PAIR-WISE SEPARABLE QUADRATIC PROGRAMMING
FOR CONSTRAINED TIME-VARYING REGRESSION ESTIMATION

O. Krasotkina
Tula State University, Tula, Russia
ko180177@yandex.ru

ABSTRACT
Estimation of time-varying regression model constrained at each time moment by linear inequalities is a natural statistical formulation of a wide class of nonstationary signal processing problems. The presence of linear constraints turns the originally quadratic three-diagonal problem of minimizing the residual squares sum, which is solvable by the linear Kalman-Bucy filtration-smoothing procedure, into that of quadratic programming, which inevitably leads to the necessity of applying much more complicated nonlinear signal processing techniques. However, the three-diagonal kind of the quadratic objective function, on one hand, and the specificity of inequality constraints imposed individually upon each vector variable in the sequence of unknown regression coefficients, on the other, essentially simplify the resulting quadratic programming problem in comparison with its standard form. We call problems of such a kind pair-wise separable quadratic programming problem. Two algorithms of nonstationary regression estimation considered in this paper are built as those of pair-wise separable quadratic programming and have linear computational complexity in contrast to polynomial 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 Kalman-Bucy filter-smoother.

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

1. Introduction
There exists a wide class of signal processing problems in which it is required to estimate the vector parameter $X = (x_t, t = 1, \ldots, N)$ of a nonstationary local model of the observed signal $Y = (y_t, t = 1, \ldots, N)$ at all the time moments. The problems of signal smoothing and nonstationary autoregression analysis are glowing examples of problems of such a kind.

We consider here the problem of nonstationary regression estimation as the canonical mathematical form of a generalized nonstationary signal processing problem which is adequate to many applications. The multidimensional signal to be analyzed $Y = ((y_t, y_i), t = 1, \ldots, N)$ consists of two components, namely, the vector succession $y_t \in \mathbb{R}^n$ and the numerical one $y_i \in \mathbb{R}$. It is assumed that, at each point of observation, the numerical component is a noisy linear function of the vector component $y_{i+1} = x_i^T y_i$ (1) with unknown regression coefficients $x_i \in \mathbb{R}^n$ relatively slowly changing in time. In particular, in the problem of nonstationary autoregression estimation $y_t = \sum_{i=1}^{n} x_i y_{t-i}$, we have $y_t = (y_{t-1} \cdots y_{t-n})^T$.

It is shown in [1] that, if no additional constraints are imposed upon the sought-for regression coefficients, the problem of estimating the hidden succession $X = (x_t, t = 1, \ldots, N)$ can be solved as that of minimizing the objective function

$$J(x_1, \ldots, x_N) = \sum_{i=1}^{N} \psi_i(x_i) + \sum_{i=2}^{N} \gamma_i(x_i, x_{i-1}) \to \min$$

with quadratic constituents

$$\psi_i(x_i) = (y_t - x_i^T y_i)^2 = (x_i - x_i^0)^T Q_i^0 (x_i - x_i^0),$$

$$Q_i^0 = y_i y_i^T, \quad x_i^0 = (y_i - y_{i-1}) Q_{i-1} U_i,$$

$$\gamma_i(x_i, x_{i-1}) = (x_i - V_i x_{i-1})^T U_i (x_i - V_i x_{i-1}),$$

where the choice of positive definite matrices $U_i$ and non-degenerate matrices $V_i$ defines the desired model of the hidden dynamics of the regression coefficients.

The quadratic optimization problem (2)-(4) boils down to that of solving the respective system of linear equations with block tri-diagonal matrix. Such a system can be solved by the usual double sweep procedure, which is completely equivalent to the commonly adopted linear Kalman-Bucy filter-smoother [2].

Both of these two forms of the same procedure are linear from two points of view. First, the procedure consists of a sequence of linear operations of polynomial computational complexity applied to the signal elements and, second, the number of such operations is linear regarding the length $N$ of the signal. The total computational complexity is proportional to $n^3 N$, where $n$ is the dimensionality of an elementary vector variable $x_i \in \mathbb{R}^n$.

In this work, we consider the problem of nonstationary regression estimation under the additional assumption that the ranges of admissible values of the regression coefficients $x_i$ are restricted by linear inequality and equality constraints imposed upon vector variables:

$$A_i x_i + c_i \geq 0, \quad A_i (m \times n), \quad c_i \in \mathbb{R}^m,$$

$$B_i x_i + f_i = 0, \quad B_i (l \times n), \quad f_i \in \mathbb{R}^l.$$  (5)

For instance, in many financial problems the sought-for regression coefficients have the sense of distribution of a resource over several kinds of investments and, so, must be nonnegative and sum up to unity. In the problem of nonstationary autoregression estimation the coefficient vector $x_i$ must belong to the stability range at each time moment.
The presence of linear inequality constraints turns the problem of constrained nonstationary regression estimation into a quadratic programming problem [3] which can no longer be solved by any linear procedure. Besides, the computational complexity of the quadratic programming problem of general kind is proportional to \( n^3N^3 \), i.e. is polynomial with respect to the number of elementary scalar variables, instead of much lesser complexity of the problem without constraints \( n^3N \).

In order to reduce the computational complexity of the constrained nonstationary regression estimation procedure, we systematically exploit the fact that the quadratic criterion (2)-(4) belongs to the class of so-called pair-wise separable functions with chain-like variables adjacency [4], which are named so because of their being sums of elementary functions of not more than two neighboring vector variables in accordance with the accepted ordering \( X = (x_t, t = 1, \ldots, N) \). In addition, the inequality constraints are imposed upon single vector variables individually at each time moment (5).

In contrast to the quadratic programming problems of general kind, we call here the above specified class of optimization problems (2)-(5) the pair-wise separable quadratic programming problems. Any algorithm of pair-wise separable quadratic programming can be immediately used as that of constrained nonstationary regression estimation.

For this specific class of quadratic programming problems, we propose two algorithms, an asymptotically strict iterative algorithm and an approximate one that finds a quasi-optimal solution of the problem in two runs through the signal. Each algorithm is nonlinear from the viewpoint of the kind of operations performed on the signal elements and linear from that of the required number of such operations with respect to the signal length.

For the sake of simplicity of presenting the main ideas, we omit in this brief paper the equality constraints (6), because their presence can easily be taken into account by reducing the dimensionality of the vector variables without any loss in the solvability of the problem by traditional linear methods.

2. Examples of constrained nonstationary regression estimation problems

The problem of nonstationary autoregression estimation. The autoregression equation \( y_t = \sum_{i=1}^{N_t} x_i y_{t-i} + \varepsilon_t = x^T y_t + \varepsilon_t \), where \( \varepsilon_t \) is white noise and \( y_t = (y_{t-1} \cdots y_{t-n})^T \), describes a stationary random process if all the roots of the characteristic polynomial \( D(z) = 1 - x_1 z - x_2 z^2 - \cdots - x_n z^n \) fall outside the unit circle in the complex plane \( |z| > 1 \).

It appears natural to model a nonstationary signal as a stationary one whose properties slowly change in time. Thus, we come to the nonstationary autoregression equation \( y_t = \hat{x}^T y_t + \varepsilon_t \), in which the parameter vector \( \hat{x} \in \mathbb{R}^n \) satisfies the stationarity conditions at each time moment \( t \). This means that the inequality \( |\varepsilon_t| > 1 \) holds for the roots of the “nonstationary” secular polynomial \( D(z) = 1 - x_1 z - x_2 z^2 - \cdots - x_n z^n \) whose coefficients change in time.

In particular, in the case of the second-order autoregression \( n = 2 \), the conditions of the instant stationarity boil down to the system of three linear inequalities [5]

\[
\begin{align*}
x_1 + x_2 &< 1, \\
x_2 - x_1 &< 1, \\
-1 &< x_1 \leq 1,
\end{align*}
\]

that determine a triangle domain in the plane of autoregression coefficients \( \hat{x}_t \in \mathbb{R}^2 \) (Fig. 1) and play the role of inequality constraints (5) in the problem of nonstationary autoregression estimation.

![Figure 1. The domain of admissible values of the autoregression coefficient for the second-order model.](image-url)

The problem of investment portfolio estimation. Investment company is a kind of financial intermediary which attracts capital of private investors and invests it in some classes of financial assets, such as stock, bonds and others securities. The set of financial assets that are in the possession of an investment company at a time moment, usually, a day, is called its investment portfolio. The main goal of any investment company is active managing the portfolio structure with the purpose of increasing the return \( y_t = (W_{t+1} - W_t)/W_t \), where \( W_t \) is the portfolio cost at day \( t \). Any investment company is obliged to publish its daily return, whereas the structure of the portfolio remains its trade secret.

In accordance with the commonly adopted RBSA model (Returns Based Style Analysis) [6,7], the periodic return \( y_t, t = 1, \ldots, N \), of an investment portfolio consisting of a number of assets is approximately represented by a dynamically changing linear combination of single factors \( y_t = (y_1^t \cdots y_n^t)^T \in \mathbb{R}^n \), \( t = 1, \ldots, N \), with real-valued nonnegative weighting coefficients \( x_t = (x_1^{(t)} \cdots x_n^{(t)})^T \), \( t = 1, \ldots, N \) :

\[
y_t = x_t^T y_t + \varepsilon_t, \quad (7)
\]

where \( \varepsilon_t \) is additive noise with zero mean value and unknown variance. The role of single factors \( y_t = (y_1^t \cdots y_n^t)^T \) is played by known returns of the respective classes of assets, and the unknown coefficients \( x_t = (x_1^{(t)} \cdots x_n^{(t)})^T \) are interpreted as numerical expression of the hidden portfolio structure.

The problem consists in estimating the sequence of weighting coefficients \( X = (x_t, t = 1, \ldots, N) \) from the
known time series of the returns \( Y = (y_1, y_2, \ldots, y_t) \) of both the company being investigated \( y_i \) and the assets classes \( y_j \) under the assumption that the coefficients are changing sufficiently smoothly. Since the sought-for variables \( x = (x_1^{(\alpha)}, \ldots, x_N^{(\alpha)})^T \), acting as regression coefficients in the linear regression model (7), have the sense of a distribution of the instant portfolio return over the set of assets classes, they must be nonnegative \( x_i^{(\alpha)} \geq 0 \), \( \ldots, x_j^{(\alpha)} \geq 0 \) and sum up to one \( \sum_{i=1}^n x_i^{(\alpha)} = 1 \).

3. The asymptotically strict iterative steepest descent algorithm

The asymptotically strict iterative algorithm is based on reformulating the original pair-wise separable quadratic programming problem into the dual form regarding the Lagrange multipliers \( \lambda = (\lambda_1^{(\alpha)}, \ldots, \lambda_N^{(\alpha)})^T \in \mathbb{R}^m \), \( \lambda_j^{(\alpha)} \geq 0 \), at inequality constraints in accordance with the Kuhn-Tucker theorem:

\[
L(x_1, \ldots, x_N, \lambda_1, \ldots, \lambda_N) \rightarrow \min_{x_i \in \mathbb{R}^n} \sum_{i=1}^n (x_i - x_i^{(0)})^T Q_i^0 (x_i - x_i^{(0)}) - \lambda_i^T (A_i x_i + c_i) + \sum_{j=1}^N (x_j - V_i x_{j-1})^T U_j (x_j - V_i x_{j-1}).
\]

Let \( x = (x_1^T, \ldots, x_N^T)^T \in \mathbb{R}^{nN} \) and \( \lambda = (\lambda_1^{(\alpha)}, \ldots, \lambda_N^{(\alpha)})^T \in \mathbb{R}^{mN} \) be the combined vectors of, respectively, all the variables and all the Lagrange multipliers, and \( x(\lambda) = \arg \min_{x \in \mathbb{R}^n} L(x, \lambda) \). To solve the initial quadratic programming problem, it is enough to find the maximum point of the dual objective function \( W(\lambda) = L(x(\lambda), \lambda) \) under inequality constraints \( \lambda \geq 0 \). For solving the dual optimization problem, we use the iterative steepest descent method applied to the function \( -W(\lambda) \).

Let \( \lambda^{(k)} \) be the approximation to the optimal vector of Lagrange multipliers at the \( k \) th iteration. To compute the next approximation, we have, first, to find, the gradient of the dual function at the current point \( h^{(k)} = -\nabla_{\lambda} W(\lambda^{(k)}) \). One of the main points of this work is the fact that, in case the signal processing for which an iterative algorithm of model estimation is highly undesirable. These are the problem of jump-point detection in the sequence of signal parameters and that of verifying the nonstationary model by the “leave-on-out” (jackknife) procedure [1]. Algorithmic solution of each of these problems requires multiple recalculating the optimal value of the goal variable \( x_t \) at a single point of the argument axis with different matrices \( U_t \) at this point. Since such a procedure is to be applied to each point \( t = 1, \ldots, N \), the computational complexity of the algorithm as a whole grows proportionally to the squared signal length \( N^2 \) in contrast to the linear computational complexity of the plain estimating the nonstationary model.

The natural procedure for the optimization of pair-wise separable objective functions like (2) is the classical dynamic programming [8,9]. In [1], in addition to the traditional “forward then back” dynamic programming procedure a new “forward against forward” scheme is proposed as a version which allows for easily preserving abrupt changes in signal parameters and is much more preferable from the computational point of view for the model verification by the “jack-knife” procedure.

The central concept of the “forward against forward” scheme of dynamic programming are the so-called left \( \bar{J}_t(x_i) = \min_{x_{i-1}} J_t(x_i, \ldots, x_N) \), and right \( \bar{J}_t^*(x_i) = \min_{x_{i+1}} J_t(x_i, \ldots, x_N) \) Bellman functions, which are related to the left \( J_t(x_i, \ldots, x_N) = \sum_{i=1}^N \psi_i(x_i) + \sum_{i=2}^N \gamma_i(x_{i-1}, x_i) \) and the right \( J_t^*(x_i, \ldots, x_N) = \sum_{i=1}^N \psi_i(x_i) + \sum_{i=1}^{N-1} \gamma_i(x_{i+1}, x_i) \) partial criterions, respectively. The left and right Bellman functions can be recurrently computed from left to right \( t = 1, \ldots, N \) and from right to left, respectively, by the easy-to-prove rules.
\( \tilde{J}_t^-(x_i) = \psi_t(x_i) + \min_{x_{i+1}} \left[ \gamma_t(x_{i+1}, x_i) + \tilde{J}_{i+1}^-(x_{i+1}) \right], \)

\( \tilde{J}_t^+(x_i) = \psi_t(x_i) + \min_{x_{i+1}} \left[ \gamma_t(x_{i+1}, x_i) + \tilde{J}_{i+1}^+(x_{i+1}) \right], \)

starting with \( \tilde{J}_t^-(x_i) = \psi_t(x_i) \) and \( \tilde{J}_N^+(x_i) = \psi_t(x_i). \)

The “meeting” of the left and right Bellman functions at some point \( t \) directly yields the optimal value of the objective variable at this point

\[ \hat{x}_t = \arg \min_{x_t} \left[ \tilde{J}_t^-(x_t) + \tilde{J}_t^+(x_t) - \psi_t(x_t) \right]. \]

Thus, the main idea of dynamic programming consists in replacing the original problem of minimizing the given function of many arguments \( J(x_1, \ldots, x_N) \) by a succession of substantially less difficult problems of minimizing functions of only one variable (9).

The classical dynamic programming procedure is fundamentally based on the assumption that all the variables take values from a finite set, what allows for recomputing the Bellman function (9) and finding the optimal values (10). However, it is shown in [1] that, if all the constituents of the pair-wise separable objective function (2) are quadratic (3), (4), and the inequality constraints (5) are absent, the Bellman functions (9) are also quadratic

\[ \tilde{J}_t^-(x_i) = \tilde{b}_i + (x_i - \hat{x}_i)^T \tilde{Q}_i (x_i - \hat{x}_i), \]

as well as the functions (10) that enable computing the optimal values of variables

\[ \tilde{J}_t^+(x_i) = \tilde{b}_i + (x_i - \hat{x}_i)^T \tilde{Q}_i (x_i - \hat{x}_i). \]

In this case, it is enough to recompute the parameters of Bellman functions \( \tilde{b}_i, \tilde{Q}_i \), and \( \hat{x}_i \). The resulting procedure of unconstrained quadratic optimization turns out to be nothing else than the double sweep procedure of solving the respective system of linear equations or, what is the same, the linear Kalman-Bucy filter applied twice, from left to right and from right to left.

It appears attractive to avoid the iterative search of the optimal decision in the problem of constrained nonstationary regression estimation by way of exploiting the dynamic-programming approach. But in the case when inequality constraints are imposed upon the sought-for regression coefficients (5), the Bellman functions will no longer be quadratic, and the dynamic programming method cannot be applied immediately.

In order to save the computational advantages of the dynamic programming procedure, we resort here to the following trick: we heuristically replace the genuine nonquadratic Bellman functions by some appropriate quadratic approximation to them.

Let the current Bellman functions at point \( t-1 \) and \( t+1 \) be quadratic \( \tilde{J}_{t-1}(x_{t-1}) \) and \( \tilde{J}_{t+1}(x_{t+1}) \). In accordance with (9), the next Bellman functions are to be found from the conditions

\[ \tilde{J}_t^-(x_i) = \psi_t(x_i) + F_t^-(x_i), \]

\[ F_t^-(x_i) = \min_{A_t, x_{i+1}, \xi_{i+1} \geq 0} \left[ \gamma_t(x_{i+1}, x_i) + \tilde{J}_{i+1}^-(x_{i+1}) \right], \]

\[ \tilde{J}_t^+(x_i) = \psi_t(x_i) + F_t^+(x_i), \]

\[ F_t^+(x_i) = \min_{A_t, x_{i+1}, \xi_{i+1} \geq 0} \left[ \gamma_t(x_{i+1}, x_i) + \tilde{J}_{i+1}^+(x_{i+1}) \right]. \]

The first summands \( \psi_t(x_i) \) (3) in \( \tilde{J}_t^-(x_i) \) and \( \tilde{J}_t^+(x_i) \) are always quadratic, but the second ones \( F_t^-(x_i) \) and \( F_t^+(x_i) \) not, because they are obtained as result of minimization of respective quadratic functions under linear inequality constraints.

The heuristic idea consists in replacing the functions \( F_t^-(x_i) \) and \( F_t^+(x_i) \) by appropriate quadratic ones

\[ F_t^-(x_i) = b_t^i + (x_i - x_{i-1})^T Q_t^-(x_i - x_{i-1}) \]

\[ F_t^+(x_i) = b_t^i + (x_i - x_{i-1})^T Q_t^+(x_i - x_{i-1}) \]

from the conditions

\[ x_t^i = \arg \min_{x_t} (x_t - \hat{x}_t^i)^T \tilde{Q}_t^i (x_t - \hat{x}_t^i), \]

\[ b_t^i = \min_{A_t, x_{i+1}, \xi_{i+1} \geq 0} \gamma_t(x_{i+1}, x_i) + \tilde{J}_{i+1}^-(x_{i+1}), \]

and

\[ x_t^i = \arg \min_{x_t} (x_t - \hat{x}_t^i)^T \tilde{Q}_t^+(x_t - \hat{x}_t^i), \]

\[ b_t^i = \min_{A_t, x_{i+1}, \xi_{i+1} \geq 0} \gamma_t(x_{i+1}, x_i) + \tilde{J}_{i+1}^+(x_{i+1}). \]

The computational complexity of this step is proportional to \( n^3 \) and, so, is of the same order as matrix inversion.

The resulting dynamic-programming-based algorithm of constrained nonstationary regression estimation differs from the classical dynamic programming procedure is fundamentally based algorithm only in that it requires solving the quadratic programming problem (11)-(12) at each of \( N \) time moments \( t = 1, \ldots, N \). The presence of this intervening operation makes the procedure nonlinear, but its computational complexity with respect to the length of the signal remains linear.

In contrast to the iterative algorithm, the approximate dynamic-programming-based one does not guarantee finding the strict solution of the problem (2)-(6), and its accuracy is to be examined experimentally.

5. Experimental results

We experimentally compared three methods of solving the problem of constrained nonstationary regression estimation: namely, the direct application of a quadratic programming tool of general kind (Mosek, http://www.mosek.com), the iterative algorithm and the dynamic-programming based algorithm. The results of comparing the computational speed of the three algorithms are shown in Fig. 2.

We did not study analytically the accuracy of the approximate optimization algorithm. Instead, we used simulated data for an experimental comparison of the solutions obtained by different methods \( \hat{x}_t, t = 1, \ldots, N \) with the “true” hidden regression coefficients sequence \( x_t, t = 1, \ldots, N \). We assessed the discrepancy between solutions by the accuracy criterion

\[ d = \sqrt{\sum_{t=1}^N \| x_t - \hat{x}_t \|^2} / \sqrt{\sum_{t=1}^N \| x_t \|^2}. \]
with simulated data, the discrepancy between the estimate obtained by the approximate algorithm and those found by the two strict ones was of the same order as that between two versions of the strict solution (Fig. 3).

![Figure 2](image.png)

**Figure 2.** Comparative dependence of time (seconds) required for solving the problem of constrained nonstationary regression estimation on the length of the signal $N$.

![Figure 3](image.png)

**Figure 3.** Comparison of the regression coefficients $\hat{x}_t$, $t=1,...,N$ estimated by different algorithms and the “true” sequence $x_t$, $t=1,...,N$.

5.2 Style Analysis of the investment portfolio: Fidelity Magellan Fund

The purpose of this section is to demonstrate how nonstationary signal processing methodology developed in previous sections could be employed to real-life mutual fund.

The Fidelity Magellan Fund is a US-domiciled mutual fund from the Fidelity family of funds. It is perhaps the world’s best known actively managed mutual fund. Jeffrey Vinik spent four years managing Magellan and during that time he produced an 83.70% cumulative return and outperformed the cumulative S&P500 return by 5.91%. Vinik is most often remembered as the manager who moved a high percentage of the portfolio out of technology stocks and into bonds and cash at the wrong time, causing Magellan to underperform its peers for the first time in the fund’s history. However, others argue that Vinik merely moved into bonds early. Vinik moved Magellan into bonds in the fall of 1995. In the seven years ending in March 2003, on a total return basis, 10-year Treasuries returned 78 percent, corporate bonds index returned 46 percent, and, with dividends reinvested, the S&P 500 returned 31 percent. In addition, Vinik’s strategy would have avoided the dot com bust.

We used the asymptotically strict iterative steepest descent algorithm produced in section 3 to determine the changes of the bonds and cash values in Fidelity’s portfolio over the period 1994 - 1997.

For our analysis we used 8 indexes provided by Lehman Brothers and Merrill Lynch. The result of the analysis is presented in the Fig. 4. Asset exposures $x_t$ of the fund for each time period are "stacked" along the Y-axis, with the sum equal to 100%.

As you can see the values of bonds (green) and cash (red) are approximately zero before the year 1995 and have maximum in the fall of 1995. The correctness of this analysis is confirmed by publicity sources.

![Figure 4](image.png)

**Figure 4.** Fidelity Magellan Fund: Estimated asset weights

### References


