CONSTRAINED NONSTATIONARY SIGNAL PROCESSING BY PAIR-WISE SEPARABLE QUADRATIC PROGRAMMING

V. Mottl, O. Krasotkina
Tula State University
Lenin Ave., 92, Tula, Russia
ko180177@uic.tula.ru

I. Muchnik
Rutgers University
P.O. Box 8018, Piscataway, NJ 08855,USA
muchnik@dimacs.rutgers.edu

ABSTRACT

There exists a wide class of nonstationary signal processing problems which invite for formulating them as those of estimating the succession of nonstationary regression parameters constrained at each time moment by linear inequalities. Such a formalization leads to the so-called pair-wise separable quadratic programming problem whose specificity consists in that, first, the quadratic objective function is block tridiagonal and, second, the inequality constraints are imposed individually upon each vector variable. The two nonstationary regression estimation algorithms considered in this paper are built as those of pair-wise separable quadratic programming and have the linear computational complexity in contrast to the polynomial computational complexity of the quadratic programming problem of general kind. The asymptotically strict iterative algorithm is based on the traditional steepest descent method of quadratic programming, whereas the fast approximate algorithm consists in a single run of a special version of the dynamic programming procedure.

KEYWORDS
Multidimensional signal processing, estimation of signal parameters, nonstationary regression, pair-wise separable quadratic programming, dynamic programming

1. INTRODUCTION

The variational approach to the synthesis of data analisys algorithms [1] consists in the systematic use of the fact that the decision-making process is always based on the search for the minimum values of certain appropriate measure $J(X\,|\,Y)$ of noncompliance between the record data array Y and the result of its processing X within the limits of a given admissible domain of variation. On has merely to select the class of objective function in such a way as to guarantee the existence of a sufficiently fast algorithm of search for the point of minimum, $\hat{X} = \arg\min J(X\,|\,Y)$ which is the result of processing.

In this paper, we discuss the optimization problem that arise when the variational approach is used to synthesize the algorithms for the analysis of signals $Y=(y_t,\ \ t_{\min} \le t \le t_{\max})$ whose essential feature consists in the fact that they are ordered along the axis of one of the arguments, as a rule, the time or space coordinate. In this paper, we consider a sufficiently wide class of problems of signal analysis along the axis of a discrete argument $Y=(y_t,\, t=1,...,N)$, in which it is required, for all points t, to estimate the current values of sufficiently smoothly changing vector parameter $\mathbf{x}_t$ of a certain local model; the sequence of these estimates $X=(\mathbf{x}_t,\, t=1,...,N)$ is the desired result.

As a rule, the special features of each applied problem determine the way of obtaining preliminary local estimates $\mathbf{x}_t^0$ which must be then matched on the basis of an a priori assumption that the hidden parameter $\mathbf{x}_t$ varies smoothly. Since the elements of a signal are linearly ordered, the variational estimation principle leads to the objective functions $J(X)=J(\mathbf{x}_1,...,\mathbf{x}_N)$, which are structured as sums of partial functions, each of which depends only on one or two neighboring arguments; i.e.,

$$J(X)=\sum_{t=1}^{N}\psi_t(\mathbf{x}_t)+\sum_{t=2}^{N}\gamma_t(\mathbf{x}_{t-1},\mathbf{x}_t)\,. \qquad (1)$$

The natural procedure for the optimization of objective functions of such a kind is the dynamic programming method, which allows one to replace the initial problem of search for the minimum of the function of many arguments by the sequence of substantially less difficult problems of minimization of intermediate functions of one argument called Bellman functions. However, the main dynamic programming procedure is fundamentally based on the assumption that the values of the objective variables make up a finite set, because the vital element of this procedure is the storage of the sequence of Bellman functions.

In this paper, the dynamic programming method is extended to the case of continuous variables for the quadratic objective functions by means of introducing the concept of a parametric family of quadratic Bellman functions. In addition, we introduce the concepts of the left an right Bellman functions, which allow us to use these functions in the construction of a natural criterion for detection of nonsmoothnesses in the variation of the parameter being estimated. The use of two-sided Bellman functions also allows one to construct a computationally efficient leave-one-out (jackknife) procedure that is used to verify a nonstationary model by means of sequential matching of the experimental value of each element of a signal with its value as predicted by the model constructed without invoking this element.

It is typical for many applications that some natural constraints are imposed upon the sought-for sequence of regression coefficients. As it can be easily shown nonstationary regression estimation problem in presence only equality constraints is lead to quadratic programming problem, consisted in minimization quadratic pair-wise separable objective function without constraints.

2. SEPARABLE OBJECTIVE FUNCTION OF GENERALIZED SIGNAL ANALYSIS PROBLEM

Examples of signal analysis problems that lead to the separable objective function can be found in [1–3]. In this paper, we consider the problem of nonstationary regression estimation, which is encountered in many applications. In this problem, the vector signal $Y=\big((y_t^0,\mathbf{y}_t),\ t=1,...,N\big)$ to be analyzed consists of two parts, namely, a vector sequence $\mathbf{y}_t\in\mathsf{R}^n$ and a numerical sequence $y_t^0\in\mathsf{R}$. The main assumption of the model consists in that the vector component y_t^0 is the signal and also in the choice of the sequence of recorded as a manifestation of some external process whose properties are not studied, and the numerical component at each point of observation is obtained as a noise-corrupted linear function of the vector component, $y_t^0=\mathbf{x}_t^T\mathbf{y}_t+\xi_t$, which is determined by the sequence of unknown vector coefficients of regression, $\mathbf{x}_t\in\mathsf{R}^n$ The problem consists in estimating the sequence of regression coefficients, $X=(\mathbf{x}_t,\ t=1,...,N)$,on the assumption that these coefficients are changing sufficiently smoothly and only seldom undergo jumplike changes. Regarding the observation noise ξ_t, it is assumed to be of the type that is known in the theory of stochastic processes as white noise of unknown intensity.

It is natural to require for the desired sequence of regression coefficients, $X=(\mathbf{x}_t,\ t=1,...,N)$ that the squared signal approximation residual be as small as possible at each point; i.e., $\psi_t(\mathbf{x}_t)=(\mathbf{x}_t-\mathbf{x}_t^0)^T\mathbf{Q}_t^0(\mathbf{x}_t-\mathbf{x}_t^0)\to\min\limits_{\mathbf{x}_t}$. This quadratic function is degenerate because the nonnegatively definite matrix $\mathbf{Q}_t^0=\mathbf{y}_t\mathbf{y}_t^T$ is degenerate. The set of values of $\mathbf{x}_t$ that ensure zero values of instantaneous residuals constitutes an affine manifold of dimension n – 1 in R^n where $\mathbf{x}_t^0=(y_t^0/\mathbf{y}_t^T\mathbf{y}_t)\ \mathbf{y}_t$ is the vector with minimum norm. It is natural to consider the quadratic function $\psi_t(\mathbf{x}_t)$ centered at the point $\mathbf{x}_t^0$ as the "diffused" local function of the instantaneous vector of regression coefficients.

On the other hand, the model assumption of the smoothness of the sequence $X=(\mathbf{x}_t,\ t=1,...,N)$ being sought for dictates the requirement that $\gamma_t(\mathbf{x}_{t-1},\mathbf{x}_t)=(\mathbf{x}_{t-1}-\mathbf{x}_t)^T\mathbf{U}_t(\mathbf{x}_{t-1}-\mathbf{x}_t)\to\min\limits_{\mathbf{x}_{t-1},\mathbf{x}_t}$ for each pair of neighboring elements of the signal, where $\mathbf{U}_t$ is a fitting sequence of positive definite $n\times n$ matrices which are identical in the simplest case; the choice of this sequence determines the required degree and character of smoothing of "diffused" local estimates.

These two conflicting requirements are compromised by a hybrid quadratic criterion (1). Many other signal analysis problems, including signal smoothing, time-spectrum signal analysis, pairwise alignment of signals of similar shape and different duration undergoing local distortions in scale along the axis of the argument, which were considered in [1–3], lead to criteria of exactly the same form. The special character of each particular problem manifests itself only in the form of the dependence of local estimators of the hidden parameters $\mathbf{x}_t^0=\mathbf{x}_t^0(Y)$ and the matrix $\mathbf{Q}_t^0=\mathbf{Q}_t^0(Y)$ on the signal and also in the choice of the sequence of smoothing matrices, $\mathbf{U}_t$.

In general, the problem of minimizing the quadratic-criterion (1) leads to the set of linear algebraic relations with a block-tridiagonal matrix standing by the unknowns; this problem is solved by the sweep method with the number of operations growing linearly as the length N of the signal increases. To understand the nature of the problem and to find the clues for synthesizing other useful algorithms of signal analysis, we consider this problem from the standpoint of applicability of the dynamic programming method to its solution; this method, while adequate for separable criteria of the form (1), was designed for solving discrete optimization problems.

3. AVOIDANCE OF EQUALITY CONSTRAINTS

In this paper, we consider a particular case of the constrained nonstationary regression problem when linear equality constraints are imposed separately upon each vector variable $\mathbf{x}_t\in\mathsf{R}^n$

$$\mathbf{B}_t\mathbf{x}_t+\mathbf{f}_t=\mathbf{0}\ ,\ t=1,...,N\ . \qquad (2)$$

If $\mathbf{B}_t(k\times n)$, $k<n$, is a full-rank matrix that has k linearly independent columns, then the constraint (2) can be written as $\bar{\mathbf{B}}_t\bar{\mathbf{x}}_t+\bar{\bar{\mathbf{B}}}_t\bar{\bar{\mathbf{x}}}_t+\mathbf{f}_t=\mathbf{0}$ or, in an equivalent form, as $\bar{\bar{\mathbf{x}}}_t=\bar{\bar{\mathbf{B}}}_t^{-1}(\bar{\mathbf{B}}_t\bar{\mathbf{x}}_t+\mathbf{f}_t)$, where matrix $\bar{\mathbf{B}}_t\left[k\times(n-k)\right]$ and nonsingular square matrix $\bar{\bar{\mathbf{B}}}_t(k\times k)$ are the submatrices of matrix $\mathbf{B}_t=(\bar{\mathbf{B}}_t,\bar{\bar{\mathbf{B}}}_t)$, and vectors

$\bar{\mathbf{x}}_t \in R^{n-k}$ and $\bar{\bar{\mathbf{x}}}_t \in R^k$ are the corresponding subvectors of vector $\mathbf{x}_t = (\bar{\mathbf{x}}_t^T, \bar{\bar{\mathbf{x}}}_t^T)^T$. Then the combined vector $\mathbf{x}_t = \mathbf{D}_t \bar{\mathbf{x}}_t + \mathbf{d}_t$, where $\mathbf{D}_t = \left(\mathbf{I}, (\bar{\bar{\mathbf{B}}}_t^{-1} \bar{\mathbf{B}}_t)^T\right)^T$ and $\mathbf{d}_t = \left(\mathbf{0}^T, (\bar{\bar{\mathbf{B}}}_t^{-1} \mathbf{f}_t)^T\right)^T$, satisfies the equality (2) for any vector $\bar{\mathbf{x}}_t$. Putting the expression for vector $\mathbf{x}_t$ into the original model $y_t^0 = \mathbf{x}_t^T \mathbf{y}_t + \xi_t$ gives an equivalent form of the linear model without equality restrictions $\bar{y}_t^0 = \bar{\mathbf{x}}_t^T \bar{\mathbf{y}}_t + \xi_t$ where $\bar{\mathbf{y}}_t = \mathbf{D}_t^T \mathbf{y}_t$, $\bar{y}_t^0 = y_t^0 - \mathbf{d}_t^T \mathbf{y}_t$.

Therefore, the presence of equality constraints does not complicate the optimization problem. Removing these constraints means decreasing the number of elements in each coefficient vector by k. The nonstationary regression estimation problem under equality constraints leads to the optimization problem that consists in minimization of a quadratic pair-wise separable objective function without any constraints. In what follows, we shall omit the upper line in vectors and matrices on the assumption that the problem consists in estimation of the equivalent nonstationary regression model $y_t^0 = \mathbf{x}_t^T \mathbf{y}_t + \xi_t$ without constraints.

4. "FORWARD THEN BACK" PROCEDURE OF DYNAMIC PROGRAMMING AND PARAMETRIC FAMILY OF QUADRATIC BELLMAN FUNCTIONS

The central idea of the dynamic programming method is the concept of a sequence of Bellman functions, $\tilde{J}_t^-(\mathbf{x}_t) = \min_{\mathbf{x}_1,\ldots,\mathbf{x}_{t-1}} J_t^-(\mathbf{x}_1,\ldots,\mathbf{x}_t)$ which are related to partial criteria $J_t^-(\mathbf{x}_1,\ldots,\mathbf{x}_t) = \sum_{s=1}^t \psi_s(\mathbf{x}_s) + \sum_{s=2}^t \gamma_s(\mathbf{x}_{s-1},\mathbf{x}_s)$ Bellman functions are sequentially computed for $t = 1,\ldots,N$ by the easy-to-prove recurrence relation $\tilde{J}_t^-(\mathbf{x}_t) = \psi_t(\mathbf{x}_t) + \min_{\mathbf{x}_{t-1}}\left[\gamma_t(\mathbf{x}_{t-1},\mathbf{x}_t) + \tilde{J}_{t-1}^-(\mathbf{x}_{t-1})\right]$ starting from $\tilde{J}_1^-(\mathbf{x}_1) = \psi_1(\mathbf{x}_1)$ Obviously, $\hat{\mathbf{x}}_N = \arg\min_{\mathbf{x}_N} \tilde{J}_N(\mathbf{x}_N)$ is the optimal value of the last variable, and other optimal values are easily found in the reverse order by the rule $\hat{\mathbf{x}}_{t-1} = \tilde{\mathbf{x}}_{t-1}(\hat{\mathbf{x}}_t) = \arg\min_{\mathbf{x}_{t-1}}\left[\gamma_t(\mathbf{x}_{t-1},\hat{\mathbf{x}}_t) + \tilde{J}_{t-1}^-(\mathbf{x}_{t-1})\right]$ if the reversed recurrence relations $\tilde{\mathbf{x}}_{t-1}(\mathbf{x}_t)$ were stored when the Bellman functions were calculated in the forward order. The procedure described above is the classical dynamic programming procedure; it is natural to call it the "forward then back" procedure.

It is clear that the reversed recurrence relations $\tilde{\mathbf{x}}_{t-1}(\mathbf{x}_t)$ can be stored in an explicit form only if the range of variables $\mathbf{x}_t$ is a finite set. However, note that if the functions $\psi_t(\mathbf{x}_t)$ and $\gamma_t(\mathbf{x}_{t-1},\mathbf{x}_t)$ in criterion (1) are quadratic, then all the Bellman functions are likewise quadratic; i.e., they can be represented in the form $\tilde{J}_t^-(\mathbf{x}_t) = \tilde{b}_t^- + (\mathbf{x}_t - \tilde{\mathbf{x}}_t^-)^T \tilde{\mathbf{Q}}_t^-(\mathbf{x}_t - \tilde{\mathbf{x}}_t^-)$. As a consequence, for the Bellman function to be found successively, it suffices to recurrently recalculate their parameters, $\tilde{b}_t^-$, $\tilde{\mathbf{x}}_t^-$ and $\tilde{\mathbf{Q}}_t^-$. These parameters completely determine the backward recurrence relations, which will be linear in this case; i.e., $\tilde{\mathbf{x}}_{t-1}(\mathbf{x}_t) = \tilde{\mathbf{x}}_{t-1}^- + \tilde{\mathbf{H}}_{t-1}(\mathbf{x}_t - \tilde{\mathbf{x}}_{t-1}^-)$, $\tilde{\mathbf{H}}_{t-1} = (\tilde{\mathbf{Q}}_{t-1}^- + \mathbf{U}_t)^{-1}\mathbf{U}_t$. The working formulas for the recalculation of Bellman function parameters can be found in [4].

It is obvious that the parameter of the last Bellman function is nothing else but the optimal value of the last variable $\hat{\mathbf{x}}_N = \tilde{\mathbf{x}}_N$. The optimal values of the variables can be easily calculated by using backward recurrence relations $\hat{\mathbf{x}}_{t-1} = \tilde{\mathbf{x}}_{t-1}^- + \tilde{\mathbf{H}}_{t-1}(\hat{\mathbf{x}}_t - \tilde{\mathbf{x}}_{t-1}^-)$.

However, below, we consider two problems that are associated with nonstationary regression estimation, but appropriate in other applications, as well, for which the classical "forward then back" scheme of dynamic programming is not adequate. In the following section, we propose another scheme where two types of Bellman functions are operating instead of one; these functions are calculated toward each other from opposite endpoints of the signal.

5. "FORWARD AGAINST FORWARD" PROCEDURE OF DYNAMIC PROGRAMMING

Along with the partial criteria $J_t^-(\mathbf{x}_1,\ldots,\mathbf{x}_t)$ that are found from the left to the right starting from the first sampling of the signal and are called left partial criteria hereafter, it is quite natural to consider the right partial criteria $J_t^+(\mathbf{x}_t,\ldots,\mathbf{x}_N) = \sum_{s=t}^N \psi_s(\mathbf{x}_s) + \sum_{s=t}^{N-1} \gamma_{s+1}(\mathbf{x}_s,\mathbf{x}_{s+1})$ which are symmetric to the left ones and are found from the end of the signal. Accordingly, in addition to the left Bellman function $\tilde{J}_t^-(\mathbf{x}_t) = \min_{x_1,\ldots,x_{t-1}} J_t^-(\mathbf{x}_1,\ldots,\mathbf{x}_t)$, we introduce the right analogs of these functions, $\tilde{J}_t^+(\mathbf{x}_t) = \min_{\mathbf{x}_{t+1},\ldots,\mathbf{x}_N} J_t^+(\mathbf{x}_t,\ldots,\mathbf{x}_N)$ which are recurrently calculated in the opposite direction; i.e., $\tilde{J}_t^+(\mathbf{x}_t) = \psi_t(\mathbf{x}_t) + \min_{\mathbf{x}_{t+1}}\left[\gamma_{t+1}(\mathbf{x}_t,\mathbf{x}_{t+1}) + \tilde{J}_{t+1}^+(\mathbf{x}_{t+1})\right]$ starting from the last right Bellman function $\tilde{J}_N^+(\mathbf{x}_N) = \psi_N(\mathbf{x}_N)$. The right Bellman functions, in the quadratic version $\tilde{J}_t^+(\mathbf{x}_t) = \tilde{b}_t^+ + (\mathbf{x}_t - \tilde{\mathbf{x}}_t^+)^T \tilde{\mathbf{Q}}_t^+(\mathbf{x}_t - \tilde{\mathbf{x}}_t^+)$, like the left ones, are completely characterized by their parameters $\tilde{b}_t^+$, $\tilde{\mathbf{x}}_t^+$ and $\tilde{\mathbf{Q}}_t^+$.

The "meeting" of the left and right Bellman functions at some point t directly yields the optimal value of the objective variable at these point, $\hat{\mathbf{x}}_t = \arg\min_{\mathbf{x}_t} \left[\tilde{J}_t^-(\mathbf{x}_t) + \tilde{J}_t^+(\mathbf{x}_t) - \psi_t(\mathbf{x}_t) \right]$. Thus, we have obtained the alternative scheme of dynamic programming, which is called the "forward against forward" procedure.

6. APPLICATION OF THE "FORWARD AGAINST FORWARD" PROCEDURE TO DETECT DISCONTINUITIES IN THE VARIATIONOF PARAMETERS BEING ESTIMATED

If the hypothesis about the relatively slow character of variation of the parameter to be estimated, $\mathbf{x}_t$, is only partly true and this parameter undergoes isolated jump-like variations at certain pairs of neighboring points $t-1$ and t, then the estimation algorithm will inevitably suppress these jumps by bringing the estimates $\hat{\mathbf{x}}_{t-1}$ and $\hat{\mathbf{x}}_t$ closer together. In many practical signal analysis problems, it is important to retain the pronounced jumps in the variation of the parameter being estimated. This can be achieved if the basic sufficiently "stringent" function $\gamma_t(\mathbf{x}_{t-1}, \mathbf{x}_t)$ that penalizes nonsmoothness is replaced at the corresponding pairs of neighboring points by more "mild" function $\gamma_t^*(\mathbf{x}_{t-1}, \mathbf{x}_t)$, but to do this, it is necessary to detect the jumps first.

A pair of left and right Bellman functions, $\tilde{J}_{t-1}^-(\mathbf{x}_{t-1})$ and $\tilde{J}_t^+(\mathbf{x}_t)$, respectively, is a convenient tool for verifying whether there is a jump between the two neighboring points $t-1$ and t. If the points of minima of these functions, $\tilde{\mathbf{x}}_t^-$ and $\tilde{\mathbf{x}}_t^+$, are close to each other, then a jump is unlikely, while a substantial discrepancy witnesses the presence of a jump. This is the main idea of the edge detection and preservation procedure presented in [3].

7. "FORWARD AGAINST FORWARD" JACKKNIFE PROCEDURE IN VERIFYING A NONSTATIONARY MODEL OF A SIGNAL

When the variational principle is applied to signal analysis, it is extremely important to verify the adequacy of the nonstationary model, which is obtained under a certain degree of smoothing. For instance, the higher the observation noise level ξ_t in the problem of estimation of the nonstationary regression model, the lower the variability of the sequence of coefficients of regression $\mathbf{x}_t$, because the adequate estimation of these coefficients is otherwise impossible. The obtained non-stationary regression model can be validated by calculating the mean square prediction error $\xi_t = y_t^0 - \hat{\mathbf{x}}_t^T \mathbf{y}_t$ over all samples of the signal and comparing it with the total variance of observations y_t^0; however, when the error is calculated at the current point t, one cannot employ the estimate $\hat{\mathbf{x}}_t$ which was obtained by using this observation.

The variance of the prediction error can be estimated by using the idea of leave-one-out (jackknife) verification procedure, which is widely applied in pat-tern recognition. Assume that the observation y_t^0 is skipped, which is easily simulated by setting $\psi_t(\mathbf{x}_t) \equiv 0$ in criterion (1). Then, the estimate of the vector of regression coefficients at this point $\mathbf{x}_t$ obtained by minimizing the criterion, is governed by interpolation between neighboring estimates due to the effect of smoothing. It is this estimate that should be used in the calculation of the error at this point.

If one uses the classical "forward then back" procedure in the minimization of a criterion, then one has to run this procedure separately for calculating the error $\xi_t = y_t^0 - \mathbf{x}_t^T \mathbf{y}_t$ in each sampling of a signal, because every time another function $\psi_t(\mathbf{x}_t)$ must vanish. On the other hand, if the "forward against forward" procedure is applied, it is sufficient to compute all right and left Bellman functions only once with nonvanished functions $\psi_t(\mathbf{x}_t)$. Then, the estimates $\hat{\mathbf{x}}_t$, which correspond to the skip of one corresponding observation y_t^0, can easily be computed all at once by the following rule: $\hat{\mathbf{x}}_t = \arg\min_{\mathbf{x}_t} \left[\tilde{J}_t^-(\mathbf{x}_t) + \tilde{J}_t^+(\mathbf{x}_t) - 2\psi_t(\mathbf{x}_t) \right]$.

REFERENCES

1. Mottl, V.V., Blinov, A.V., Kopylov, A.V., and Kostin,A.A.,Computer-aided Signal and Image Processing: A Uni-versal Variational Approach, *Journal of Journals: Review of Global Scientific Achievement*s, 1998, vol. 2, no. 1, pp. 23–30.
2. Mottl, V., Blinov, A., Kopylov, A., and Muchnik, I., Variational Methods in Signal and Image Analysis, *Proc. 14th Int. Conf. on Pattern Recognitio*n, Brisbane, 1998, vol. 1, pp. 525–527.
3. Mottl, V., Kopylov, A., and Kostin, A., Edge-preserving in Generalized Smoothing of Signals and Images, *Proc. 14th Int. Conf. on Pattern Recognition,* Brisbane, 1998, vol. 2, pp. 1579–1581.
4. Muchnik, I. and Mottl, V., Bellman Functions on Trees for Segmentation, Generalized Smoothing, Matching and Multialignment in Massive Data Sets. *DIMACS Technical Report* 98-15, 1998. Center for Discrete Math-ematics and Theoretical Computer Science, Rutgers Univ., the State Univ. of New Jersey, USA. ftp://dimacs.rutgers.edu/pub/dimacs/Techical-Reports/TechReports/ 1998/98-15.ps.gz