Signal Processing Techniques for Integrated Navigation Systems

Luís Gomes
Instituto Superior Técnico (IST) and
Institute for Systems and Robotics (ISR),
Lisbon, Portugal
Email: lois_gomes@hotmail.com

***Abstract*— A non-iterative methodology for the interpolation and regularization of multidimensional sampled signals with missing data resorting to Principal Component Analysis is introduced. Based on unbiased estimators for the mean and covariance of signals, the Principal Component Analysis is performed and the signal is interpolated and regularized. The optimal solution is obtained from a weighted least mean square minimization problem, and upper and lower bounds are provided for the mean square interpolation error. This solution is a refinement to the previously introduced method in [19], where three extensions are exploited: i) mean substitution for covariance estimation, ii) Tikhonov regularization method and, iii) dynamic principal components selection. The resulting method will be applied to bathymetric data, acquired at sea in the passage between the islands of Faial and Pico, Azores. Based on the properties of the least squares solution, an estimate for the reconstruction error is proposed, allowing for the assessment on the performance of the overall method even in the case where the original signal is not known. The results obtained pave the way to the use of the proposed framework in a number of sensor fusion problems, in the presence of missing data. In control systems, there are a number of situations where we face the challenge of controlling a system from an incomplete signal obtained through sensors. The application of the interpolation based on PCA to Kalman filtering represents a starting point of improvement on the control over an unreliable communication channel. An exploratory empirical assessment was performed to determined the impact of such strategy on Kalman filtering. A number of adaptations to the method proposed in the first part of this work were conducted, which includes the development of recursive estimators for the mean and covariance in order to obtain a recursive framework of the reconstruction with PCA. This work opens the door to future analysis in greater depth and formality. For each situation were conducted performance assessment studies where conducted in several situations.**

***Index Terms*— Principal Components Analysis - Missing Data - Reconstruction - Regularization - Bathymetric Data - Kalman Filtering**

I. Introduction

The problem of interpolation of multidimensional sampled signals with missing data is central in a series of engineering problems. Autonomous robotic surveying [22], underwater positioning, remote sensing, digital communications (subject to bursts of destructive interferences), estimation and control in networked systems, and computer vision (when occlusions occurs) are a few of a multitude of examples where data is not available at uniform temporal/spatial rates.

The scientific community has been active for long time in solving interpolation and reconstruction problems, see [2], [4], [14], [33] and the references therein for an in-depth repository of available techniques. Iterative methods such as P-G algorithm [7], [20], [21] and the EM algorithm [24] are the most commonly used. However, the iterative characteristics of these methods, with the correspondent computational burden, the restricted domain of application to bandlimited signals, and the low convergence rates verified, preclude its use in a number of relevant applications.

Primarily motivated by a terrain based navigation problem for underwater autonomous robotic activities [22], [18], this paper extends previous work presented in [19], where a new methodology was proposed for the interpolation of signals with missing data, that departed from the aforementioned approaches. In this work a non-iterative methodology for the regularized interpolation of multidimensional sampled signals with missing data based on PCA is proposed. Resorting to unbiased estimators for the mean and covariance of multidimensional signals, corrupted by zero-mean noise, the Principal Component Analysis is straightforward to be computed. The signal interpolation is tackled in the components space, formulating a weighted least squares minimization problem with known optimal solution. Moreover, based on PCA properties corrected upper and lower bounds (relative to [19]) for the mean square interpolation error and the interval of validity of the proposed method are provided. Relative to the basic solution previously proposed in [19], three refinements are exploited:

i. mean substitution,
ii. Tikhonov regularization and,
iii. dynamic principal components selection.

It is important to remark that not only the interval of validity of the resulting methods are extended but also outperform the basic one.

It is nowadays common the widespread use of ASCs, ROVs, and AUVs in a number of missions at sea, see [23] and references therein. Those robotic vehicles carrying powerful computers, mass storage media, and state-of-the-art sensors and transducers, have endowed the scientists with tools that can collect massive quantities of data on relevant marine quantities.

There is a problem central to all information acquired during the survey missions: data is not available at uniform temporal and spatial rates. Other domains where this phenomena occurs are remote sensing, digital communications (subject to bursts of destructive interferences), estimation and control in networked systems, and computer vision (when occlusions occur), just to name a few. Thus, the problem of interpolation of multidimensional sampled signals with missing data is central in a series of applications.

The proposed method is applied to bathymetric data acquired during tests at sea. Based on estimators for the mean and covariance of signals corrupted by zero-mean noise, the PCA decomposition is performed and the signal is interpolated and regularized. The optimal solution is obtained from a regulated weighted least mean square minimization problem, and not only upper and lower bounds are provided, but also an estimate of the mean square interpolation error is introduced. This solution exploits the refinements found adequate for the problem at hand. The generalization of the proposed method to other multidimensional signals, such as magnetometers and gradiometers, is immediate.

The passage between the islands of Faial and Pico is probably the best-known shallow-water area in the Azores. A number of studies using mechanically scan and sidescan sonars [23] and multibeam sonars [16], [28] have already covered the main aspects of seabed morphology and character, resulting in a good knowledge of the general distribution of the different bottom types and features. The Espalamaca-Madalena ridge is one of those structures of great interest for the marine geologists and was selected to test the capabilities of the sensor fusion technique central to this work. The results obtained pave the way to the use of the proposed framework in a number of sensor fusion problems, in the presence of missing data. Ultimately, this method aims at overcoming the limitations faced today in marine data fusion problems.

Nowadays, control theory and communication channels have been come across in a wide range of situations. Let us consider the example of controlling a vehicle based on the position and velocity estimated from a network sensor, whose communications relays on a channel where measurements are lost or suffer such delay that are considered lost [12], [31]. Thus, a number of questions arises, namely, the minimum rate of arrival of observations that guaranties a stable control, and the performance degradation that occurs from the absence of measurements.

Previous studies concerning this problem were conducted in [8], [13], [17], [32]. Even so, in [26], we find a theoretical supported analysis of the existence of a limit of absent observations for which the state estimation error covariance converges, within the discrete-time framework. Additionally, upper and lower bounds of the expected error covariance were found resorting to the state estimation with Kalman filtering [1], [6]. This is a recursive estimator which is widely used in a number of applications.

Motivated by the high applicability of the problem and the requirement to obtain a better performance, the goal of the work developed is to assess the impact of including a reconstruction tool to mitigate the unreliable communication channel, as an effort to eliminate the setbacks produced by the discontinue arrival of measurements. This tool resorts to the interpolation and regularization of multidimensional signals based on PCA, which corresponds to the focus of the work in this dissertation. The Kalman filtering embraces the recursive framework, for which the PCA interpolator was not taken into consideration. Therefore, a large part of work is devoted to the required adaptations. A major modification is the introduction of recursive estimators for the mean and covariance of discrete stochastic signals.

II. Interpolation using PCA

The purpose of this section is to describe a methodology, supported on PCA, allowing the interpolation of multidimensional sampled signals with missing data, corrupted by zero mean noise, based on the following assumption, central to the rest of this work:

Assumption II.1:

The missing information on the multidimensional sampled signals are negligible and the available samples, in a number greater than to the selected number of principal components, are representative of the original signal.

A. Mean and Covariance Estimators with Missing Data

New estimators are proposed to account for missing data and an indicator index $\mathbf{l}$ is introduced, on which is applied the same stacking operation as in the multidimensional signals.

The index $\mathbf{l}_i(j), j = 1, \ldots, N$ is set to 1 if the j^{th} component of signal $\mathbf{x}_i$ is available and zero otherwise. In the latter, the component $\mathbf{x}(j)$ is also set to zero, without loss of generality. The estimators for the mean and covariance of multidimensional signals with missing data are now presented.

Lemma 1: Given a set of M signals $\mathbf{x}_i, i = 1, \ldots, M$, with associated indexes $\mathbf{l}_i$, the auxiliary vector of counters $\mathbf{c} = \sum_{i=1}^{M} \mathbf{l}_i$, and $\mathbf{C} = \sum_{i=1}^{M} \mathbf{l}_i \mathbf{l}_i^T$:

i) the estimator for the j^{th} component of the ensemble mean is

$$\mathbf{m}_x(j) = \frac{1}{\mathbf{c}(j)} \sum_{i=1}^{M} \mathbf{l}_i(j)\mathbf{x}_i(j), \quad j = 1, \ldots, N; \qquad (1)$$

ii) the estimator for the covariance element $\mathbf{R}_{xx}(j,k), j,k = 1, \ldots, N$, given $\mathbf{y}_i(j) = \mathbf{x}_i(j) - \mathbf{l}_i(j)\mathbf{m}_x(j)$, can be computed from

$$\mathbf{R}_{xx}(j,k) = \frac{1}{\mathbf{C}(j,k) - 1} \sum_{i=1}^{M} \mathbf{l}_i(j)\mathbf{l}_i(k)\mathbf{y}_i(j)\mathbf{y}_i(k). \quad (2)$$

Proof: It resorts only to basic signal processing tools and is omitted here (see [9] and [10] for details). ∎

B. Solution to the Interpolation Problem

To solve the interpolation problem central to this paper, consider that each signal $\mathbf{x}_i$ is obtained from the original signal $\mathbf{r}_i$ due to missing data, verifying the relation

$$\mathbf{x}_i = \mathbf{L}_i \mathbf{r}_i, \qquad (3)$$

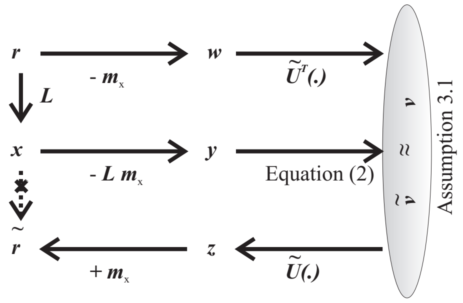


Fig. 1. Diagram describing the interpolation of sampling signals with missing data.

where $\mathbf{L}_i \in \mathcal{R}^{N\times N}$ is a diagonal matrix, filled with the indicator index $\mathbf{l}_i$. The interpolation operation is formulated as finding $\tilde{\mathbf{r}}_i$ that minimizes the weighted l_2 norm of the error. However, due to the existence of missed samples, it is only possible to compute the estimation error on the components of the signal which are known. Thus, the correct form of formulating the problem is to consider only the interpolation error for the available elements.

Lemma 2: Considering the original signal $\mathbf{r}_i$, from which there is only available a signal with samples indexed by $\mathbf{L}_i$, the optimal interpolated signal $\tilde{\mathbf{r}}_i$ (in the minimum error energy sense) can be obtained solving the weighted least mean square problem

$$\min_{\tilde{\mathbf{r}}_i \in \mathcal{R}^N} \|\mathbf{L}_i(\tilde{\mathbf{r}}_i - \mathbf{r}_i)\|^2_{2,\mathbf{W}}, \tag{4}$$

where the solution based on PCA is given by

$$\tilde{\mathbf{v}}_i = (\tilde{\mathbf{U}}^T \mathbf{L}_i \mathbf{W} \mathbf{L}_i \tilde{\mathbf{U}})^{-1} \tilde{\mathbf{U}}^T \mathbf{L}_i \mathbf{W} \mathbf{y}_i. \tag{5}$$

Proof: Given that a minimum energy estimation error problem can be formulated and solved as a weighted least mean square error optimization, (4) is written as

$$\min_{\tilde{\mathbf{v}}_i \in \mathcal{R}^n} \|\mathbf{L}_i(\tilde{\mathbf{U}}\tilde{\mathbf{v}}_i + \mathbf{m}_x) - \mathbf{L}_i\mathbf{r}_i\|^2_{2,\mathbf{W}} = $$
$$= \|\mathbf{L}_i\tilde{\mathbf{U}}\tilde{\mathbf{v}}_i + \mathbf{L}_i\mathbf{m}_x - \mathbf{L}_i\mathbf{r}_i\|^2_{2,\mathbf{W}},$$

resorting to the approximated PCA projection $\tilde{\mathbf{r}}_i = \tilde{\mathbf{U}}\tilde{\mathbf{v}}_i + \mathbf{m}_x$. Through the relations $\mathbf{x}_i = \mathbf{L}_i\mathbf{r}_i$ and $\mathbf{y}_i = \mathbf{x}_i - \mathbf{L}_i\mathbf{m}_x$, the following minimization is then obtained

$$\min_{\tilde{\mathbf{v}}_i \in \mathcal{R}^n} \|\mathbf{L}_i\tilde{\mathbf{U}}\tilde{\mathbf{v}}_i - \mathbf{y}_i\|^2_{2,\mathbf{W}}.$$

This is a weighted version of a linear least square problem, for which a well-known solution exists, resulting in (5), where the relations $\mathbf{L}\mathbf{L}^T = \mathbf{L}$ and $\mathbf{L}^T = \mathbf{L}$ were used. ■

From the previous assumption, the principal components can be computed with negligible degradation, and the signal can finally be reconstructed using the relation

$$\tilde{\mathbf{r}}_i = \tilde{\mathbf{U}}\tilde{\mathbf{v}}_i + \mathbf{m}_x. \tag{6}$$

The relations among the underlying signals are depicted in Fig. 1. Note that the aforementioned assumption can be interpreted as a change on the focus of the data from sample rates to the amount of information available.

According to optimal stochastic minimization techniques [11], the knowledge on the stochastic process characteristics allows for the optimal choice of the weight

$$\mathbf{W} = \mathbf{R}_{xx}^{-1}. \tag{7}$$

Nevertheless, the covariance matrix is estimated from an incomplete data set, which may lead to numerical problems when performing the inverse. Next, an approximated and more robust numerical solution is proposed.

Lemma 3: The inverse, $\mathbf{W}$, of a covariance matrix $\mathbf{R}_{xx}$, which characterizes a stochastic process with principal components $\tilde{\mathbf{U}}$, can be obtained approximately and with no explicit computation of the inverse by,

$$\tilde{\mathbf{W}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}^{-1}\tilde{\mathbf{U}}^T, \tag{8}$$

where $\tilde{\mathbf{\Lambda}} \in \mathcal{R}^{n\times n}$ is the diagonal matrix, whose k^{th} diagonal element is λ_k and contains the first n principal components.

Proof: The covariance matrix can be decomposed as

$$\mathbf{R}_{xx} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T,$$

where $\mathbf{\Lambda} \in \mathcal{R}^{N\times N}$ is the diagonal matrix, whose k^{th} diagonal element is λ_k and contains all N eigenvalues. An approximation of the covariance matrix is obtained using the approximated PCA [15], i.e.

$$\tilde{\mathbf{R}}_{xx} = \tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{U}}^T.$$

Consider now the hypothesis that the inverse of the approximated covariance matrix is obtained from $\tilde{\mathbf{R}}_{xx}^{-1} = \tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}^{-1}\tilde{\mathbf{U}}^T$. If the hypothesis corresponds in fact to the approximated inverse of the covariance matrix, the relation $\mathbf{R}_{xx}\tilde{\mathbf{R}}_{xx}^{-1} = \tilde{\mathbf{I}}$, where $\tilde{\mathbf{I}} \in \mathcal{R}^{N\times N}$ is approximately the identity matrix, must be verified by the following equation,

$$\mathbf{R}_{xx}\tilde{\mathbf{R}}_{xx}^{-1} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T\tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}^{-1}\tilde{\mathbf{U}}^T.$$

Considering algebraic properties and the orthogonality of both $\mathbf{U}$ and $\tilde{\mathbf{U}}$, the above expression is equivalent to

$$\mathbf{R}_{xx}\tilde{\mathbf{R}}_{xx}^{-1} = \tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}^{-1}\tilde{\mathbf{U}}^T = \tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Lambda}}^{-1}\tilde{\mathbf{U}}^T = \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T = \tilde{\mathbf{I}}.$$

Therefore is possible to conclude that,

$$\tilde{\mathbf{W}} = \tilde{\mathbf{R}}_{xx}^{-1} = \tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}^{-1}\tilde{\mathbf{U}}^T.$$

■

It is important to remark that the approximation is more accurate as the number of components increases. In the limit case where all principal components are used, $\mathbf{U}\mathbf{U}^T = \mathbf{I}$, meaning that $\tilde{\mathbf{W}} = \mathbf{W}$.

A complexity analysis to the proposed methods previously introduced revealed that the underlying complexity is $O(\eta J N^2 M)$. This is a consequence of the application of (5) to the unknown samples, therefore it depends on the amount of missing samples. Note that the matrix $\tilde{\mathbf{U}}^T\mathbf{L}_i\mathbf{W}\mathbf{L}_i\tilde{\mathbf{U}}$ to be inverted has dimension $n \times n$, presenting reduced computational complexity, given the choice of $n << N$. Interestingly enough, this result can be interpreted as a generalization of the classical Yen interpolator [33] and the minimax-optimal interpolators [4].

C. Lower and Upper Bounds

In order to evaluate the quality of the signal's interpolation, the variance of the interpolation error per signal sample, σ^2, is defined as,

$$\sigma^2 \equiv \frac{E[\|\tilde{\mathbf{r}} - \mathbf{r}\|]}{N-1}. \tag{9}$$

Scaling the bounds to a per sample basis, leads to the result

$$\eta \frac{\sum_{i=n+1}^{N} \lambda_i}{N-1} \leq \sigma^2 \leq \eta \frac{\sum_{i=1}^{N} \lambda_i}{N-1}. \tag{10}$$

A simple validation on the bounds for certain values of η can be made. Consider the extreme case where $\eta = 0$, i.e. there is no missing samples. The bounds for this case are both null, which is correct because no interpolation error is present and consequently, a null value for σ^2 is obtained. Now consider the case $\eta \approx 1$, which means that almost no samples of the signal are available and the interpolation will correspond to the signal's variance as stated in the upper bound for high levels of η.

Assumption II.1 can be interpreted as providing conditions when the interpolation is well posed or when the corresponding numerical tools can be applied. The number of samples available are required to be greater than to the selected number of principal components, i.e. $N(1-\eta) > n$. This leads to the following validity interval deduced from Assumption II.1,

$$0 \leq \eta < \frac{N-n}{N}. \tag{11}$$

Interestingly enough, no limitation on the amount of missing data was found for the application of the method.

III. Extensions to the Interpolation Solutions

In this section, refinements to the solutions proposed in Section II and in [19] are presented.

A. Mean Substitution Method

When dealing with the estimation of covariance based on missing data, several method are available [9]. An important group of such methods corresponds to the techniques that represents no extra computational complexity. A classical technique is the mean substitution method. The missing sample on the j^{th} component of the i^{th} variable is replaced by the corresponding component of the mean, i.e. the missing value $\mathbf{x}_i(j)$ is filled with the value $\mathbf{m}_x(j)$. Although originally the data set has missing samples, due to the replacement of the missing samples by the mean, the estimator which do not account for missing data are now applicable.

B. Dynamic Principal Components Selection

Consider now the case when Assumption II.1 is not verified, i.e. the available samples are less than or equal to the assigned number of principal components n. Under this situation, the solution to the minimization problem is ill-conditioned resulting a violation on the validity interval given by (11). An alternative approach is suggested next, to be applied to those cases. The number of components used for the computation of the minimizing solution is set to the nearest integer below the current number of available samples.

As a result of this procedure, the proposed reconstruction algorithm is extended to any amount of missing samples. Assumption II.1. is always verified, given the adjustment on the principal components used relative to the existing information. Note that the lower and upper bounds in (10) remain valid throughout the whole interval $\eta \in [0, 1[$.

C. Tikhonov Regularization

A commonly used technique is the Tikhonov regularization, for which a well-known solution exists [29]. With the purpose of ensuring a suitable reconstruction of the signal, it is desired a smooth transition between the available and the recovered samples. To satisfy this requirement, a regularization term can be added to the reformulated minimization problem. The first order difference matrix $\mathbf{D} =\in \mathcal{R}^{(N-1)\times N}$ and the auxiliary matrix $\overline{\mathbf{L}}_i \in \mathcal{R}^{N\times N}$, which is a diagonal matrix filled with the complementary values of the indicator index $\mathbf{l}_i$, i.e.

$$\overline{\mathbf{L}}_i = \mathbf{I} - \mathbf{L}_i, \tag{12}$$

are considered.

Lemma 4: Considering the original signal $\mathbf{r}_i$, from which there is only available a signal with samples indexed by $\mathbf{L}_i$, the optimal interpolated and regularized signal $\tilde{\mathbf{r}}_i$, given the auxiliary matrices $\mathbf{D} \in \mathcal{R}^{(N-1)\times N}$ and $\overline{\mathbf{L}}_i \in \mathcal{R}^{N\times N}$, can be obtained solving the weighted least mean square problem with two terms expressed as

$$\min_{\tilde{\mathbf{v}}_i \in \mathcal{R}^n} \|\mathbf{L}_i(\tilde{\mathbf{r}}_i - \mathbf{r}_i)\|_{2,\mathbf{W}}^2 + \|\alpha \mathbf{D}(\mathbf{L}_i \mathbf{r}_i + \overline{\mathbf{L}}_i \tilde{\mathbf{r}}_i)\|_2^2, \tag{13}$$

with the solution that can be obtained resorting also to the PCA decomposition as

$$\tilde{\mathbf{v}}_i = (\tilde{\mathbf{U}}^T \mathbf{L}_i \mathbf{W} \mathbf{L}_i \tilde{\mathbf{U}} + \alpha^2 \mathbf{\Gamma}^T \mathbf{\Gamma})^{-1} (\tilde{\mathbf{U}}^T \mathbf{L}_i \mathbf{W} \mathbf{y}_i - \alpha^2 \mathbf{\Gamma}^T \mathbf{\Delta}), \tag{14}$$

where $\mathbf{\Gamma} \equiv \mathbf{D}\overline{\mathbf{L}}_i \tilde{\mathbf{U}}$ is the regularization matrix, $\mathbf{\Delta} \equiv \mathbf{D}(\mathbf{y}_i + \mathbf{m}_x)$, and α is a regularization parameter.

Proof: The minimization problem in (13), including the specified regularization term, can be written as in Lemma 2,

$$\min_{\tilde{\mathbf{v}}_i \in \mathcal{R}^n} \|\mathbf{L}_i \tilde{\mathbf{U}} \tilde{\mathbf{v}}_i - \mathbf{y}_i\|_{2,W}^2 + \|\alpha \mathbf{D}(\mathbf{L}_i \mathbf{r}_i + \overline{\mathbf{L}}_i \tilde{\mathbf{r}}_i)\|_2^2,$$

which can be rewritten, using the relation $\tilde{\mathbf{r}}_i = \tilde{\mathbf{U}}\tilde{\mathbf{v}}_i + \mathbf{m}_x$, as

$$\min_{\tilde{\mathbf{v}}_i \in \mathcal{R}^n} \|\mathbf{L}_i \tilde{\mathbf{U}} \tilde{\mathbf{v}}_i - \mathbf{y}_i\|_{2,W}^2 + \|\alpha \mathbf{D}\overline{\mathbf{L}}_i \tilde{\mathbf{U}} \tilde{\mathbf{v}}_i + \alpha \mathbf{D}(\mathbf{L}_i \mathbf{r}_i + \overline{\mathbf{L}}_i \mathbf{m}_x)\|_2^2.$$

Considering that $\mathbf{L}_i \mathbf{r}_i + \overline{\mathbf{L}}_i \mathbf{m}_x = \mathbf{y}_i + \mathbf{m}_x$ and the definitions of $\mathbf{\Gamma}$ and $\mathbf{\Delta}$ above, this minimization problem results in the compact form

$$\min_{\tilde{\mathbf{v}}_i \in \mathcal{R}^n} \left\| \begin{pmatrix} \mathbf{L}_i \tilde{\mathbf{U}} \\ \alpha \mathbf{\Gamma} \end{pmatrix} \tilde{\mathbf{v}}_i - \begin{pmatrix} \mathbf{y}_i \\ -\alpha \mathbf{\Delta} \end{pmatrix} \right\|_{2,W}^2.$$

Solving the previous least mean square problem, leads to

$$\tilde{\mathbf{v}}_i = \left[\begin{pmatrix} \mathbf{L}_i \tilde{\mathbf{U}} \\ \alpha \mathbf{\Gamma} \end{pmatrix}^T \begin{pmatrix} \mathbf{L}_i \tilde{\mathbf{U}} \\ \alpha \mathbf{\Gamma} \end{pmatrix} \right]^{-1} \begin{pmatrix} \mathbf{L}_i \tilde{\mathbf{U}} \\ \alpha \mathbf{\Gamma} \end{pmatrix}^T \begin{pmatrix} \mathbf{y}_i \\ -\alpha \mathbf{\Delta} \end{pmatrix}.$$

In order to accomplish an expression in a standard form, the following statements are necessary,

$$\begin{pmatrix} \mathbf{L}_i\tilde{\mathbf{U}} \\ \alpha\mathbf{\Gamma} \end{pmatrix}^T = \begin{pmatrix} \tilde{\mathbf{U}}^T\mathbf{L}_i & \alpha\mathbf{\Gamma}^T \end{pmatrix},$$

$$\begin{pmatrix} \mathbf{L}_i\tilde{\mathbf{U}} \\ \alpha\mathbf{\Gamma} \end{pmatrix}^T \begin{pmatrix} \mathbf{L}_i\tilde{\mathbf{U}} \\ \alpha\mathbf{\Gamma} \end{pmatrix} = \tilde{\mathbf{U}}^T\mathbf{L}_i\mathbf{L}_i\tilde{\mathbf{U}} + \alpha^2\mathbf{\Gamma}^T\mathbf{\Gamma},$$

$$\text{and } \begin{pmatrix} \mathbf{L}_i\tilde{\mathbf{U}} \\ \alpha\mathbf{\Gamma} \end{pmatrix}^T \begin{pmatrix} \mathbf{y}_i \\ -\alpha\mathbf{\Delta} \end{pmatrix} = \tilde{\mathbf{U}}^T\mathbf{L}_i\mathbf{y}_i - \alpha^2\mathbf{\Gamma}^T\mathbf{\Delta}.$$

When applied to the presented solution, it follows,

$$\tilde{\mathbf{v}}_i = (\tilde{\mathbf{U}}^T\mathbf{L}_i\mathbf{L}_i\tilde{\mathbf{U}} + \alpha^2\mathbf{\Gamma}^T\mathbf{\Gamma})^{-1}(\tilde{\mathbf{U}}^T\mathbf{L}_i\mathbf{y}_i - \alpha^2\mathbf{\Gamma}^T\mathbf{\Delta}),$$

and since is considered the weighted least mean square problem, the solution is presented in (14). ■

The regularization parameter acts as a scaling factor involving the least square term and the regularization term of the minimization problem. For $\alpha = 0$, equation (14) reduces to the unregulated least squares solution presented in (5). A number of advantages associated to the application of a regularization technique can be delineated;

i) more adequate results are obtained, accordingly to the choice of the regularization term, which privileges suitable solutions;
ii) the unregulated solution of equation (5) may result in an amplification of the corresponding interpolation error, in case of severe lack of samples, leading to an inaccurate result;
iii) there is no significative increment on the computational complexity, as the matrix $(\tilde{\mathbf{U}}^T\mathbf{L}_i\mathbf{W}\mathbf{L}_i\tilde{\mathbf{U}} + \alpha^2\mathbf{\Gamma}^T\mathbf{\Gamma})$ to be inverted, preserves the dimension of the unregulated solution.

IV. Error Estimate for PCA Interpolator

When applying the interpolation method to a signal where samples are missing, there is no possibility to determine the interpolation error without knowing the original signal. However, in what follows, a technique is introduced that indirectly computes an estimate of the variance of the interpolation error for each estimated sample, as defined in (9).

A. Error Estimate Computation

Central to the work developed in this chapter are the following lemmas for the interpolation error estimation with PCA. The results are presented for the regulated solutions from Lemma 4, respectively.

Lemma 5: Consider the original signal $\mathbf{r}_i$ and reconstructed signal $\tilde{\mathbf{r}}_i$, obtained as the solution to the Tikhonov regulated weighted least mean squares problem, based on PCA. For the interpolation error $\mathbf{e}_i = \tilde{\mathbf{r}}_i - \mathbf{r}_i$, the error covariance can be computed, without explicit information about the original signal, as

$$\tilde{\mathbf{R}}_{\mathbf{e}_i\mathbf{e}_i} = (\mathbf{P}_i - \mathbf{I})\mathbf{R}_{xx}(\mathbf{P}_i - \mathbf{I})^T + \mathbf{Q}_i\mathbf{Q}_i^T, \quad (15)$$

where

$$\begin{aligned} \mathbf{P}_i &= \tilde{\mathbf{U}}\mathbf{V}_i(\tilde{\mathbf{U}}^T\mathbf{L}_i\mathbf{W}\mathbf{L}_i - \alpha^2\mathbf{\Gamma}^T\mathbf{D}\mathbf{L}_i), & (16)\\ \mathbf{Q}_i &= \alpha^2\tilde{\mathbf{U}}\mathbf{V}_i\mathbf{\Gamma}^T\mathbf{D}\mathbf{m}_x, & (17)\\ \textit{and}\quad \mathbf{V}_i &= (\tilde{\mathbf{U}}^T\mathbf{L}_i\mathbf{W}\mathbf{L}_i\tilde{\mathbf{U}} + \alpha^2\mathbf{\Gamma}^T\mathbf{\Gamma})^{-1}, & (18)\end{aligned}$$

for the sake of compactness.

Proof: From the relations depicted in Fig. 1 it follows that:

$$\mathbf{e}_i = \tilde{\mathbf{r}}_i - \mathbf{r}_i = \tilde{\mathbf{U}}\tilde{\mathbf{v}}_i + \mathbf{m}_x - \mathbf{w}_i - \mathbf{m}_x = \tilde{\mathbf{U}}\tilde{\mathbf{v}}_i - \mathbf{w}_i.$$

Given the interpolation problem solution from (14) and $\mathbf{y}_i = \mathbf{L}_i\mathbf{w}_i$ leads to:

$$\mathbf{e}_i = \tilde{\mathbf{U}}\mathbf{V}_i(\tilde{\mathbf{U}}^T\mathbf{L}_i\mathbf{W}\mathbf{L}_i\mathbf{w}_i - \alpha^2\mathbf{\Gamma}^T\mathbf{\Delta}) - \mathbf{w}_i.$$

Using the relation $\mathbf{\Delta} = \mathbf{D}(\mathbf{y}_i + \mathbf{m}_x)$, and organizing the previous equation, is obtained,

$$\begin{aligned}\mathbf{e}_i = [&\tilde{\mathbf{U}}\mathbf{V}_i(\tilde{\mathbf{U}}^T\mathbf{L}_i\mathbf{W}\mathbf{L}_i - \alpha^2\mathbf{\Gamma}^T\mathbf{D}\mathbf{L}_i) - \mathbf{I}]\mathbf{w}_i - \\ &- \tilde{\mathbf{U}}\mathbf{V}_i\alpha^2\mathbf{\Gamma}^T\mathbf{D}\mathbf{m}_x,\end{aligned}$$

which can be written in a simplified form as

$$\mathbf{e}_i = (\mathbf{P}_i - \mathbf{I})\mathbf{w}_i - \mathbf{Q}_i.$$

The covariance of $\mathbf{e}_i$, $\mathbf{R}_{\mathbf{e}_i\mathbf{e}_i}$ (note that $E\{\mathbf{e}_i\} = 0$), is computed through

$$\begin{aligned}\mathbf{R}_{\mathbf{e}_i\mathbf{e}_i} &= E\{\mathbf{e}_i\mathbf{e}_i^T\} = \\ &= E\{[(\mathbf{P}_i - \mathbf{I})\mathbf{w}_i - \mathbf{Q}_i][\mathbf{w}_i^T(\mathbf{P}_i - \mathbf{I})^T - \mathbf{Q}_i^T]\},\end{aligned}$$

using the distribution property, the fact that both $\mathbf{Q}_i$ and $\mathbf{P}_i$ are deterministic and because $E\{\mathbf{w}_i\} = E\{\mathbf{r}_i - \mathbf{m}_x\} = E\{\mathbf{r}_i\} - \mathbf{m}_x = 0$, it follows

$$\begin{aligned}\mathbf{R}_{\mathbf{e}_i\mathbf{e}_i} &= (\mathbf{P}_i - \mathbf{I})E\{\mathbf{w}_i\mathbf{w}_i^T\}(\mathbf{P}_i - \mathbf{I})^T + E\{\mathbf{Q}_i\mathbf{Q}_i^T\} - \\ &\quad -(\mathbf{P}_i - \mathbf{I})E\{\mathbf{w}_i\}\mathbf{Q}_i^T - \mathbf{Q}_iE\{\mathbf{w}_i^T\}(\mathbf{P}_i - \mathbf{I})^T \\ &= (\mathbf{P}_i - \mathbf{I})E\{\mathbf{w}_i\mathbf{w}_i^T\}(\mathbf{P}_i - \mathbf{I})^T + \mathbf{Q}_i\mathbf{Q}_i^T.\end{aligned}$$

For the missing data problem, it is not possible to compute the covariance of the original signal. However, there are efficient estimators available, already introduced, that allow the computation of an estimate of the covariance of the original signal. Based on this assumption, $E\{\mathbf{w}_i\mathbf{w}_i^T\} \simeq \mathbf{R}_{xx}$, with the result

$$\tilde{\mathbf{R}}_{\mathbf{e}_i\mathbf{e}_i} = (\mathbf{H}_i - \mathbf{I})\mathbf{R}_{xx}(\mathbf{H}_i - \mathbf{I})^T + \mathbf{B}_i\mathbf{B}_i^T.$$

■

Finally, the variance of the interpolation error of the j^{th} interpolated sample is determined resorting to the j^{th} diagonal element of $\mathbf{R}_{\mathbf{e}_i\mathbf{e}_i}$. Thus, is guarantied that no information about the original signal is required, due to the fact that all information is included indirectly by $\mathbf{R}_{xx}$. Despite this, the error estimation covariance is dependent of the index i, which departs to the fact that the interpolation error is influenced by the missing data sequence, $\mathbf{L}_i$, that for obvious reasons varies in function of i.

V. Performance Assessment and Results

After introducing the framework for interpolation based on PCA, a performance evaluation is conducted. In Fig. 2 is validated the estimation of the interpolation error based on an $1D$ audio signal and 20 Monte Carlo experiments, while in Fig. 3 a comparison is performed for the $1D$ interpolation with the Papoulis-Gercheberg algorithm and the averaging method. The same procedure is replied to a $2D$ signal, in this case a 8-bit gray image, in Fig. 4 and Fig. 5.

From the survey mission to the Espalamaca ridge, the obtained data was processed so that the PCA interpolation could be possible. Thus, resulting in the image depicted in Fig. 6. After the reconstruction based on PCA, the data is shown is Fig. 7. A comparison of the after and before reconstruction is presented in Fig. 8 where both situations are integrated in the same plot.

Finally, as an additional example, is evaluated the impact of performing Kalman filtering in case of intermittent observations. A number of elements were added to the proposed interpolator, namely, FILO stacking and recursive estimators for the mean and covariance. A statistical analysis is depicted for both stable and unstable systems in Fig. 9 and Fig. 10, respectively.

VI. Conclusions

A new methodology to interpolate and regularize sampled signals with missing data is presented, supported on estimates from two efficient estimators for the mean and covariance of the underlying signals. Three refinements to the basic method in [19] are included with positive impact on the overall performance:

i. mean substitution,
ii. Tikhonov regularization and,
iii. dynamic principal components selection.

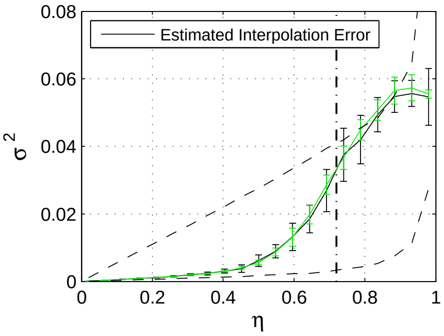
These extensions naturally increased the numerical robustness of the interpolation method and removed the original limitations on the interval of validity, thus paving the way to the application of the present methods to a number of real problems in the interpolation of multidimensional signals. Tight upper and lower bounds were presented and validated through a series of tests, with improved performance when compared with local averaging and the P-G methods. No bandlimited nor Gaussian noise assumptions are required for the signals and noise present, respectively. Sensitivity studies on a series of parameters in the estimators revealed a graceful degradation on the interpolation performance. Ultimately, the application of the proposed methodology to data obtained in a series of surveying missions at sea, with unmanned underwater vehicles, is expected to be the key enabling tool to tackle terrain based navigation problems with feature based techniques [18].

A framework for reconstructing bathymetric data acquired at sea tests was successfully introduced. The data processing necessary for the reconstruction based on PCA was found and employed on the survey performed at Espalamaca, Azores. It included the integration of discretization of the surveyed area and the LTP coordinate systems. Motivated by the indispensable reconstruction quality assessment of the data interpolation, estimators for the error were delineated and is validated not only by a proof, but also in an evaluating example of application. It represents a major contribution, which complements the methodology of interpolation with PCA, allowing off-line evaluation of the expected performance of the tool. The work developed for the Espalamaca survey data establishes an important vector of improvement in nowadays data acquisition with state-of-the-art sensors and transducers included in ASCs, ROVs, and AUVs for sea survey missions.

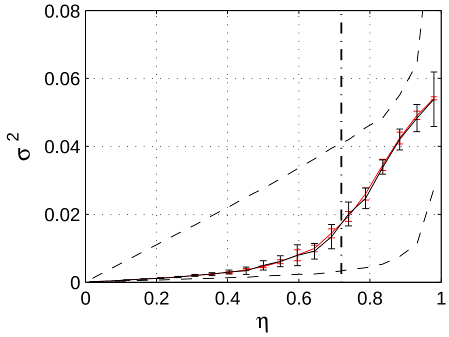
Also networked control systems, the subject of missing data is a relevant matter under study. In a number of situations we face the challenge of controlling a system from a incomplete signal obtained through sensors. The application of the interpolation based on PCA on Kalman filtering represents a starting point of improvement on the control over an unreliable communication channel. An exploratory empirical assessment was performed in order to determined the impact of such strategy on Kalman filtering. It was successfully implemented a recursive framework of the reconstruction with PCA where a number of transformations were conducted, which includes the development of recursive estimators for the mean and covariance, where the no-stationarity was taken in consideration by the incorporation of a forgetting factor. When considering stable systems the PCA interpolation showed a major improvement compared to previous methods. However, if the system is unstable, the PCA reconstruction reveled to be inadequate in such situations. Nonetheless, this work opens the door to future analysis in greater depth and formality.

VII. Acknowledgments

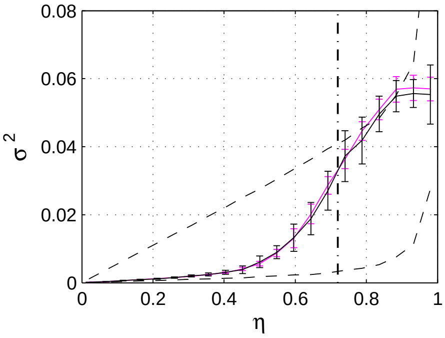
I would also like to thank to João Alves, Luís Sebastião, and Manuel Rufino from ISR/IST for their valuable contribution to the development, testing, and operation of the IRIS survey tool. A special thanks to Carlos Silvestre, PI of the MEDIRES project, and António Pascoal, PI of the project MAROV, that challenged the team to study the Espalamaca ridge in detail.



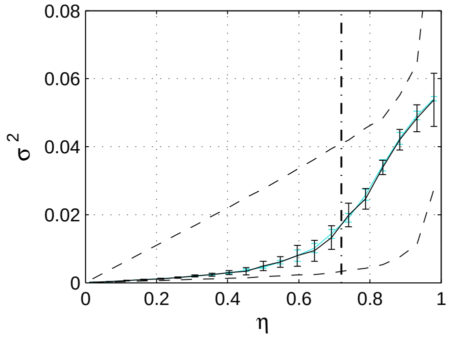
(a) Lemma 1 and Equation 5.



(b) Lemma 1 and Tikhonov Regularization



(c) Mean Substitution and Equation 5.



(d) Mean Substitution and Tikhonov Regularization.

Fig. 2. Error variance for the interpolation method with PCA, including the estimation of the error depicted in black, for $\eta \in [0.02, 0.98]$, $N = 25$ and $n = 7$. The error bars show the $\pm$ one standard deviation across the 20 runs for each η. For the regularized solutions $\alpha = 0.7\eta^2$.

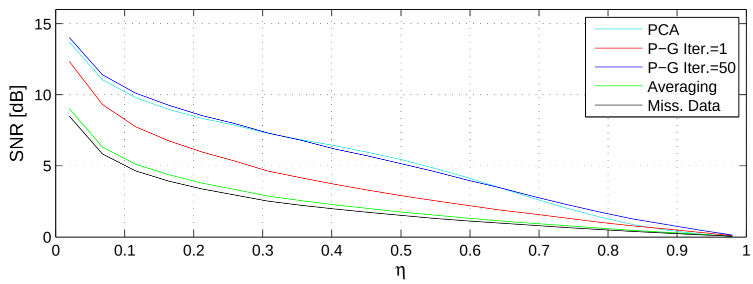
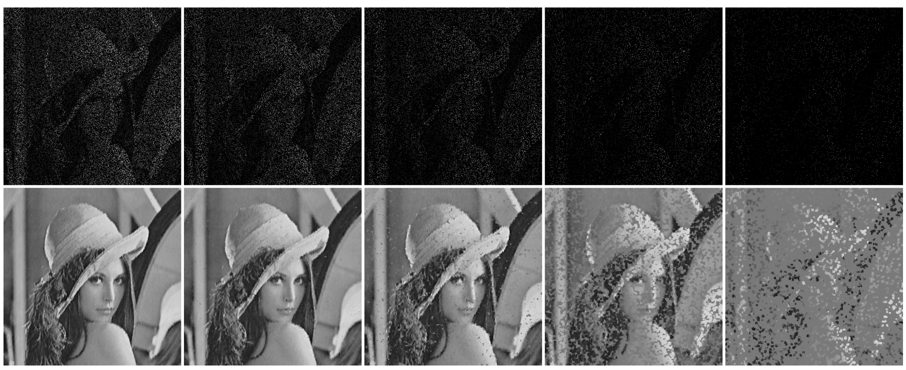


Fig. 3. 1D Signal interpolation $SNR(\mathbf{r}, \tilde{\mathbf{r}})$ for $\eta \in [0.02, 0.98]$ with the PCA interpolation method (in cyan), the alternative methods and with no interpolation applied (in black).



η=0.80 η=0.85 η=0.90 η=0.95 η=0.98

Fig. 4. 2D Signal interpolation with the PCA interpolation method (lower row). In the upper row, the corresponding images with missing data for the interval $\eta \in \{0.8, \ldots, 0.98\}$.

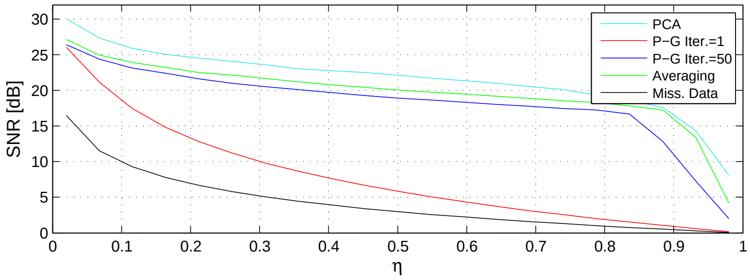


Fig. 5. 2D Signal interpolation $SNR(\mathbf{r}, \tilde{\mathbf{r}})$ for $\eta \in [0.02, 0.98]$ with the PCA interpolation method (in cyan), the alternative methods, and with no interpolation applied (in black).

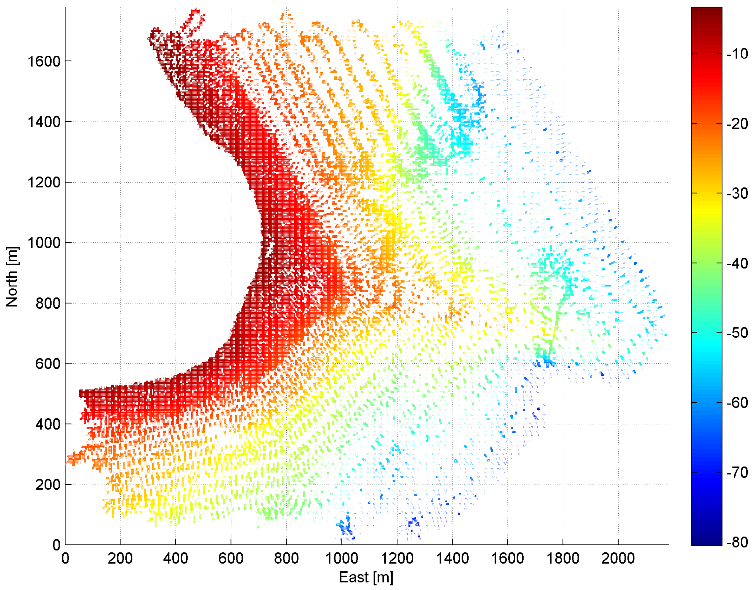


Fig. 6. Data from the survey mission to the Espalamaca converted to a matrix layout. The origin is located at latitude $38.5379°$ and longitude $-28.6097°$ with a resolution of 1254×1536 for an area of $H = 1779.6m$ height by $W = 2180.8m$ width.

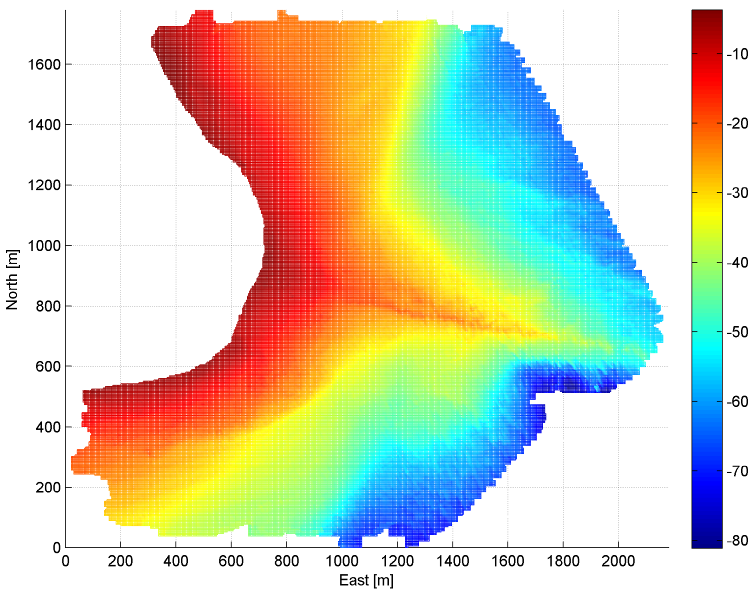


Fig. 7. Data with the PCA interpolation method, where $N = 17 * 17 = 289$, $n = 34$ and $\alpha_i = 0.1\eta_i^2$.

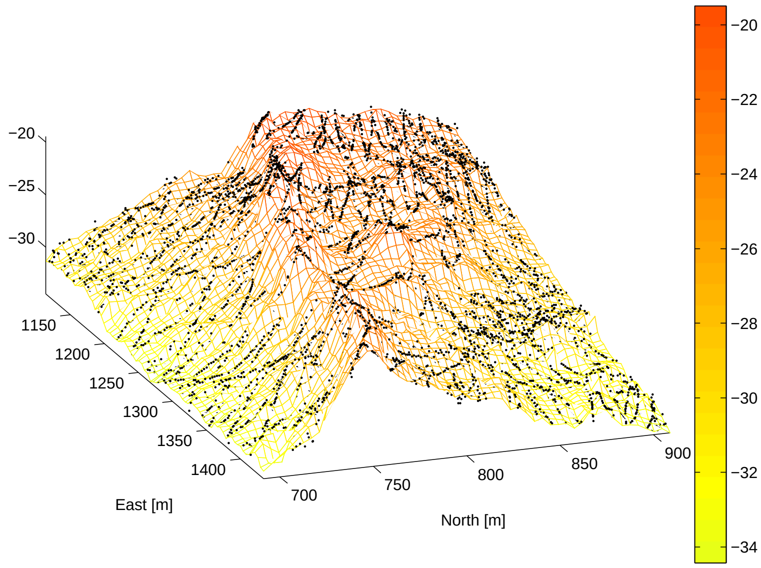


Fig. 8. 3D visualization of the data interpolated with PCA for a portion of the surveyed area and the measurements previously to the grid discretization and reconstruction (in black).

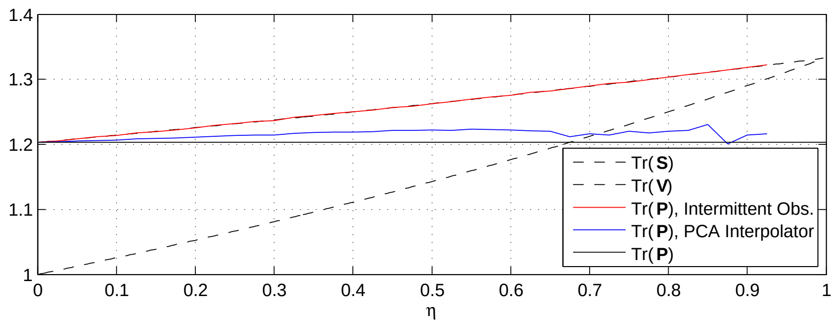


Fig. 9. Limit over time of the mean covariance error, for a SISO stable system $\mathbf{A} = -0.5$, over a set of missing observations ratios. The several approaches are included, with the corresponding bounds (in black).

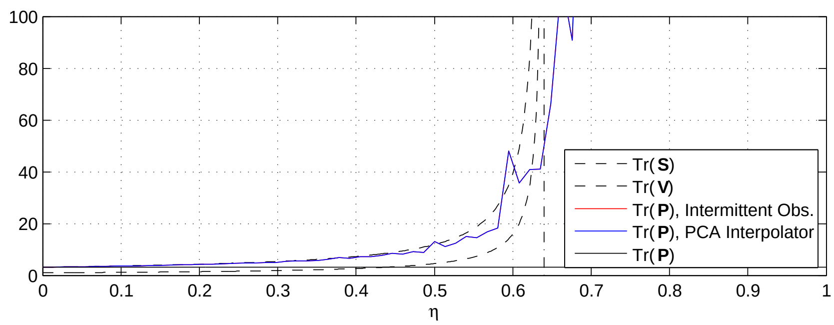


Fig. 10. Limit over time of the mean covariance error, for a SISO unstable system $\mathbf{A} = -1.25$, over a set of missing observations ratios. The several approaches are included, with the corresponding bounds (in black).

REFERENCES

[1] B. Anderson, and J. Moore, *Optimal Filtering*, Prentice-Hall, 1979

[2] J. Benedetto and P. Ferreira, *Modern Sampling Theory: Mathematics and Applications,* Birkhuser, 2000.

[3] V. Blanz, and T. Vetter, *Reconstructing the Complete 3D Shape of Faces from Partial Information,* it+ti Oldenburg, Verlag, 2002.

[4] H. Choi, and D. Munson, *Analyis and Design of Minimax-Optimal Interpolators,* IEEE Trans. on Signal Processing, vol.46, n.6, 1998.

[5] J. Farrell and M.Barth, *The Global Positioning System & Inertial Navigation,* McGraw Hill, 1999.

[6] A. Gelb, J. Kasper Jr., R. Nash Jr., C. Price, and A. Sutherland Jr., *Applied Optimal Estimation*, chapter 4, MIT Press, 1974

[7] R. Gerchberg, *Super-resolution through error energy reduction,* Optica Acta, vol.21, n.9, pp.709-720, 1974.

[8] M. Hadidi, and S. Schwartz, *Linear recursive state estimators under uncertain observations,* IEEE Transaction on Information Theory, vol. 24, no. 6, pp. 944-948, 1979.

[9] I. Jolliffe, *Principal Component Analysis,* Springer-Verlag, 2002.

[10] S. Kay, *Fundamentals of Statistical Signal Processing,* Prentice-Hall, 1993.

[11] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation,* Prentice-Hall, 2000.

[12] J. Lygeros, D. Godbole, and S. Sastry, *Verified hybrid controllers for automated vehicles,* IEEE Transactions on Automatic Control, vol. 43(4), 1998

[13] M. Mariton, *Jump Linear Systems in Automatic Control*, Marcel Dekker, 1990.

[14] F. Marvasti, *Nonuniform Sampling Theory and Practice Series: Information Technology: Transmission, Processing and Storage,* Springer-Verlag, 2001.

[15] A. Mertins, *Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and Applications,*John Wiley & Sons, 1999.

[16] N.C. Mitchell, T. Schmitt, F. Tempera, E. Isidro, F. Cardigos, J. Figueiredo, and J.C. Nunes, *Coastal processes around volcanic islands - dendritic lava flows, landslides and terraces around the central Azores islands,* Geophysical Research Abstracts, Vol. 7, 00177, 2005.

[17] N. Nahi, *Optimal recursive estimation with uncertain observation,* IEEE Transaction on Information Theory, vol. 15, no. 4, pp. 457-462, 1969.

[18] P. Oliveira, *MMAE terrain reference navigation for underwater vehicles using PCA,* accepted for publication on a special number of the International Journal of Control, in Navigation, guidance and control of uninhabited underwater vehicles, vol. 80, n. 7, pp.1008-1017, 2007.

[19] P. Oliveira, *Interpolation of Signals with Missing Data Using PCA,* Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing - ICASSP06, Toulouse, France, 2006.

[20] A. Papoulis, *A new algorithm in spectral analysis and band-limited extrapolation,* IEEE Transactions on Circuits Syst. vol.19, pp.735-742, 1975.

[21] A. Papoulis, *A new method in image restoration,* JSTAC, Paper VI-3, 1973.

[22] A. Pascoal, P. Oliveira, C. Silvestre, A. Bjerrum, A. Ishoy, J-P. Pignon, G. Ayela, and C. Petzelt, *MARIUS: An Autonomous Underwater Vehicle for Coastal Oceanography,* IEEE Robotics & Automation Magazine, 1997.

[23] A. Pascoal, C. Silvestre and P. Oliveira, *Vehicle and Mission Control of Single and Multiple Autonomous Marine Robots*, Advances in Unmanned Marine Vehicle (Roberts, G. N. and Sutton, R. (Ed)). ISBN 0-86341-450-8. IEE Control Series 69. pp. 353-386, January 2006.

[24] S. Roweis, *EM Algorithms for PCA and SPCA,* in *Advances in Neural Information Processing Systems,* vol.10, 1998.

[25] L. Silva, J. Santos, M. Neves, C. Silvestre, P. Oliveira, and A. Pascoal, *Tools for the diagnosis and automated inspection of semi-submerged structures*, 13th International Harbour Congress, Belgium, March-April, 2003.

[26] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, *Kalman Filtering with Intermittent Observations*, in *IEEE Transactions on Automatic Control, Special Issue on Networked Control Systems*, Setember, 2004.

[27] H-Y. Shum, K. Ikeuchi, and R. Reddy, *Principal Component Analysis with Missing Data and Its Application to Polyhedral Object Modeling,* IEEE Trans. Pattern Analyis and Machine Intelligence, vol.17, n.9, 1995.

[28] F. Tempera, N.C. Mitchell, T. Schmitt, E. Isidro, F. Cardigos, J. Figueiredo, and J.C. Nunes, *Dendritic lava flows, landslides and terraces around the central Azores islands,* American Geophysical Union - Fall Meeting, San Francisco (USA), 2004.

[29] A. Tikhonov, A. Goncharsky, V. Stepanov, and A. YagolaKluwer, *Numerical Methods for the Solution of Ill-posed Problems,* Academic Publishers, 1990.

[30] M. Unser, *Sampling - 50 Years After Shannon,* Proceedings of the IEEE, vol.88, n.4, 2000.

[31] P. Varaiya, *Smart cars on smart road: Problems of Controls,* IEEE Transactions on Automatic Control, vol. 38(2), 1993

[32] Y. Wang, D. Ho, and X. Liu, *Variance-constrained filtering for uncertain stochastic systems with missing measurements,* IEEE. Trans. Automat. Control, vol. 48, no. 7, pp. 1254-1258, 2003.

[33] J. Yen, *On nonuniform sampling of bandlimited signals,* IRE Transactions on Circuit Theory, vol. CT-3, 1956.