \| v_n(\omega) - v(\omega) \| \cdot \| q \| \cdot \| u(\omega) \| + \| v(\omega) \| \cdot \| q \| \cdot \| u_n(\omega) - u(\omega) \| \rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$  for almost all  $\omega \in \Omega$ . Therefore  $J_u J_v^T q$  is strongly  $\Sigma$ -measurable. Similarly  $J_v J_v^T q$  is strongly  $\Sigma$ -measurable. It follows that

$$\begin{aligned} \| \mathcal{E}[J_u J_v^T q] \|^2 & \leq (\mathcal{E}[\| \langle v, q \rangle u \|])^2 \leq \| q \|^2 (\mathcal{E}[\| u \| \cdot \| v \|])^2 \\ & \leq \| q \|^2 \mathcal{E}[\| u \|^2] \mathcal{E}[\| v \|^2] < \infty \end{aligned}$$

and likewise that  $\| \mathcal{E}[J_v J_v^T q] \|^2 < \infty$ .

The operators  $\mathcal{E}[J_v J_u^T] \in \mathcal{L}(X, Y)$  and  $\mathcal{E}[J_u J_v^T] \in \mathcal{L}(Y, X)$  are defined by setting  $\mathcal{E}[J_v J_u^T]p = \mathcal{E}[\langle u, p \rangle v]$  and  $\mathcal{E}[J_u J_v^T]q = \mathcal{E}[\langle v, q \rangle u]$  for each  $p \in X$  and  $q \in Y$ . We have

$$\begin{aligned} \langle p, \mathcal{E}[J_u J_v^T]q \rangle &= \langle p, \mathcal{E}[\langle v, q \rangle u] \rangle = \mathcal{E}[\langle p, u \rangle \langle v, q \rangle] \\ &= \langle \mathcal{E}[\langle p, u \rangle v], q \rangle = \langle \mathcal{E}[J_v J_u^T]p, q \rangle \end{aligned}$$

and hence  $\mathcal{E}[J_u J_v^T]^T = \mathcal{E}[J_v J_u^T]$ . The self-adjoint operator  $\mathcal{E}[J_v J_v^T] \in \mathcal{L}(Y, Y)$  can be defined in a similar way.

### 5. The main results

We are now ready to return to the problem posed in the introduction. Suppose  $u \in L^2(\Omega, X)$  and  $v \in L^2(\Omega, Y)$ . For each  $F \in \mathcal{L}(Y, X)$ , the linear transformation  $\mathcal{M}_F \in \mathcal{L}(L^2(\Omega, Y), L^2(\Omega, X))$  is defined by  $[\mathcal{M}_F v](\omega) = Fv(\omega)$  for each  $\omega \in \Omega$ . Once again it is customary to write  $Fv$  rather than  $\mathcal{M}_F v$  since we then have  $[Fv](\omega) = Fv(\omega)$  for each  $\omega \in \Omega$ . We wish to solve the following problem.

**PROBLEM 1.** Let  $u \in L^2(\Omega, X)$  be an unknown random function and  $v \in L^2(\Omega, Y)$  an observable random function and suppose that  $\mathcal{E}[J_u J_v^T]$  and  $\mathcal{E}[J_v J_v^T]$  are known. Let  $Q : \mathcal{L}(Y, X) \mapsto \mathbb{R}$  be defined by  $Q(F) = \mathcal{E}[\|u - Fv\|^2]$  for each  $F \in \mathcal{L}(Y, X)$ . We wish to find  $F_0 \in \mathcal{L}(Y, X)$  such that  $Q(F_0) \leq Q(F)$  for all such  $F$ .

**REMARK 2.** In spite of an apparent close similarity between Problem 1 and the problem solved by Hua and Liu [9] we note that Problem 1 is formulated in a more general vector space. Hua and Liu assume the random functions are realised in finite dimensional vector spaces and the optimal estimator is constructed using matrices. Problem 1 allows realisation of the random functions in a separable Hilbert space and the optimal estimator is constructed using Hilbert-Schmidt operators.

**LEMMA 2.** *The null space  $\mathcal{N}(\mathcal{E}[J_v J_v^T])$  of the operator  $\mathcal{E}[J_v J_v^T] \in \mathcal{L}(Y, Y)$  is a subspace of the null space  $\mathcal{N}(\mathcal{E}[J_u J_v^T])$  of the operator  $\mathcal{E}[J_u J_v^T] \in \mathcal{L}(Y, X)$ .*

**PROOF.** Suppose  $q_{\mathcal{N}} \in \mathcal{N}(\mathcal{E}[J_v J_v^T])$ . Then  $\langle q_{\mathcal{N}}, \mathcal{E}[\langle v, q_{\mathcal{N}} \rangle v] \rangle = 0$  and hence  $\mathcal{E}[\langle v, q_{\mathcal{N}} \rangle^2] = 0$ . But for each  $p \in X$  we have

$$\begin{aligned} |\langle p, \mathcal{E}[J_u J_v^T]q_{\mathcal{N}} \rangle| &= |\mathcal{E}[\langle p, u \rangle \langle v, q_{\mathcal{N}} \rangle]| \\ &\leq (\mathcal{E}[\langle p, u \rangle^2])^{1/2} (\mathcal{E}[\langle v, q_{\mathcal{N}} \rangle^2])^{1/2} = 0. \end{aligned}$$

Therefore  $\mathcal{E}[J_u J_v^T]q_{\mathcal{N}} = 0$  and hence  $q_{\mathcal{N}} \in \mathcal{N}(\mathcal{E}[J_u J_v^T])$ .

COROLLARY 4.  $\mathcal{E}[J_u J_v^T] \mathcal{E}[J_v J_v^T]^\dagger \mathcal{E}[J_v J_v^T] = \mathcal{E}[J_u J_v^T]$ .

PROOF. If  $q_{\mathcal{N}} \in \mathcal{N}(\mathcal{E}[J_v J_v^T])$ , then  $\mathcal{E}[J_u J_v^T] \mathcal{E}[J_v J_v^T]^\dagger \mathcal{E}[J_v J_v^T] q_{\mathcal{N}} = 0$  and since the previous lemma shows that  $q_{\mathcal{N}} \in \mathcal{N}(\mathcal{E}[J_u J_v^T])$ , we have also  $\mathcal{E}[J_u J_v^T] q_{\mathcal{N}} = 0$ . On the other hand, if  $q_{\mathcal{Q}} \in \mathcal{R}(\mathcal{E}[J_v J_v^T]) = \mathcal{N}(\mathcal{E}[J_v J_v^T])^\perp$ , then there exists  $k \in Y$  such that  $q_{\mathcal{Q}} = \mathcal{E}[J_v J_v^T]^\dagger k$ . Hence

$$\begin{aligned} \mathcal{E}[J_u J_v^T] \mathcal{E}[J_v J_v^T]^\dagger \mathcal{E}[J_v J_v^T] q_{\mathcal{Q}} &= \mathcal{E}[J_u J_v^T] \mathcal{E}[J_v J_v^T]^\dagger \mathcal{E}[J_v J_v^T] \mathcal{E}[J_v J_v^T]^\dagger k \\ &= \mathcal{E}[J_u J_v^T] q_{\mathcal{Q}}. \end{aligned}$$

The desired result follows from the fact that any element of  $Y$  can be written in the form  $q = q_{\mathcal{N}} + q_{\mathcal{Q}}$ .

REMARK 3. If  $u \in L^2(\Omega, X)$  and  $x = u(\omega)$  for some  $\omega \in \Omega$  then the operators  $J_x J_x^T$  and  $J_x^T J_x$  are each nuclear operators and the trace is well defined. The trace is used to establish the next identity and the subsequent theorem. It is therefore necessary to know that the operators concerned are nuclear operators.

For each  $F \in \mathcal{L}(Y, X)$ , we have

$$Q(F) = \mathcal{E}[\|u - Fv\|^2] = \mathcal{E}[\text{tr}(J_{u-Fv} J_{u-Fv}^T)] = \mathcal{E}[\text{tr}((J_u - FJ_v)[J_u - FJ_v]^T)].$$

THEOREM 3. *The solution to Problem 1 is given by*

$$F_0 = \mathcal{E}[J_u J_v^T] \mathcal{E}[J_v J_v^T]^\dagger + K[I - (\mathcal{E}[J_v J_v^T])^{1/2} (\mathcal{E}[J_v J_v^T]^\dagger)^{1/2}],$$

where  $K \in \mathcal{L}(Y, X)$  is an arbitrary bounded linear operator. The corresponding uniquely defined minimum value of  $Q(F)$  is

$$Q(F_0) = \text{tr}\{\mathcal{E}[J_u J_u^T] - \mathcal{E}[J_u J_v^T] \mathcal{E}[J_v J_v^T]^\dagger \mathcal{E}[J_v J_u^T]\}.$$

PROOF. If we write  $\Delta Q(F) = Q(F) - Q(F_0)$ , then

$$\begin{aligned} \Delta Q(F) &= \text{tr}\{F \mathcal{E}[J_v J_v^T] F^T - \mathcal{E}[J_u J_v^T] F^T - F \mathcal{E}[J_v J_u^T] \\ &\quad + \mathcal{E}[J_u J_v^T] \mathcal{E}[J_v J_v^T]^\dagger \mathcal{E}[J_v J_u^T]\} \\ &= \text{tr}\{(F - \mathcal{E}[J_u J_v^T] \mathcal{E}[J_v J_v^T]^\dagger) \mathcal{E}[J_v J_v^T] (F - \mathcal{E}[J_u J_v^T] \mathcal{E}[J_v J_v^T]^\dagger)^T\} \\ &= \|(F - \mathcal{E}[J_u J_v^T] \mathcal{E}[J_v J_v^T]^\dagger) (\mathcal{E}[J_v J_v^T])^{1/2}\|_{HS}^2 \end{aligned}$$

where the norm is the Hilbert-Schmidt norm. Hence  $Q(F) - Q(F_0) \geq 0$  with equality if and only if  $F = F_0$ .

**COROLLARY 5.** *The best estimate  $\hat{u}$  of  $u$  using a bounded linear estimator on the observed signal  $v$  is given by*

$$\hat{u} = \mathcal{E}[J_u J_v^T] \mathcal{E}[J_v J_v^T]^\dagger v + K [I - (\mathcal{E}[J_v J_v^T])^{1/2} (\mathcal{E}[J_v J_v^T]^\dagger)^{1/2}] v,$$

where  $K \in \mathcal{L}(Y, X)$  is arbitrary. The minimum norm estimate is given by setting  $K = \mathbb{O}$ .

**REMARK 4.** The papers by Hua and Liu [9] and Yamashita and Ogawa [18] consider a finite-dimensional problem and use a truncated singular-value decomposition to minimise  $Q(F) = \mathcal{E}[\|u - Fv\|^2]$  subject to a restriction on the rank of the matrix  $F$ . We believe our result can be used to solve a similar problem for random vectors with realisations in a separable Hilbert space.

### 6. Analysis of the generic example

The generic example will be used to demonstrate the construction of an optimal estimator. In this example random functions are represented by infinite-dimensional random vectors. We show that the optimal filter can be represented using infinite-dimensional matrices with suitable limits on the size of the matrix coefficients. Let  $X = l^2$ . Suppose that we wish to estimate the random function

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \end{pmatrix} : \Omega \mapsto X$$

on the basis of an observed function

$$v(\omega) = \begin{pmatrix} v_1(\omega) \\ v_2(\omega) \\ v_3(\omega) \\ v_4(\omega) \\ 0 \\ \vdots \end{pmatrix}$$

with realisations in a four-dimensional subspace  $Y \subseteq X$ . We assume that  $v = Au$ , where

$$A = \begin{pmatrix} A_{11} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} \quad \text{and} \quad A_{11} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

and where  $\mathbb{O}$  is an infinite-dimensional zero submatrix. Therefore  $v_1 = u_1 + u_2$ ,  $v_2 = u_2 + u_3$ ,  $v_3 = u_3 + u_4$ ,  $v_4 = u_4 + u_1$  and  $v_k = 0$  for all  $k \geq 5$ .

To find the best estimate  $\hat{u}$  of  $u$  using a linear estimator on  $v$ , we need to define some special operators. For each  $u : \Omega \mapsto l^2$  and  $v : \Omega \mapsto l^2$  and each  $y \in Y$ , the functions  $J_u J_v^T y : \Omega \mapsto X$  and  $J_v J_v^T y : \Omega \mapsto Y$  are defined by

$$[J_u J_v^T y](\omega) = \langle v(\omega), y \rangle u(\omega) \quad \text{and} \quad [J_v J_v^T y](\omega) = \langle v(\omega), y \rangle v(\omega)$$

for each  $\omega \in \Omega$ . We suppose that the random variables  $u_k$  are pairwise independent with  $\mathcal{E}[u_k] = \rho_k$  and  $\mathcal{E}[(u_k - \rho_k)^2] = \sigma_k^2$ . In practice this could occur as a deterministic function with coefficients  $\{\rho_k\}$  and an additive noise term  $\xi_k = u_k - \rho_k$ .

We also suppose that  $|\rho_k| \leq R/k$  for some fixed constant  $R > 0$ . We can now calculate

$$\mathcal{E}[J_u J_v^T y] = \begin{pmatrix} \sigma_1^2(y_1 + y_4) \\ \sigma_2^2(y_1 + y_2) \\ \sigma_3^2(y_2 + y_3) \\ \sigma_4^2(y_3 + y_4) \\ 0 \\ \vdots \end{pmatrix} + \begin{pmatrix} \rho_1[(\rho_1 + \rho_2)y_1 + (\rho_2 + \rho_3)y_2 + (\rho_3 + \rho_4)y_3 + (\rho_4 + \rho_1)y_4] \\ \rho_2[(\rho_1 + \rho_2)y_1 + (\rho_2 + \rho_3)y_2 + (\rho_3 + \rho_4)y_3 + (\rho_4 + \rho_1)y_4] \\ \rho_3[(\rho_1 + \rho_2)y_1 + (\rho_2 + \rho_3)y_2 + (\rho_3 + \rho_4)y_3 + (\rho_4 + \rho_1)y_4] \\ \rho_4[(\rho_1 + \rho_2)y_1 + (\rho_2 + \rho_3)y_2 + (\rho_3 + \rho_4)y_3 + (\rho_4 + \rho_1)y_4] \\ \rho_5[(\rho_1 + \rho_2)y_1 + (\rho_2 + \rho_3)y_2 + (\rho_3 + \rho_4)y_3 + (\rho_4 + \rho_1)y_4] \\ \vdots \end{pmatrix}$$

and

$$\mathcal{E}[J_v J_v^T y] = \begin{pmatrix} (\sigma_1^2 + \sigma_2^2)y_1 + \sigma_2^2 y_2 + \sigma_1^2 y_4 \\ \sigma_2^2 y_1 + (\sigma_2^2 + \sigma_3^2)y_2 + \sigma_3^2 y_3 \\ \sigma_3^2 y_2 + (\sigma_3^2 + \sigma_4^2)y_3 + \sigma_4^2 y_4 \\ \sigma_1^2 y_1 + \sigma_4^2 y_3 + (\sigma_4^2 + \sigma_1^2)y_4 \\ 0 \\ \vdots \end{pmatrix} + \begin{pmatrix} (\rho_1 + \rho_2)[(\rho_1 + \rho_2)y_1 + (\rho_2 + \rho_3)y_2 + (\rho_3 + \rho_4)y_3 + (\rho_4 + \rho_1)y_4] \\ (\rho_2 + \rho_3)[(\rho_1 + \rho_2)y_1 + (\rho_2 + \rho_3)y_2 + (\rho_3 + \rho_4)y_3 + (\rho_4 + \rho_1)y_4] \\ (\rho_3 + \rho_4)[(\rho_1 + \rho_2)y_1 + (\rho_2 + \rho_3)y_2 + (\rho_3 + \rho_4)y_3 + (\rho_4 + \rho_1)y_4] \\ (\rho_4 + \rho_1)[(\rho_1 + \rho_2)y_1 + (\rho_2 + \rho_3)y_2 + (\rho_3 + \rho_4)y_3 + (\rho_4 + \rho_1)y_4] \\ 0 \\ \vdots \end{pmatrix}$$

for all  $y \in Y$ . This enables us to write down a matrix representation for each of these operators. Note that these representations are essentially infinite matrices with some limit on the size of the matrix coefficients. In this case the size of the coefficients is limited by the inequality

$$\sum_{i,j=1}^{\infty} (\rho_i \rho_j)^2 < \frac{R\pi^4}{36}.$$

If we define

$$S = \begin{pmatrix} S_{11} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix}, \quad \text{where } S_{11} = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \\ 0 & 0 & 0 & \sigma_4 \end{pmatrix} \quad \text{and } \rho = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \end{pmatrix},$$

then

$$\mathcal{E}[J_u J_v^T] = [SS^T + \rho\rho^T]A^T \quad \text{and} \quad \mathcal{E}[J_v J_v^T] = A[SS^T + \rho\rho^T]A^T.$$

We now show that the operator  $\mathcal{E}[J_v J_v^T]$  is not invertible and calculate the generalised inverse. Define an orthogonal transformation

$$U = \begin{pmatrix} U_{11} & \mathbb{O} \\ \mathbb{O} & I \end{pmatrix}, \quad \text{where } U_{11} = \begin{pmatrix} 1/2 & -1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & 1/2 & 1/2 \\ -1/2 & -1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \end{pmatrix}$$

and observe that

$$U\mathcal{E}[J_v J_v^T]U^T = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & \sum_{i=1}^4 \sigma_i^2 & \sigma_4^2 - \sigma_2^2 & \sigma_1^2 - \sigma_3^2 & 0 & \dots \\ 0 & \sigma_4^2 - \sigma_2^2 & \sigma_2^2 + \sigma_4^2 & 0 & 0 & \dots \\ 0 & \sigma_1^2 - \sigma_3^2 & 0 & \sigma_1^2 + \sigma_3^2 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} + \begin{pmatrix} 0 \\ \sum_{i=1}^4 \rho_i \\ \rho_4 - \rho_2 \\ \rho_1 - \rho_3 \\ 0 \\ \vdots \end{pmatrix} (0 \quad \sum_{i=1}^4 \rho_i \quad \rho_4 - \rho_2 \quad \rho_1 - \rho_3 \quad 0 \quad \dots).$$

Using an appropriate partition we can therefore write

$$U\mathcal{E}[J_v J_v^T]U^T = \begin{pmatrix} \mathbb{O} & \mathbb{O} & \mathbb{O} \\ \mathbb{O} & P & \mathbb{O} \\ \mathbb{O} & \mathbb{O} & \mathbb{O} \end{pmatrix},$$

where

$$P = \begin{pmatrix} \sum_{i=1}^4 \sigma_i^2 & \sigma_4^2 - \sigma_2^2 & \sigma_1^2 - \sigma_3^2 \\ \sigma_4^2 - \sigma_2^2 & \sigma_2^2 + \sigma_4^2 & 0 \\ \sigma_1^2 - \sigma_3^2 & 0 & \sigma_1^2 + \sigma_3^2 \end{pmatrix} \\ + \begin{pmatrix} \sum_{i=1}^4 \rho_i \\ \rho_4 - \rho_2 \\ \rho_1 - \rho_3 \end{pmatrix} \begin{pmatrix} \sum_{i=1}^4 \rho_i & \rho_4 - \rho_2 & \rho_1 - \rho_3 \end{pmatrix}.$$

Since

$$\begin{vmatrix} \sum_{i=1}^4 \sigma_i^2 & \sigma_4^2 - \sigma_2^2 & \sigma_1^2 - \sigma_3^2 \\ \sigma_4^2 - \sigma_2^2 & \sigma_2^2 + \sigma_4^2 & 0 \\ \sigma_1^2 - \sigma_3^2 & 0 & \sigma_1^2 + \sigma_3^2 \end{vmatrix} = \sum_{1 \leq i < j < k \leq 4} 4\sigma_i^2 \sigma_j^2 \sigma_k^2 > 0,$$

it follows that  $P^{-1}$  exists and

$$\mathcal{E}[J_v J_v^T]^\dagger = U \begin{pmatrix} \mathbb{O} & \mathbb{O} & \mathbb{O} \\ \mathbb{O} & P^{-1} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} & \mathbb{O} \end{pmatrix} U^T.$$

It has been shown that the minimum norm best estimate  $\hat{u}$  of  $u$  using a linear estimator on the observed function  $v$  is given by

$$\hat{u} = \mathcal{E}[J_u J_u^T] \mathcal{E}[J_v J_v^T]^\dagger v.$$

In this example we have seen that the estimator can be easily computed. In general the estimator involves an infinite-dimensional matrix and our implementation must necessarily be a truncation of the true optimal estimator. The approximation can be made as accurate as we please.

## References

- [1] A. V. Balakrishnan, *Applied functional analysis*, Applications of Mathematics 3 (Springer, New York, 1976).
- [2] W. B. Davenport and W. L. Root, *An introduction to the theory of random signals and noise* (McGraw-Hill, New York, 1958).

- [3] N. Dunford and J. T. Schwartz, *Linear operators. Part I. General theory* (Wiley, New York, 1988).
- [4] L. E. Franks, *Signal theory* (Prentice-Hall, Englewood Cliffs, N.J., 1966).
- [5] P. R. Halmos, *Measure theory*, University Series in Higher Mathematics, 12th printing (Van Nostrand, Princeton, 1968).
- [6] S. Haykin, *Adaptive filter theory* (Prentice-Hall, Englewood Cliffs, N.J., 1991).
- [7] P. G. Howlett and A. P. Torokhti, "A methodology for the constructive approximation of nonlinear operators defined on noncompact sets", *Numer. Funct. Anal. Optimiz.* **18** (1997) 343–365.
- [8] P. G. Howlett and A. P. Torokhti, "Weak interpolation and approximation of non-linear operators on the space  $\mathcal{C}([0, 1])$ ", *Numer. Funct. Anal. Optimiz.* **19** (1998) 1025–1043.
- [9] Y. Hua and W. Q. Liu, "Generalized Karhunen-Loève transform", *IEEE Signal Process. Lett.* **5** (1998) 141–142.
- [10] D. Kazakos, "Optimal constrained representation and filtering of signals", *Signal Process.* **5** (1983) 347–353.
- [11] I. W. Sandberg, "On approximation of linear functionals on  $L_p$  spaces", *IEEE Trans. Circuits Systems I Fund. Theory Appl.* **42** (1995) 402–404.
- [12] I. W. Sandberg, "Uniform approximation of multidimensional myopic maps", *IEEE Trans. Circuits Systems I Fund. Theory Appl.* **44** (1997) 477–485.
- [13] I. W. Sandberg, "Notes on uniform approximation of time-varying systems on finite time intervals", *IEEE Trans. Circuits Systems I Fund. Theory Appl.* **45** (1998) 863–864.
- [14] I. W. Sandberg, "A representation theorem for linear systems", *IEEE Trans. Circuits Systems I Fund. Theory Appl.* **45** (1998) 578–583.
- [15] I. W. Sandberg, "Separation conditions and approximation of continuous-time approximately finite memory systems", *IEEE Trans. Circuits Systems I Fund. Theory Appl.* **46** (1999) 820–826.
- [16] H. W. Sorenson, *Parameter estimation, principles and problems* (Marcel Dekker, New York, 1980).
- [17] A. P. Torokhti and P. G. Howlett, "On the constructive approximation of non-linear operators in the modelling of dynamical systems", *J. Austral. Math. Soc. Ser. B.* **39** (1997) 1–27.
- [18] Y. Yamashita and H. Ogawa, "Relative Karhunen-Loève transform", *IEEE Trans. Signal Process.* **44** (1996) 371–378.
- [19] K. Yosida, *Functional Analysis*, 5th ed. (Springer, New York, 1978).
- [20] L. H. Zou and J. Lu, "linear associative memories with optimal rejection to coloured input noise", *IEEE Trans. Circuits Systems II Analogue and Digital Signal Process.* **44** (1997) 990–1000.