Missing Data Reconstruction based on Projection Manifolds

Alexander Bernstein
Laboratory of Cognitive Methods of
Data Analysis and Modelling
Institute of System Analysis
9, 60th October Anniversary pr., Moscow
Russia
a.bernstein@irias.ru

Evgeniy Burnaev
Laboratory of Information Transmission
Theory and Control
Institute of Information Transmission Problems
19, Bolshoy Karetny per., Moscow
Russia
burnaev@iitp.ru

Abstract

New method for missing values reconstruction based on topological properties of data is elaborated. Proposed technique generalizes method for missing data reconstruction based on ordinary principal components by taking into account nonlinearities in a learning sample. Application of elaborated methodology to the design of metamodels enables to get very accurate surrogates which then can be used for construction of systems with desired reliability.

1 Introduction

Problem of missing data reconstruction arises very frequently in different areas of applied science. It is usually assumed that missing data appears due to some technical problems rather than due to changes in the state of the system generating data. For example, missing values can appear due to inaccessibility of measurements or expensiveness of their acquisition, malfunctioning of measuring device, etc. Depending on the problem missing values or some functions of missing values should be reconstructed. In both cases reconstruction methods are based on assumption that there is some redundant information in the data sample expressed by some dependencies among features of observations.

Missing values can be reconstructed using topological structure of the data sample: missing coordinates of the multivariate vector are filled with the values obtained by its projection on the manifold near to which (in the sense of geometrical proximity) the data sample is located. This manifold is constructed using dimension reduction procedure and has dimension which is smaller than the dimension of the sample vectors.

In Aivazyan et al. 1983 the method realizing this idea is described. This method is linear as it is based on ordinary principal component analysis (PCA), i.e. nonlinear dependencies among features of observations are not taken into account.

The aim of the present work is to generalize the method based on PCA to the nonlinear case. Elaborated missing value reconstruction procedure is successfully applied to the problem of inverse design of airfoils(see also Hazarika et al. 1998).

2 Problem statement

Let our observations be represented as n-dimensional vectors $\mathbf{Y} = (y_1, y_2, \ldots, y_n) \in R^n$. We denote the known sample of observations of size N by $D_N = \{\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_N\}$.

Suppose that new observation $\mathbf{Y}$ is from the same class of observations as available sample D_N. Let us denote subvectors of $\mathbf{Y}$ by

$$\mathbf{Y} = (\mathbf{X}, \mathbf{Z}) = (X_1, X_2, \ldots, X_s, Z_1, Z_2, \ldots, Z_{n-s}),\ \mathbf{X} \in R^s,\ \mathbf{Z} \in R^{n-s},\ 1 \le s < n,$$

where $\mathbf{X}$ is known and $\mathbf{Z}$ is unknown.

The problem of missing data reconstruction is to generate such procedure $\Pi : R^s \to R^{n-s}$ using the sample D_N that its output $\mathbf{Z}^* = \Pi(\mathbf{X})$ is close in some sense to the real vector $\mathbf{Z}$.

The main steps of topological method for missing data reconstruction can be summarized as follows:

1. Using nonlinear dimension reduction procedure construct nonlinear manifold which is close to the data sample.

2. Estimate missing coordinates of a new observation by the projection of the observation onto the manifold.

In the following two sections mathematical formulations of these two steps are given.

3 Mathematical concepts of topological reconstruction

3.1 Dimension reduction

The procedure Σ used for dimension reduction performs two basic functions (see Bernstein et al. 2008-a, 2008-c, 2008-d)

- constructing a compressed description $\lambda = \lambda(\mathbf{Y})$ with a smaller dimension m for a given vector $\mathbf{Y}$ with the dimension n;
- reconstructing the "full-size" vector $\mathbf{Y}(\lambda)$ from a given compressed description λ.

The procedure $\Sigma = \{m, K_m, K_m^-\}$ used for dimension reduction consists of:

- dimension m of the compressed description;
- compression transformation $K_m : \mathbf{Y} \to K_m(\mathbf{Y}) = \lambda \in R^m$ that transforms the vector $\mathbf{Y}$ to the m-dimensional compressed vector $\lambda = (\lambda_1, \ldots, \lambda_m)$;
- reconstruction transformation $K_m^- : \lambda \to K_m^-(\lambda) = \mathbf{Y}$ that transforms the m-dimensional compressed vector λ to the full-size vector $\mathbf{Y}$.

Let us denote the reconstructed vector obtained by applying the reconstruction transformation K_m^- to the result of compression $K_m(\mathbf{Y})$ of the initial vector $\mathbf{Y}$ as $\mathbf{Y}^* = K_m^-(K_m(\mathbf{Y})) = (Y_1^*, \ldots, Y_n^*)$.

Let

$$\Delta(\mathbf{Y}) = \Delta(\mathbf{Y}|\Sigma) = \mathbf{Y}^* - \mathbf{Y} = (\Delta_1, \ldots, \Delta_n)$$

be a vector consisting of recovery errors and let

$$\rho(\Delta) = \sqrt{\sum_{i=1}^{n} \Delta_i^2}$$

be the distance between the initial vector $\mathbf{Y}$ and the recovered vector $\mathbf{Y}^*$. The distance $\rho(\Delta(\mathbf{Y}|\Sigma))$ characterizes the quality of the dimension reduction procedure Σ applied to a specific vector $\mathbf{Y}$.

For a given data sample D_N the integral sample characteristic of the dimension reduction procedure Σ calculated from this data set is

$$R = R(\Sigma|D_N) = \frac{1}{N} \sum_{i=1}^{N} \rho(\Delta(\mathbf{Y}_i|\Sigma)).$$

3.2 Topological reconstruction of missing data

Let us assume that we construct some dimension reduction procedure Σ using data sample D_N. Then we can define a manifold (nonlinear in general)

$$\Omega_m = \left\{\mathbf{Y} \in R^n : \mathbf{Y} = K_m^-(\lambda),\, \lambda \in \Lambda \subset R_m\right\},$$

which is a closest manifold among manifolds from considered class (the class of manifolds is defined by used dimension reduction procedure).

Let us denote (see also section 2) by $\mathbf{Z}$ the unknown subvector of the given vector $\mathbf{Y}$ and represent the reconstruction transformation in the form

$$K_m^-(\lambda) = \left(K_{\mathbf{X},m}^-(\lambda), K_{\mathbf{Z},m}^-(\lambda)\right),\ K_{\mathbf{X},m}^-(\lambda) \in R^s,\ K_{\mathbf{Z},m}^-(\lambda) \in R^{n-s}$$

in compliance with the known and unknown parts of the vector $\mathbf{Y}$.

In order to estimate the unknown subvector $\mathbf{Z}$ we project the vector $\mathbf{Y}$ onto the manifold Ω_m.

Mathematically this projection Z^* can be found as follows. We should solve the following optimization problem

$$(\lambda^*, \mathbf{Z}^*) = \underset{\lambda, \mathbf{Z}}{\operatorname{argmin}} \left(\left\| \mathbf{X} - K^-_{\mathbf{X},m}(\lambda) \right\|^2 + \left\| \mathbf{Z} - K^-_{\mathbf{Z},m}(\lambda) \right\|^2 \right), \tag{1}$$

where $\| \cdot \|$ is a euclidian norm. It is obvious that the pair

$$\mathbf{Z}^* = K^-_{\mathbf{Z},m}(\lambda^*),\ \lambda^*(\mathbf{X}) = \underset{\lambda}{\operatorname{argmin}} \left\| \mathbf{X} - K^-_{\mathbf{X},m}(\lambda) \right\|^2$$

is the solution of (1). Thus, missing data reconstruction procedure takes the form

$$\mathbf{Z}^* = \Pi(\mathbf{X}) = K^-_{\mathbf{Z},m}(\lambda^*(\mathbf{X})).$$

4 Dimension reduction based on projection manifold

Let us describe special nonlinear dimension reduction procedure $\Sigma(NLPCA)$ which generalizes linear dimension reduction procedure PCA and allows to construct projection manifold Ω_m of dimension m. We will use this manifold for topological reconstruction of missing data (see section 3).

The procedure $\Sigma(NLPCA) = \{\Sigma, K_{NLPCA}, K^-_{NLPCA}\}$ makes two transformations:

- compression transformation $K_{NLPCA,m} : \mathbf{Y} \to K_m(\mathbf{Y}) = \lambda \in R^m$,
- reconstruction transformation $K^-_{NLPCA,m} : \lambda \to K^-_m(\lambda) = \mathbf{Y}^* \in R^n$.

According to the papers of Bernstein et al. 2008-a, 2008-d reconstruction transformation defines projection manifold Ω_m, and compression transformation is simply the projection of the vector $\mathbf{Y}$ onto this manifold.

Nonlinear manifold Ω_m is constructed by nonlinear approximation of vectors of residuals. These vectors are located in the orthogonal complement to the plane defined by first m principal components obtained using PCA. Approximators are used for construction of Ω_m.

Dimension reduction procedure $\Sigma(NLPCA_m)$ is indexed by m equal to the number of chosen principal components, coordinates of which parameterize nonlinear manifold Ω_m.

Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be n-dimensional eigenvectors corresponding to the eigenvalues $\{\mu_1, \ldots, \mu_n\}$ of the sample covariance matrix calculated using D_N with $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$. Let the transformation $K_{PCA,n} : \mathbf{Y} \to K_n(\mathbf{Y}) = \lambda \in R^n$ convert each vector $\mathbf{Y}_r$ of the data set D_N into a n-dimensional vector

$$\lambda^{(r)} = \lambda(\mathbf{Y}_r) = \left(\lambda_1^{(r)}, \ldots, \lambda_n^{(r)} \right),\ r = 1, 2, \ldots, N,$$

where

$$\lambda_i^{(r)} = (\mathbf{Y}_r - \mathbf{Y}_{\text{mean}}, \mathbf{e}_i),\ i = 1, 2, \ldots, n,\ \mathbf{Y}_{\text{mean}} = \sum_{i=1}^{N} \mathbf{Y}_i / N.$$

Assuming that the intrinsic dimension of the data is m, $m < n$, one may expect the components of the vector $\lambda^{(r)}$ with the indices $i > m$ to be a functions of the variables $\lambda_1^{(r)}, \ldots, \lambda_m^{(r)}$. It means that for any $r = 1, 2, \ldots, N$

$$\lambda_{m+1}^{(r)} = g_{m+1}\left(\lambda_1^{(r)}, \ldots, \lambda_m^{(r)} \right), \ldots, \lambda_n^{(r)} = g_n\left(\lambda_1^{(r)}, \ldots, \lambda_m^{(r)} \right).$$

These dependencies define an m-dimensional surface Ω_m, located in the n-dimensional plane which is a nonlinear span of the first m principal components

$$\Omega_m = \{\omega(\lambda_1, \ldots, \lambda_m) | \lambda_1 \in R, \ldots, \lambda_m \in R\}, \tag{2}$$

where

$$\omega(\lambda_1, \ldots, \lambda_m) = (\lambda_1, \ldots, \lambda_m, g_{m+1}(\lambda_1, \ldots, \lambda_m), \ldots, g_n(\lambda_1, \ldots, \lambda_m)). \tag{3}$$

The functions $g_k(\lambda_1, \ldots, \lambda_m)$, $k = m+1, \ldots, n$ should be reconstructed based on the initial sample D_N. To do this we should perform the approximation of multidimensional nonlinear dependencies. It is important that the functions g_k must be looked for in the class of nonlinear dependencies. Otherwise we would get a simple redefinition of the basis of principal components.

The multidimensional nonlinear dependencies are approximated using new techniques of approximation developed by Bernstein et al. 2008-b, Burnaev et al. 2009.

Obtaining the approximation of the functions $g_k(\lambda_1, \ldots, \lambda_m)$ we get a parametric description of the m-dimensional manifold Ω_m using m principal components for manifold parametrization according to (2), (3). The manifold Ω_m is described in terms of initial coordinates using the vector-function

$$\mathbf{Y} = \mathbf{Y}(\lambda_1, \ldots, \lambda_m) = \mathbf{Y}_{\text{mean}} + \sum_{i=1}^{m} \lambda_i \mathbf{e}_i + \sum_{i=m+1}^{n} g_i(\lambda_1, \ldots, \lambda_m)\mathbf{e}_i$$

of the variables $(\lambda_1, \ldots, \lambda_m)$, where $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ are vectors of the PCA basis. This vector-function defines the reconstruction transformation $K^{-}_{NLPCA,m}(\lambda) = \mathbf{Y}$.

5 Experiments

In order to check the efficiency of the proposed procedure we used it for inverse design of airfoils (see also Bernstein et al. 2008-c). The sample size was equal to 5418, each vector of pressure distribution has the dimension $n = 59$. The quality of data reconstruction procedure is estimated by the mean absolute error (L_1) between the initial and reconstructed vectors of airfoil descriptions.

Table 1: Errors of reconstruction of airfoil descriptions.

Error	PCA ($m = 7$)	Projection Manifold ($m = 7$)
L_1	0.0036	0.0027

Acknowledgements

This work was partially supported by the RFBR grants 09-01-00584-a, 09-07-00180-a.

References

[1] Aivazyan, S.A., Yenyukov, I.S., Meshalkin, L.D. (1983). *Applied Statistics. Bases of Modelling and Initial Data Processing.* Moscow, Finansy i statistica. (In Russian)

[2] Bernstein, A.V., Burnaev, E.V., Dorofeev, E.A., Sviridenko, Yu.N., Chernova, S.S. (2008-a). Cascade Procedure for Dimension Reduction. *Proceedings of the 11th National Conference on Artificial Intelligence CII-08*, 28 September - 3 October, Dubna, 241-250. (In Russian)

[3] Bernstein, A.V., Kuleshov, A.P., Burnaev, E.V. (2008-b). An Approach for Approximation of Multidimensional Dependences. *Selected papers of IV International Conference "Parallel Computations and Control Problems" (PACO '2008)*, 27-29 October, Moscow, 56-62.

[4] Bernstein, A.V., Kuleshov, A.P. (2008-c). Cognitive Technologies in Problem of Dimension Reduction of Description of Geometrical Objects. *Information Technologies and Computing Systems, 2.*

[5] Bernstein, A.V., Kuleshov, A.P. (2008-d). Construction of Orthogonal Nonlinear Manifolds in Problems of Dimension Reduction. *Proceedings of VII International school-seminar "Multidimensional Statistical Analysis and Econometrics"*, 21-30 September, Tsahkadzor. (In Russian)

[6] Burnaev, E., Belyaev, M., Prixodko, P., Chernova, S. (2009). Neural Approximation based on Regression and Boosting. *In the present book.*

[7] Hazarika, N., Tuncer, I.H., Lowe, D. (1998). An Inverse Design Procedure for Airfoils using Artificial Neural Networks. *Proceedings of the 21st ICAS (International Council of the Aeronautical Sciences) Congress*, 13-18 September, Melbourne, Australia.

Neural Approximation based on Regression and Boosting

Evgeniy Burnaev
Laboratory of Information Transmission Theory and Control
Institute of Information Transmission Problems
19, Bolshoy Karetny per., Moscow
Russia
burnaev@iitp.ru

Svetlana Chernova
Laboratory of Cognitive Methods of Data Analysis and Modelling
Institute of System Analysis
9, 60th October Anniversary pr., Moscow
Russia
chernova@irias.ru

Mikhail Belyaev
Faculty of Radio Engineering and Technical Cybernetics
Moscow Institute of Physics and Technology
9, Institutskiy per., Dolgoprudniy
Moscow Region, Russia
belyaevmichel@gmail.com

Pavel Prihodko
Faculty of Radio Engineering and Technical Cybernetics
Moscow Institute of Physics and Technology
9, Institutskiy per., Dolgoprudniy
Moscow Region, Russia
prikhodkop@live.ru

Abstract

New multidimensional approximation technique NARB (Neural Approximation based on Regression and Boosting) is elaborated. This technique is a combination of Regression Analysis and Resilient Propagation learning with subsequent construction of Artificial Neural Nets ensemble on basis of gradient leveraging. Application of NARB to the design of nonlinear approximators enables to get very accurate surrogate models which then can be used for construction of systems with desired reliability.

1 Introduction

The problem of determining the analytical description for a set of data arises in numerous sciences and applications and can be referred to as data modelling or system identification. Artificial Neural Networks (ANN) are a convenient means of representation because they are universal approximators that can learn data by example (see Sutton et al. 1998, Werbos et al. 1992). They can be easily trained to map multidimensional nonlinear functions because of their parallel architecture.

The problem of multidimensional function modelling is very widespread. For example in Gavard et al. 1996 ANN is used to predict the properties of metal on basis of its structure. In Linse et al. 1993 ANN is used to approximate the dependency of aerodinamic characteristics of an airplane on its layout and flight regime. In Narendra et al. 1990 ANN is used for the identification and control of dynamical systems, mapping the input-output representation of an unknown system and, possibly, its control law.

However ANN with rather simple structure is used for approximation in the mentioned papers. Usually this is two-layer ANN in which the first layer has hyperbolic transfer function and the second layer has linear transfer function. We will call such ANN as simple ANN in the sequel.

The aim of the present work is to describe the general methodology for multidimensional function approximation which outperform methods of approximation based on simple ANN both in learning time and accuracy. The proposed methodology is based on combined use of simple ANN, Linear Regression, Resilient Propagation learning and ensemble construction based on gradient leveraging. The proposed methodology originates from the work of Bernstein et al. 2008.

2 Problem Statement

Let us consider some multidimensional function

$$y = f(\mathbf{X}), \mathbf{X} \in R^M \ (M \gg 1), y \in R^1. \tag{1}$$

There is the sample $\Sigma = \{y_i, \mathbf{X}_i\}_{i=1}^N$ of the function values (1). The sample is divided randomly into two parts $\Sigma = \Sigma_{learn} \bigcup \Sigma_{test}$, where the size of the subsample Σ_{test} (testing set) is equal to $10 - 20\%$ of the

size of the sample Σ. Approximation problem consists in construction of such multidimensional function $y = \tilde{f}(\mathbf{X})$, $\mathbf{X} \in R^M, y \in R$ using the sample Σ_{learn} for which some error measure

$$D\left(f, \tilde{f} | \Sigma_{test}\right) \leq \varepsilon,$$

where ε is a desired error. Usually quadratic error measure

$$D\left(f, \tilde{f} | \Sigma_{test}\right) = \sqrt{\sum_{\{y, \mathbf{X}\} \in \Sigma_{val}} \left(y - \tilde{f}(\mathbf{X})\right)^2}$$

is used. The proposed methodology of approximator construction consists of several parts each described in separate section.

3 Hybrid learning algorithm for ANN

The simple ANN is used as a building block for elaborated multidimensional approximator. The error backpropagation method has been greatly used for the supervised training of simple ANN (see Haykin 2006). However, as it is well-known, this method has a slow convergence. Several techniques have been developed to speed up this method, such as, among others, second order algorithms, adaptive step size methods, appropriate weights initialization etc (see Tarhov 2005, Haykin 2006, Riedmiller et al. 1993). In this section hybrid learning algorithm for simple ANN is described based on Regression Analysis and Resilient Propagation. The use of Regression Analysis allows to considerably decrease the learning time. This effect can be explained by the fact that on each iteration of the hybrid algorithm the obtained weight values of the second layer are optimal.

Let us describe how Regression Analysis can be used to estimate the second layer weights of the simple ANN. The simple ANN realizes the following transformation of the input vector $\mathbf{X}$

$$y_{ANN}(\mathbf{X}) = \sigma\left(\mathbf{X} \cdot \mathbf{W}^{\mathrm{T}} + \mathbf{d}\right) \cdot \mathbf{V} + b, \tag{2}$$

where $\mathbf{X} \in R^M$ is the input row vector, $\mathbf{W} \in R^{m \times M}$ and $\mathbf{d} \in R^m$ are the matrix and row vector of weights of the hidden layer, $\mathbf{V} \in R^m$ (column vector) and $b \in R$ are the weights of the second layer, $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid activation function, which acts on each component of the row vector $\mathbf{X} \cdot \mathbf{W}^{\mathrm{T}} + \mathbf{d}$ separately, m is equal to the number of neurons in the hidden layer.

Let us divide randomly the learning set Σ_{learn} into subsets S_{train} and S_{val}. The first set is used for parameter estimation of ANN and the second one is used for control of generalization ability of ANN. The weights of the simple ANN are defined as the solution of the following optimization problem

$$E = E(\mathbf{W}, \mathbf{d}, \mathbf{V}, b) = \sum_{\{y, \mathbf{X}\} \in S_{train}} \left(y - y_{ANN}(\mathbf{X})\right)^2 \to \min_{\mathbf{W}, \mathbf{d}, \mathbf{V}, b}. \tag{3}$$

Suppose that we know the values of the weights $\mathbf{W}$ and $\mathbf{d}$. Since the function $y_{ANN}(\mathbf{X})$ in (2) depends on $\mathbf{V}$ and b linearly, then the values of this weights can be obtained using Regression Analysis (for some fixed $\mathbf{W}$ and $\mathbf{d}$). Let $N_{train} = \#(S_{train})$ be the training sample size, $\mathbf{Q} = \mathbf{Q}(S_{train}) \in R^{N_{train} \times m}$ be the matrix with rows equal to $\sigma\left(\mathbf{X} \cdot \mathbf{W}^{\mathrm{T}} + \mathbf{d}\right)$ for $\mathbf{X} \in S_{train}$, $\mathbf{Y} \in R^{N_{train}}$ is a column vector of function values $y \in S_{train}$, $\overline{y} = \frac{1}{N_{train}} \sum_{y \in S_{train}} y$, $\overline{\mathbf{Q}}$ is a row vector elements of which are equal to the sample means calculated using the columns of the matrix $\mathbf{Q}$, $\widetilde{\mathbf{Q}}$ is a "centered" matrix $\mathbf{Q}$ (from each column of the matrix $\mathbf{Q}$ the corresponding component of the vector $\overline{Q}$ was subtracted), analogously, $\widetilde{\mathbf{Y}}$ is a "centered" (using $\overline{y}$) vector $\mathbf{Y}$. Using Regression Analysis we get that

$$\mathbf{V} = \left(\widetilde{\mathbf{Q}}^{\mathrm{T}} \widetilde{\mathbf{Q}}\right)^{-1} \cdot \left(\widetilde{\mathbf{Q}}\right)^{\mathrm{T}} \cdot \widetilde{\mathbf{Y}}, \; b = \overline{y} - \overline{\mathbf{Q}} \cdot \mathbf{V}. \tag{4}$$

We use Resilient Propagation (see Riedmiller et al. 1993) for weights $\mathbf{W}$ and $\mathbf{d}$ adjustment. It follows from (4) that the weights of the second layer are the functions of the weights of the first layer. However it can be easily proved that the result of gradient calculation does not depend on whether the weights $\mathbf{W}$ and $\mathbf{d}$ depend on the weights $\mathbf{V}$ and b according to formula (4) or not.

Thus the hybrid learning algorithm for simple ANN can be described as follows:

1. The weights $\mathbf{W}$ and $\mathbf{d}$ are initialized using special initialization algorithm (see Nguyen et al. 1990), the weights $\mathbf{V}$ and b are calculated using (4).

2. The weights $\mathbf{W}$ and $\mathbf{d}$ are adjusted using Resilient Propagation (see Riedmiller et al. 1993).

3. The the weights $\mathbf{V}$ and b are calculated using (4) etc.

4. If for the user-defined number of consecutive iterations the error on the set S_{val} does not decrease below its current minimal value and/or one of the stopping criteria of Resilient Propagation is fulfilled the hybrid learning algorithm is stopped. Otherwise goto 2.

4 Ensemble of approximators

Proposed hybrid learning algorithm for simple ANN outperforms standard algorithms (see Tarhov 2005, Haykin 2006, Riedmiller et al. 1993) both in learning time and accuracy. However the hybrid algorithm has typical shortcomings of standard learning algorithms – the result depends on the initialization, only part of the learning data is used for weights adjustment, the division of Σ_{learn} into S_{train} and S_{test} is random. In order to smooth over the influence of these shortcomings we propose to construct ANN ensemble on basis of specially elaborated gradient leveraging algorithm, based on ideas of boosting and bagging methods (see Breiman et al. 1994, Friedman 1999, Friedman et al. 1999). This algorithm can be described as follows:

1. Let the initial output of the ensemble equal to $F_0(\mathbf{X}) = 0$.

2. For $B = 1, 2, \ldots$:

   (a) Let us define a new sample $\Sigma_{learn,B} = \{y^B, \mathbf{X}\}$, where

   $$y^B = B \cdot y - (B-1) \cdot F_{B-1}(\mathbf{X})$$

   for $\{y, \mathbf{X}\} \in \Sigma_{learn}$. We divide the learning set $\Sigma_{learn,B}$ into two parts $S_{train,B}$ and $S_{val,B}$ randomly.

   - Train the B-th approximator $f_B(\mathbf{X})$ of the ensemble using the sample $\Sigma_{learn,B}$ and the hybrid learning algorithm. The set $S_{train,B}$ is used for weights adjustment. If for the user-defined number of consecutive iterations the error on the set $S_{val,B}$ does not decrease below its current minimal value and/or one of the stopping criteria of Resilient Propagation is fulfilled the hybrid learning algorithm is stopped. Otherwise continue.
   - The output of the ensemble is set to $F_B(\mathbf{X}) = \frac{B-1}{B} \cdot F_{B-1}(\mathbf{X}) + \frac{1}{B} \cdot f_B(\mathbf{X})$.
   - The algorithm for ensemble construction is stopped if for the user-defined number of iterations the error $D\left(f, F_B | \Sigma_{learn}\right)$ does not decrease below its current minimal value. Otherwise continue.

Usage of the proposed algorithm for ensemble construction enables us to adapt to the whole learning set Σ_{learn} and decrease the random variations (induced by random weights initialization and random learning set splitting) since each new approximator in the ensemble will correct the errors of the previous approximators.

5 Experiments

In order to check the efficiency of NARB technique we used it to approximate the dependency of aerodynamic characteristics of an airplane on its layout and flight regime.

We use two types of relative errors – mean relative absolute error ($\mathrm{E}_{\mathrm{mean}}$) and 95%-quantile relative absolute error ($\mathrm{E}_{\mathrm{Q}(0.95)}$). These errors can be calculated as follows

$$\mathrm{E}_{\mathrm{mean}} = \mathrm{Mean}\left\{\left|y - F_{NARB}(\mathbf{X})\right| / y, \{y, \mathbf{X}\} \in \Sigma_{test}\right\},$$

$$\mathrm{E}_{\mathrm{Q}(0.95)} = \mathrm{Max}_{0.95}\left\{\left|y - F_{NARB}(\mathbf{X})\right| / y, \{y, \mathbf{X}\} \in \Sigma_{test}\right\}$$

Table 1: Errors of approximation of lateral pressure coefficient on learning and testing sets.

Error	ANN		ANN + hybrid learning		NARB	
	Σ_{learn}	Σ_{test}	Σ_{learn}	Σ_{test}	Σ_{learn}	Σ_{test}
E_{mean} (in %)	2.29	2.78	2.01	2.45	1.15	1.61
$E_{Q(0.95)}$ (in %)	3.09	3.75	2.83	3.37	1.58	2.21

(maximal error for the set Σ_{test} after deleting of 5% of data points with maximal errors).

It can be seen from the table 1 that the errors of approximation on basis of NARB technique is significantly lower compared to errors obtained using simple ANN. On average 85 minutes is needed to learn approximator using NARB technique and 278 minutes is needed to learn approximator using standard error backpropagation algorithm.

Another experiments with NARB technique can be found in Burnaev et al. 2009.

Acknowledgements

This work was partially supported by the RFBR grants 09-01-00584-a, 09-07-00180-a.

References

[1] Bernstein, A.V., Kuleshov, A.P., Burnaev, E.V. (2008). An Approach for Approximation of Multidimensional Dependences. *Selected papers of IV International Conference "Parallel Computations and Control Problems" (PACO '2008)*, 27-29 October, Moscow, Institute for Control Problems, 56-62.

[2] Burnaev, E., Grihon, S. (2009). Construction of the Metamodels in Support of Stiffened Panel Optimization. *In the present book.*

[3] Breiman, L. (1994). Bagging Predictors. *Technical Report No. 421, Department of Staticstics, University of California.*

[4] Friedman, J.H. (1999). Greedy function approximation: a gradient boosting machine. *Technical Report, Department of Staticstics, Stanford University.*

[5] Friedman, J.H., Hastie, T., Tibshirani, R. (1999). Additive logistic regression: a statistical view of boosting. *Technical Report, Department of Staticstics, Stanford University.*

[6] Gavard, L., Bhadeshia, H.K.D.H., MacKay, D.J.C., Suzuki, S. (1996). Bayesian neural network model for austenite formation in steels. *Mater. Sci. Technol. 12*, 453-463.

[7] Haykin, S. (2006). *Neural Networks: A Comprehensive Foundation Second Edition.* Williams.

[8] Linse, D., Stengel, R.F. (1993). Identification of aerodynamic coefficients using computational neural networks. *J. Guid. Control Dyn. 16 (6)*, 1018–1025.

[9] Narendra, K.S., Parthasarantly, K. (1990). Identification and control of dynamical systems using neural networks. *IEEE Trans. Neural Netw. 1(1)*, 4-27.

[10] Nguyen, D., Widrow, B. (1990). Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights. *Proc. Intl. Joint Conf. on Neural Networks III*, 21-26, San Diego, CA.

[11] Riedmiller, M., Braun, H. (1993). A direct adaptive method for faster backpropagation learning: the RPROP algorithm. *Proceedings of the IEEE International Conference on Neural Networks.*

[12] Sutton, R.S., Barto, A.G. (1998). *Reinforcement learning.* Cambridge MA, MIT Press.

[13] Tarhov, D.A. (2005). *Neural Networks. Models and Algorithms. Vol. 18.* Moscow, Radio Engineering. (In Russian)

[14] Werbos, R.J. (1992). Neurocontrol and supervised learning: an overview and evaluation. *Handbook of Intelligent Control. Ed. by White, D.A., Sofge, D.A.,* 65–86. New York, Van Nostrand.

Construction of the Metamodels in Support of Stiffened Panel Optimization

Evgeniy Burnaev
Laboratory of Information Transmission Theory and Control
Institute for Information Transmission Problems
19, Bolshoy Karetny per., Moscow
Russia
burnaev@iitp.ru

Stéphane Grihon
Engineering Structure Analysis
AIRBUS France
316, Route de Bayonne, Toulouse
France
stephane.grihon@airbus.com

Abstract

Metamodels are a valuable tool to support a wide scope of activities in modern engineering design, especially design optimization, and allows to get products with prescribed level of reliability at a significantly lower time compared to expensive simulation processes. In this work we construct metamodels for approximation of optimization constraints by performing regression of static mechanical criteria, namely buckling and collapse reserve factors of a stiffened panel. Due to the highly non-linear behavior of these functions with respect to loadings and design variables the use of conventional methods of approximation (Kriging, Artificial Neural Networks, Support Vector Regression, Multivariate Nonparametric Regression, Polynomial Regression) did not allow us to get sufficient quality on the whole design space. We show how metamodel with very high quality can be constructed on basis of newly developed approximation technique NARB (Neural Approximation based on Regression and Boosting).

1 Introduction

Aeronautical structures are mainly made of stiffened panels, i.e. thin shells (also called skin) enforced with stiffeners (respectively called frames and stringers) in both orbital and longitudinal directions (Samuelides et al. 2009, Merval et al. 2006). For the sake of study the whole structure is divided into elementary parts called Super Stiffeners, consisting in the theoretically union of a stringer and two half panels. These basic structures are subject to highly non-linear phenomena such as buckling, collapse and damage tolerance.

In order to determine the optimal size of these super stiffeners, static mechanical criteria must be computed using dedicated software that is based on non-linear calculation. Thus, the analysis and the dimension estimation of the whole structure is currently computed by running a two-level study: at a global level, a Finite Element (FE) analysis run on the whole FE model provides internal loads – applied to each S-Stiffener; at a local level, these loads are used to compute static mechanical criteria. Most of these criteria are formulated using Reserve Factors (RF): a structure is validated provided all its RFs are greater than 1.

Therefore detailed design of an aircraft fuselage requires a two-level loop: first, numerous local optimizations are run on isolated super stiffeners in order to size them respecting mechanical criteria, depending on current internal loads. But changes in local geometry from initial to optimal design involve a new load distribution in the whole structure; an update step must then be performed to take these changes into account. This bilevel optimization process is then repeated until convergence of load distribution in the whole structure.

Local optimizations are computed using local methods that require gradients of the constraint functions – which can only be obtained by finite differences. Values of mechanical stability constraints are computed using dedicated software through analytical calculation. A call to this software takes up to a second; as a consequence the need for finite difference calculations in each numerous local optimization greatly increases the time between two update steps. Much time could be saved by using a reduced model instead of performing straightforward computing (Forrester et al. 2008). This time saving is the aim of the present study and is of great importance to engineers working in Airbus Structural Analysis Framework.

In this study we focus on the metamodelling of static instability phenomena which are the buckling and the collapse of a super stiffener. For this we use specially designed approximation technique called NARB (Neural Approximation based on Regression and Boosting; see Bernstein et al. 2008, Burnaev

et al. 2009) and achieve significantly better results compared to the results obtained by conventional methods (Kriging, Artificial Neural Networks, Support Vector Regression, Multivariate Nonparametric Regression, Polynomial Regression; see Bishop 2006, Chiles et al. 1999) and by method of Merval et al. 2006 specially designed for the considered problem.

2 Problem statement

As described in the previous section the aim of the study is to perform approximation of several mechanical criteria that are constraints of some local optimization problem (see Merval et al. 2006 for details). These criteria are formulated as Reserve Factors (RFs) so that a structure is feasible if its RFs are greater than 1. In this study we focus on a T-shaped composite stringer joint to composite panels (see figure 1). RFs of interest give information on: stringer local buckling RF_{STR}; minimum RF for buckling (skin, stringer, inter-rivert) RF_{BCK}, which expresses the collapse of the whole super stiffener. Thus RF_{BCK} is a continuous function with discontinuous gradient.

Values of these criteria according to design and loading variables are supplied by internal Airbus software to be used in a new project focused on optimization of composite stiffened panels. This static calculation software assesses instability phenomena such as buckling, local buckling and post buckling solving an energy method (Rayleigh-Ritz method). It gives information on damage tolerance margins too.

RF_{BCK} and RF_{STR} depend on both the geometry of the structure and the applied loading. Here, super Stiffener's geometry is defined by (see figure 1): left and right panel thickness t_{lp} and t_{rp}; stringer flange thickness and width t_f and t_f; stringer web thickness and height t_w and h_w; panel width and length.

In our study panel width and length and flange width are constant, so that geometry is defined by 5 continuous variables. The super stiffener is subject to two-dimensional loading defined by: compression force F_{comp} applied to the super stiffener (applying on both panel and stringer); transverse flux N_{YY} and shear flux N_{XY}, both applied to the panel only. Thus, the input space is 8-dimensional in this first application case (5 geometrical variables and 3 loading ones).

The aim of the study was to perform regression of function RF_{BCK} and RF_{STR} in the following domain: t_{lp} and t_{rp} in $[2, 20]$; t_f and t_w in $[2, 30]$; h_w in $[30, 80]$; F_{comp} in $[-1.5E6, -1000]$; N_{YY} in $[-5000, 5000]$; N_{XY} in $[0, 5000]$. Length unity is millimetre, forces are expressed in Newton and fluxes in N/mm.

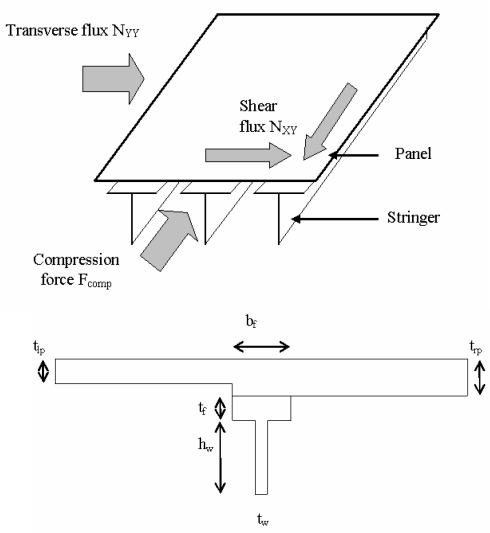

Figure 1: Schematic layout of the T-shaped composite stringer joint to composite panels.

3 Construction of the surrogate model

Mathematically the process of construction of surrogate model $\widehat{y} = F_{surr}(\mathbf{X})$ for some unknown function $y = F(\mathbf{X})$ consists of several steps:

- Automatic generation of the training sample (input vectors are generated for example using Latin Hypercube space filling design, see Forrester et al. 2008). The output is the set of data points $\Sigma_{estim} = \{y_i, \mathbf{X}_i\}_{i=1}^N$, where y_i is the value of RF for the corresponding input vector $\mathbf{X}_i \in \Re^8$.

- Let $y_{\min}$ and $y_{\max}$ be the up and the low limits of the output value $y = F(\mathbf{X})$ (can be estimated using generated sample or/and knowledge of the physical mechanism underlying the function $y = F(\mathbf{X})$). The interval of variation of the output variable $y \in [y_{\min}, y_{\max}]$ is divided into K subintervals, i.e. $[y_{\min}, y_{\max}] = \bigcup_{i=1}^K [y_{i-1}, y_i]$, where $y_0 = y_{\min}$, $y_K = y_{\max}$. In general case the subintervals can intersect. The selection of particular decomposition of the interval of variation $[y_{\min}, y_{\max}]$ into subintervals is done by human expert on the basis of the physical sense of approximated dependency. For example when approximating $\mathrm{RF_{STR}}(\mathbf{X})$ or $\mathrm{RF_{BCK}}(\mathbf{X})$ the accuracy domain $A_\varepsilon = \{\mathbf{X} : 1 - \varepsilon \le F(\mathbf{X}) \le 1 + \varepsilon\}$ is included into such decomposition. This allows to substantially increase the accuracy of approximation in this domain (important for applications).

- The sample is divided into such K subsamples $\Sigma_{estim} = \bigcup_{i=1}^K \Sigma_i$ that in each of the subsample Σ_{estim}^i there are only those values of $\{y, \mathbf{X}\}$ for which $y \in [y_{i-1}, y_i]$, $i = 1, \ldots, K$.

- On basis of each subsample Σ_{estim}^i an approximator $F_{approx}^i(\mathbf{X})$ is constructed using NARB method described in Bernstein et al. 2008, Burnaev et al. 2009.

- On basis of the sample $\Sigma_{estim} = \{y_i, \mathbf{X}_i\}_{i=1}^N$ a classifier (see Bishop 2006) $F_{class}(\mathbf{X})$ is constructed. This classifier for each $\mathbf{X}$ gives the subset of the constructed approximators $F_{approx}^j(\mathbf{X})$, $j \in J(\mathbf{X}) \subseteq \{1, \ldots, K\}$ which should be used to predict the value y of the reserve factor at the given point $\mathbf{X}$.

Two steps are done in order to predict the value of $y = F(\mathbf{X})$ by the value $F_{surr}(\mathbf{X})$ of the surrogate model for the given $\mathbf{X}$:

- Using the classifier $F_{class}(\mathbf{X})$ the corresponding subset $F_{approx}^j(\mathbf{X})$, $j \in J(\mathbf{X}) \subseteq \{1, \ldots, K\}$ of constructed approximators are chosen.

- Using $F_{approx}^j(\mathbf{X})$, $j \in J(\mathbf{X}) \subseteq \{1, \ldots, K\}$ the value of the $y = F(\mathbf{X})$ is estimated according to the formula $F_{surr}(\mathbf{X}) = \sum_{j \in J(\mathbf{X})} \omega(j, \mathbf{X}) F_{approx}^j(\mathbf{X})$, where $\omega(j, \mathbf{X})$, $j \in J(\mathbf{X}) \subseteq \{1, \ldots, K\}$ are weights such that $\sum_{j \in J(\mathbf{X})} \omega(j, \mathbf{X}) = 1$ and $\omega(j, \mathbf{X}) > 0$ for all $j \in J(\mathbf{X}) \subseteq \{1, \ldots, K\}$.

4 Experiments

4.1 Quality of local approximation

The learning and testing samples of sizes 35000 and 115000 correspondingly were obtained in which outputs $\mathrm{RF_{BCK}}$ and $\mathrm{RF_{STR}}$ were restricted to $[0.5, 1.5]$, i.e. inputs belong to the accuracy domain A_ε with $\varepsilon = 0.5$.

In order to compare different approximation techniques we use two type of errors: mean absolute error ($\mathrm{E_{mean}}$) and 95%-quantile absolute error ($\mathrm{E_{Q(0.95)}}$). These errors can be calculated as follow. Let Σ be some set of data points $\{y, \mathbf{X}\}$, then $\mathrm{E_{mean}} = \mathrm{Mean}\left\{|y - F_{surr}(\mathbf{X})|, \{y, \mathbf{X}\} \in \Sigma\right\}$ and $\mathrm{E_{Q(0.95)}} = \mathrm{Max}_{0.95}\left\{|y - F_{surr}(\mathbf{X})|, \{y, \mathbf{X}\} \in \Sigma\right\}$ (maximal error for the set Σ after deleting of 5% of data points with maximal errors).

In the table 1 there are errors of approximation for the data from the learning and testing sets, obtained using NARB, Kriging, Artificial Neural Networks (ANN), Support Vector Regression (SVR), Multivariate Nonparametric Regression (MNR), Linear Regression (LR).

It can be seen from the table 1 that NARB method significantly outperforms all other conventional methods for both RFs.

Table 1: Errors of approximation of RF_{BCK} and RF_{STR} on the testing set.

RF	Error	NARB	Kriging	ANN	SVR	MNR	LR
RF_{BCK}	E_{mean}	0.011	0.026	0.032	0.153	0.112	0.161
	$E_{Q(0.95)}$	0.036	0.081	0.087	0.395	0.323	0.403
RF_{STR}	E_{mean}	0.009	0.020	0.030	0.149	0.104	0.143
	$E_{Q(0.95)}$	0.027	0.068	0.083	0.401	0.301	0.351

4.2 Quality of surrogate models

In order to estimate the quality of the surrogate models obtained using NARB and methodology of surrogate model construction, described in the previous section, we generate the learning and testing samples of sizes 100000 and 1550000 correspondingly. We decompose the interval of variation of RF into 4 intervals. One of these intervals corresponds to the accuracy domain $A_{0.5}$. In the table 2 there are errors of the surrogate models on the testing set for RF_{BCK} and RF_{STR}. These errors are calculated for the whole testing set as well as for its subsets, obtained imposing restrictions on the values of corresponding RF.

Table 2: Errors of the surrogate models for RF_{BCK} and RF_{STR} on the testing set.

RF	Error	$\forall$ RF	RF $\in [0.5, 1.5]$	RF < 0.5	RF > 1.5
RF_{BCK}	E_{mean}	0.012	0.011	0.007	0.023
	$E_{Q(0.95)}$	0.038	0.037	0.023	0.057
RF_{STR}	E_{mean}	0.010	0.009	0.005	0.013
	$E_{Q(0.95)}$	0.030	0.028	0.009	0.037

It can be seen from the table 2 that the errors of the surrogate model is comparable in the accuracy domain $A_{0.5}$ with the errors of the local approximator (see table 1) constructed using NARB.

Acknowledgements

This work was partially supported by the RFBR grants 09-01-00584-a, 09-07-00180-a.

References

[1] Bernstein, A.V., Kuleshov, A.P., Burnaev, E.V. (2008). An Approach for Approximation of Multidimensional Dependences. *Selected papers of IV International Conference "Parallel Computations and Control Problems" (PACO '2008)*, 27-29 October, Moscow, Institute for Control Problems, 56-62.

[2] Bishop, C.M. (2006). *Pattern recognition and machine learning*. Springer.

[3] Burnaev, E., Belyaev, M., Prixodko, P., Chernova, S. (2009). Neural Approximation based on Regression and Boosting. *In the present book*.

[4] Chiles, J.-P., Delfiner, P. (1999). *Geostatistics. Modeling Spatial Uncertainty*. Wiley.

[5] Forrester, A., Sobester, A., Keane, A. (2008). *Engineering Design via Surrogate Modelling. A Practical Guide*. Wiley.

[6] Merval, A., Samuelides, M., Grihon, S. (2006). Application of response surface methodology to stiffened panel optimization. In: *47th conference on AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Newport, Rhode Island.

[7] Samuelides, M., Bettebghor, D., Grihon, S., Merval, A., Morlier, J. (2009). Modèles réduits en optimisation multiniveau de structures aéronautiques. In: *9e Colloque national en calcul des structures*, 25-29 Mai, Giens, France.