

LD-BASED PARAMETER IDENTIFICATION APPROACH FOR THE PAIRWISE MARKOV MODELS

Andrey Tsyganov¹

¹Department of Higher Mathematics
Ulyanovsk State Pedagogical University
Russian Federation
tsyganov.av@ulspu.ru

Julia Tsyganova^{2,1}

² Institute of Data Science and Artificial Intelligence
Innopolis University
Russian Federation
y.tsyganova@innopolis.ru

Article history:

Received 24.06.2026, Accepted 20.09.2026

Abstract

This paper addresses the problem of parameter identification in pairwise Markov models, where all system matrices may depend on an unknown vector parameter. A novel, numerically stable approach based on the LD decomposition of the error covariance matrices is proposed. The main theoretical result is formulated as a theorem that justifies the exact gradient evaluation using the LD factorization. Two efficient algorithms are presented: one for computing the identification criterion only, and another for the simultaneous computation of the identification criterion and its gradient. The proposed methods are implemented in MATLAB and validated through Monte Carlo studies. The first experiment considers a canonical pairwise model with a scalar parameter, while the second one demonstrates the identification of the wave speed in the one-dimensional wave equation discretized by a finite-difference scheme. The results confirm that the LD-based gradient-type parameter identification algorithm provides accurate parameter estimates with high numerical robustness. The approach can be applied for adaptive filtering and parameter estimation in distributed parameter systems.

Key words

Pairwise Markov model, pairwise Kalman filter, MWGS orthogonalization, LD factorization, parameter identification.

1 Introduction and problem statement

In statistical signal processing, the main focus is on the problem of estimating an unknown random process $\{x_k\}_{k \in \mathbb{N}}$ based on observed data $\{y_k\}_{k \in \mathbb{N}}$. The classical approach to solving is based on hidden Markov models (HMMs), in which an unobservable sequence is assumed to be Markovian.

In [Pieczynski, 2003], pairwise Markov chains (PMCs) are proposed, which are a generalization of HMMs. The difference between PMCs and classical models is that the Markov property is given not to a separate hidden process, but to a pair $(x; y)$, which makes it possible to describe more complex dependencies between variables. As shown in [Pieczynski and Desbouvries, 2003], the possibility of Kalman filtering remains for pairwise Markov models (PMMs), which led to the development of the pairwise Kalman filter (PKF).

At this moment, the research and development of new computational methods for discrete-time stochastic systems represented by pairwise Markov models remains highly relevant due to the importance of modern practical applications, such as signal and image processing [Derrode and Pieczynski, 2004], image segmentation [Rafi et al., 2011], multi-channel information processing using evidence theory [Pieczynski, 2007], as well as stock index forecasting [Gorynin et al., 2017].

When solving mathematical and computer simulation problems, an additional difficulty arises due to the so-called parameter uncertainty. That is, when the structure of a model is known, but the numerical values of its parameters are unknown or may change abruptly during the operation of the simulated object. In this case, the problem of parameter identification arises [Ljung, 1999], which in practice is usually solved by numerical optimization of a chosen identification criterion. The implementation of gradient-type numerical optimization methods requires the calculation of the gradient of the identification criterion.

It is also well known that in the software implementation of algorithms requiring complex computations, a serious problem is the impact of machine round off errors on the quality of the result [Golub and Van Loan, 2013]. Their accumulation sometimes becomes critical, leading to the loss of numerical stability. It should be

noted that the problem of numerical degradation in classical identification procedures based on the maximum likelihood is well-known and can manifest in various ways, for instance, through the instability of the ML-EM algorithm [Pham et al., 2024], as well as through the tendency to generate incorrect (non-positive definite) covariance matrices during blind estimation of the parameters of pairwise Markov models [Nemesin and Derrode, 2012]. The problem of increasing the numerical stability of the TLS-based identification method is also discussed in [Ivanov et al., 2024]. Therefore, constructing new numerically stable computational methods makes it possible to obtain high-quality solutions to the parameter identification problem of pairwise Markov models.

In this paper, we consider a Gaussian Markov process $\{t_k\}_{k=0}^N$ modeled as a linear pairwise Markov stochastic system [Nemesin and Derrode, 2012] with parameter uncertainty:

$$\begin{bmatrix} x_{k+1} \\ y_k \end{bmatrix} = \underbrace{\begin{bmatrix} F_{x,x}(\theta) & F_{x,y}(\theta) \\ F_{y,x}(\theta) & F_{y,y}(\theta) \end{bmatrix}}_{\mathcal{F}(\theta)} \underbrace{\begin{bmatrix} x_k \\ y_{k-1} \end{bmatrix}}_{t_k} + \underbrace{\begin{bmatrix} w_k^x \\ w_k^y \end{bmatrix}}_{w_k} \quad (1)$$

where $\theta \in \mathbb{R}^p$ is an unknown model parameter, $x_k \in \mathbb{R}^{n_x}$ is the hidden (unobservable) state vector, $y_k \in \mathbb{R}^{n_y}$ are the available measurements, $\{w_k\}$ is a sequence of normally distributed random vectors of dimension $n_x + n_y$ with zero means and covariance matrix

$$\mathcal{Q}(\theta) = \begin{bmatrix} Q_{x,x}(\theta) & Q_{x,y}(\theta) \\ Q_{x,y}^T(\theta) & Q_{y,y}(\theta) \end{bmatrix} > 0. \quad (2)$$

The initial state $x_0 \sim \mathcal{N}(\bar{x}_0(\theta), P_0(\theta))$ is independent of $\{w_k\}$, $P_0(\theta) > 0$. The parameter uncertainty of model (1), (2) means that all elements of matrices $\mathcal{F}(\theta)$, $\mathcal{Q}(\theta)$, $P_0(\theta) > 0$, as well as vector $\bar{x}_0(\theta)$, may depend on the unknown parameter θ .

We formulate the parameter identification problem for the pairwise Markov model (1), (2), which consists in computing estimates of parameter θ from the available measurements y_k , $k = 0, 1, \dots, N$.

One common approach to solving this problem is to apply numerical methods for constrained minimization of an identification criterion of the form

$$\hat{\theta}_{\min} = \underset{\theta \in D(\theta)}{\operatorname{argmin}} J(\theta, Y_0^N) \quad (3)$$

where θ is the unknown parameter to be identified, $D(\theta)$ is the appropriate domain for θ , and $Y_0^N = \{y_0, \dots, y_N\}$ is the available measurement information [Gibbs, 2011]. The identification criterion (3) is chosen depending on the class of the stochastic system considered [Ljung, 1999]. The domain $D(\theta)$ of admissible parameter values is determined by the convergence conditions of estimates when solving the optimization problem (3). For discrete-time linear stochastic systems, the general convergence conditions of gradient-type identification methods are well known [Ljung, 1978].

To solve the parameter identification problem in the class of pairwise Markov models, we consider an identification criterion based on the negative log-likelihood function:

$$J_{\text{PKF}}(\theta, Y_0^N) = \frac{Nn_y}{2} \ln(2\pi) + \frac{1}{2} \sum_{k=1}^N \left[\ln \det R_{e,k}(\theta) + \|e_k(\theta)\|_{R_{e,k}^{-1}(\theta)}^2 \right] \quad (4)$$

where the innovation vector $e_k(\theta)$ and its covariance matrix $R_{e,k}(\theta)$ are obtained by the pairwise Kalman filter [Pieczynski, 2003; Pieczynski and Desbouvries, 2003; Kulikova and Tsyganova, 2017], which unlike the conventional Kalman filter [Grewal and Andrews, 2015], accounts for the pairwise Markov property. Here we present the pairwise Kalman filter as Algorithm 1.

Algorithm 1. Conventional Pairwise Kalman Filter (PKF).

Input: $\mathcal{F}$, $\mathcal{Q}$, $\bar{x}_0$, P_0 , $\{y_0, \dots, y_N\}$.

Output: Estimates $\{\hat{x}_{k|k}\}_{k=1}^N$ and covariance matrices $\{P_{k|k}\}_{k=1}^N$.

1. Initialization:

$$\hat{F}_{x,x} = F_{x,x} - Q_{x,y} Q_{y,y}^{-1} F_{y,x}, \quad (5)$$

$$\hat{F}_{x,y} = F_{x,y} - Q_{x,y} Q_{y,y}^{-1} F_{y,y}, \quad (6)$$

$$\hat{Q}_{x,x} = Q_{x,x} - Q_{x,y} Q_{y,y}^{-1} Q_{y,x}, \quad (7)$$

$$\hat{x}_{0|0} = \bar{x}_0, \quad P_{0|0} = P_0. \quad (8)$$

2. Main loop for $k = 0$ to $N - 1$

2.1. Time update:

$$\hat{x}_{k+1|k} = \hat{F}_{x,x} \hat{x}_{k|k} + Q_{x,y} Q_{y,y}^{-1} y_k + \hat{F}_{x,y} y_{k-1} \quad (9)$$

(for $k = 0$ the term y_{-1} is omitted),

$$P_{k+1|k} = \hat{Q}_{x,x} + \hat{F}_{x,x} P_{k|k} \hat{F}_{x,x}^T. \quad (10)$$

2.2. Measurement update:

$$e_{k+1} = y_{k+1} - F_{y,x} \hat{x}_{k+1|k} - F_{y,y} y_k, \quad (11)$$

$$R_{e,k+1} = Q_{y,y} + F_{y,x} P_{k+1|k} F_{y,x}^T, \quad (12)$$

$$K_{k+1} = P_{k+1|k} F_{y,x}^T R_{e,k+1}^{-1}, \quad (13)$$

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1} e_{k+1}, \quad (14)$$

$$P_{k+1|k+1} = P_{k+1|k} - K_{k+1} R_{e,k+1} K_{k+1}^T. \quad (15)$$

3. End

Remark 1. In Algorithm 1, for simplicity of presentation we do not explicitly indicate the possible dependence of the model matrices on parameter θ .

For the constrained minimization of criterion (4), a gradient-based approach is efficient, which requires the computation of the gradient of criterion (4) at each step of the parameter identification algorithm [Gibbs, 2011]. The update of the unknown parameter θ estimate in the standard first-order gradient-type method is performed according to the iterative expression

$$\theta^{j+1} = \theta^j - \beta_j \nabla_{\theta} J(\theta^j, Y_0^N) \quad (16)$$

where θ^j is the estimate of the unknown parameter at the j -th iteration of the algorithm. In (16), ∇_{θ} denotes the gradient operator $[\partial/\partial\theta_1 \mid \cdots \mid \partial/\partial\theta_p]^T$, which is applied to (4) at $\theta = \theta^j$. The scalar step parameter β_j is chosen so as to satisfy $J(\theta^{j+1}, Y_0^N) \leq J(\theta^j, Y_0^N) + \varepsilon$, where ε is a positive number that can be selected in various ways depending on the particular minimization method [Nocedal and Wright, 2006].

For the pairwise Markov model (1), (2), the computation of the gradient $\nabla_{\theta} J(\theta, Y_0^N)$ requires the implementation of the pairwise Kalman filter, augmented by so-called “sensitivity equations”. These equations define how to calculate the derivatives of the state estimate $\hat{x}_k$ and the error covariance matrix P_k with respect to θ .

The described approach to solving the parameter identification problem makes it possible to compute estimates of the unknown parameter θ simultaneously with the estimation of the hidden state vector x_k of the pairwise Markov model. The quality of the obtained solution can be improved by developing numerically efficient methods for computing both the identification criterion itself and its gradient. By numerical efficiency we primarily mean the stability of the computational scheme with respect to machine round off errors. It is known that, for discrete-time stochastic systems with additive Gaussian noise, this property is fully ensured by an approach to constructing discrete filtering algorithms based on modified weighted Gram–Schmidt orthogonalization (so-called MWGS orthogonalization) [Bierman, 1977; Grewal and Andrews, 2015].

The main goal of our work is to develop a new approach to solving the parameter identification problem in the class of pairwise Markov models based on the use of modified weighted Gram–Schmidt orthogonalization and LD decomposition of parameterized matrices.

Achieving this goal involves finding solutions to the following sub-problems:

1. Construction of the identification criterion based on the pairwise LD filter.
2. Development of a new algorithm for computing the derivatives with respect to the model parameter θ in the pairwise LD filter based on original methods for differentiating parameterized matrices.
3. Construction of a new method for computing the gradient of the identification criterion based on the

pairwise LD filter.

Furthermore, in this work we demonstrate the application of the developed approach to solving the problem of parameter identification in a wave equation model using noisy measurements.

The following sections describe in detail the solutions obtained.

2 Pairwise filtering algorithms

Algorithm 1 is the “conventional” pairwise Kalman filter, as it has a similar structure to the conventional Kalman filter for hidden Markov models.

It is easy to see that Algorithm 1, like the conventional Kalman filter, involves matrix subtraction and inversion operations, which can affect the numerical stability of its computer implementation in the presence of round off errors. In particular, if matrices $R_{e,k}$ become singular or nearly singular, or if the error covariance matrix $P_{k|k}$ loses positive definiteness, the algorithm may fail and the estimates diverge.

To prevent numerical instability mentioned above, there exist methods based on orthogonal matrix transformations that allow the construction of numerically stable modifications of discrete filtering algorithms [Grewal and Andrews, 2015; Kailath et al., 2000]. For discrete-time stochastic systems with additive Gaussian noises, such methods are well known.

For the class of pairwise Markov models, numerically stable modifications of Kalman-type pairwise filtering algorithms were previously proposed in [Nemesin and Derrode, 2012; Kulikova, 2017; Kulikova and Tsyganova, 2017; Kulikova et al., 2020]. One efficient approach to constructing such algorithms is the use of MWGS orthogonalization. This method was first proposed in [Thornton, 1976] for computing the UD factors of the error covariance matrix during the time update in Bierman’s UD algorithm [Bierman, 1977].

It should be noted that there are two types of decomposition of a rectangular matrix using the MWGS method: the LD decomposition (using the forward MWGS procedure) and the UD decomposition (using the backward MWGS procedure).

For pairwise Markov models (1), (2), the problem of constructing UD-based pairwise filtering and parameter identification algorithms was solved in [Nemesin and Derrode, 2012; Kulikova and Tsyganova, 2017; Kulikova et al., 2020]. However, the LD versions of such algorithms have not been presented until now.

Extending the theory of pairwise filtering and complementing the class of parameter identification methods for pairwise Markov models, we propose in this paper a new approach to constructing gradient methods for parameter identification of such models with parameter uncertainty. This approach is based on the numerically stable modified weighted Gram–Schmidt orthogonalization method using the LD decomposition.

Consider the numerically stable LD modification of the pairwise Kalman filter, which uses the LD decomposition of the error covariance matrices in the form $P = L_P D_P L_P^T$, where L_P is a unit lower triangular matrix and $D_P > 0$ is a diagonal matrix.

In [Tsyganov, 2026], the pairwise LD filter was first proposed. The main idea behind the LD-based pairwise filtering algorithm is to transform block pre-arrays: $\mathcal{A}$ (rectangular block matrix) and $\mathcal{D}_{\mathcal{A}}$ (block diagonal matrix). Applying the forward MWGS orthogonalization procedure yields block post-arrays: $\mathcal{L}$ (lower triangular block matrix) and $\mathcal{D}_{\mathcal{L}}$ (block diagonal matrix) such that

$$\mathcal{A}^T = \mathcal{L} \mathcal{B}^T \quad \text{and} \quad \mathcal{A}^T \mathcal{D}_{\mathcal{A}} \mathcal{A} = \mathcal{L} \mathcal{D}_{\mathcal{L}} \mathcal{L}^T \quad (17)$$

where $\mathcal{B}$ is the MWGS transformation matrix.

We denote by MWGS-LD the forward modified weighted Gram–Schmidt orthogonalization procedure [Tsyganov and Tsyganova, 2025b]. Algorithm 2 presents the pairwise LD filter.

Algorithm 2. LD-based Pairwise Kalman Filter (LD-PKF).

Input: $\mathcal{F}$, $\mathcal{Q}$, $\bar{x}_0$, P_0 , $\{y_0, \dots, y_N\}$.

Output: Estimates $\{\hat{x}_{k|k}\}_{k=1}^N$ and LD factors $\{L_{P_{k|k}}, D_{P_{k|k}}\}_{k=1}^N$.

1. Initialization: Compute the auxiliary matrices as in the conventional PKF. Using modified Cholesky decomposition factorize: $P_0 = L_{P_0} D_{P_0} L_{P_0}^T$, $\hat{Q}_{x,x} = L_Q D_Q L_Q^T$, $Q_{y,y} = L_R D_R L_R^T$. Set $\hat{x}_{0|0} = \bar{x}_0$, $L_{P_{0|0}} = L_{P_0}$, $D_{P_{0|0}} = D_{P_0}$.

2. Main loop for $k = 0$ to $N - 1$

2.1. Time update:

$$\hat{x}_{k+1|k} = \hat{F}_{x,x} \hat{x}_{k|k} + Q_{x,y} Q_{y,y}^{-1} y_k + \hat{F}_{x,y} y_{k-1}. \quad (18)$$

(For $k = 0$ the term y_{-1} is omitted.)

Form the pre-arrays:

$$\mathcal{A}_{\text{pred}} = \begin{bmatrix} L_{P_{k|k}}^T & \hat{F}_{x,x}^T & L_Q^T \end{bmatrix}^T, \quad \mathcal{D}_{\text{pred}} = \begin{bmatrix} D_{P_{k|k}} & 0 \\ 0 & D_Q \end{bmatrix}. \quad (19)$$

Compute the LD factors of the predicted covariance matrix:

$$[L_{P_{k+1|k}}, D_{P_{k+1|k}}] = \text{MWGS-LD}(\mathcal{A}_{\text{pred}}, \mathcal{D}_{\text{pred}}). \quad (20)$$

2.2. Measurement update: Form the pre-arrays

$$\mathcal{A}_{\text{upd}} = \begin{bmatrix} L_R & F_{y,x} L_{P_{k+1|k}} \end{bmatrix}^T, \quad \mathcal{D}_{\text{upd}} = \begin{bmatrix} D_R & 0 \\ 0 & D_{P_{k+1|k}} \end{bmatrix}. \quad (21)$$

Compute the joint LD factors

$$[\mathcal{L}, \mathcal{D}] = \text{MWGS-LD}(\mathcal{A}_{\text{upd}}, \mathcal{D}_{\text{upd}}). \quad (22)$$

Extract the required matrices

$$L_{R_{e,k+1}} = \mathcal{L}(1 : n_y, 1 : n_y), \quad (23)$$

$$D_{R_{e,k+1}} = \mathcal{D}(1 : n_y, 1 : n_y), \quad (24)$$

$$L_{P_{k+1|k+1}} = \mathcal{L}(n_y + 1 : \text{end}, n_y + 1 : \text{end}), \quad (25)$$

$$D_{P_{k+1|k+1}} = \mathcal{D}(n_y + 1 : \text{end}, n_y + 1 : \text{end}), \quad (26)$$

$$(K_{k+1} L_{R_{e,k+1}}) = \mathcal{L}(n_y + 1 : \text{end}, 1 : n_y). \quad (27)$$

Compute the normalized innovations

$$\bar{e}_{k+1} = L_{R_{e,k+1}}^{-1} e_{k+1} \quad (28)$$

where

$$e_{k+1} = y_{k+1} - F_{y,x} \hat{x}_{k+1|k} - F_{y,y} y_k. \quad (29)$$

Update the state estimate

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + (K_{k+1} L_{R_{e,k+1}}) \bar{e}_{k+1}. \quad (30)$$

Store $\hat{x}_{k+1|k+1}$ and LD factors $L_{P_{k+1|k+1}}$, $D_{P_{k+1|k+1}}$.

3. End

Remark 2. Using (17), one can show that the PKF and LD-PKF algorithms are algebraically equivalent.

The advantage of the proposed LD modification of the pairwise filtering algorithm is its block-matrix form, which allows updating the LD factors of the error covariance matrices using the numerically stable MWGS orthogonalization procedure. In particular, it was demonstrated in [Tsyganov et al., 2023] that the forward MWGS procedure and the LD filters based on it have a computational advantage over the backward procedure and the previously developed UD filters. The same advantages hold for the numerically stable modifications of the pairwise Kalman filter.

Now we proceed to describe our approach for solving the problem of parameter identification of pairwise Markov models, which is based on the numerically stable LD modification of the pairwise Kalman filter.

3 Main results

Our main results present new solutions to the three interrelated problems listed in the first section of this paper. Together, they form an approach to constructing gradient-type parameter identification algorithms based on the modified weighted Gram–Schmidt orthogonalization and pairwise Kalman-type filtering.

To solve the first problem (see Section 1), we construct an algorithm for computing the identification criterion in terms of the pairwise LD filter.

To solve the second problem, we apply the original method for computing derivatives in the LD decomposition of parameterized matrices [Tsyganova and Tsyganov, 2018], which allows computing the partial derivatives of the LD factors. The corresponding formulas are added to Algorithm 2.

To solve the third problem, we develop a new method for computing the gradient of the identification criterion in terms of the “differentiated” pairwise LD filter.

3.1 Construction of the identification criterion in terms of the pairwise LD filter

To formulate the identification criterion in terms of Algorithm 2, we need to express the quantities $\det R_{e,k}(\theta)$ and $\|e_k(\theta)\|_{R_{e,k}^{-1}(\theta)}^2$ appearing in (4) through the quantities available in the pairwise LD filter.

Proposition 1. *The identification criterion (4) expressed in terms of the pairwise LD filter (Algorithm LD-PKF) takes the form*

$$J_{LD-PKF}(\theta, Y_0^N) = \frac{Nn_y}{2} \ln(2\pi) + \frac{1}{2} \sum_{k=1}^N \left[\ln \det D_{R_{e,k}}(\theta) + \|\bar{e}_k(\theta)\|_{D_{R_{e,k}}^{-1}(\theta)}^2 \right] \quad (31)$$

where the right-hand side of (31) contains quantities that are directly available in the LD-PKF algorithm at each step of its execution.

Proof. Let us show that

$$\det R_{e,k}(\theta) = \det D_{R_{e,k}}(\theta), \quad (32)$$

$$\|e_k(\theta)\|_{R_{e,k}^{-1}(\theta)}^2 = \|\bar{e}_k(\theta)\|_{D_{R_{e,k}}^{-1}(\theta)}^2. \quad (33)$$

For simplicity, we omit the dependence on parameter θ .

First, we prove (32). From (17) we have

$$R_{e,k} = L_{R_{e,k}} D_{R_{e,k}} L_{R_{e,k}}^T.$$

Hence,

$$\det R_{e,k} = \det \left(L_{R_{e,k}} D_{R_{e,k}} L_{R_{e,k}}^T \right) = \det D_{R_{e,k}},$$

because $L_{R_{e,k}}$ is the triangular matrix with a unit diagonal.

Now rewrite (33) in the equivalent form:

$$e_k^T R_{e,k}^{-1} e_k = \bar{e}_k^T D_{R_{e,k}}^{-1} \bar{e}_k. \quad (34)$$

Equality (34) holds due to (28) and the expression for the inverse matrix:

$$R_{e,k}^{-1} = L_{R_{e,k}}^{-T} D_{R_{e,k}}^{-1} L_{R_{e,k}}^{-1}.$$

Thus Proposition 1 is proved. $\square$

Using this result, we construct Algorithm 3, which computes the value of the identification criterion (31) for a given parameter θ .

Algorithm 3. Computation of $J_{LD-PKF}(\theta, Y_0^N)$.

Input: $\mathcal{F}(\theta)$, $\mathcal{Q}(\theta)$, $\bar{x}_0(\theta)$, $P_0(\theta)$, $\{y_0, \dots, y_N\}$.

Output: Value of $J_{LD-PKF}(\theta, Y_0^N)$.

1. Initialization: Compute $\hat{F}_{x,x}$, $\hat{F}_{x,y}$, $\hat{Q}_{x,x}$ according to (5)–(7). Using modified Cholesky decomposition factorize: $P_0 = L_{P_0} D_{P_0} L_{P_0}^T$, $\hat{Q}_{x,x} = L_Q D_Q L_Q^T$, $Q_{y,y} = L_R D_R L_R^T$. Set $\hat{x}_{0|0} = \bar{x}_0$, $L_{P_{0|0}} = L_{P_0}$, $D_{P_{0|0}} = D_{P_0}$, $J_{LD-PKF} = 0$.

2. Main loop for $k = 0$ to $N - 1$

2.1. Time update: Apply equations (18)–(20).

2.2. Measurement update: Apply equations (21)–(30). Compute the contribution to J_{LD-PKF} :

$$\Delta J_{LD-PKF} = \sum_{i=1}^{n_y} \left(\ln(D_{R_{e,k+1}})_{ii} + \frac{(\bar{e}_{k+1})_i^2}{(D_{R_{e,k+1}})_{ii}} \right), \quad (35)$$

$$J_{LD-PKF} = J_{LD-PKF} + \Delta J_{LD-PKF}. \quad (36)$$

3. End

$$J_{LD-PKF} = J_{LD-PKF} + \frac{Nn_y}{2} \ln(2\pi). \quad (37)$$

3.2 Computing derivatives with respect to the model parameter θ in the pairwise LD filter

Computing the gradient of the identification criterion (31) requires knowledge of the derivatives of the quantities that depend on θ , namely $L_{R_{e,k}}(\theta)$, $D_{R_{e,k}}(\theta)$, $\bar{e}_k(\theta)$.

To compute them we apply the original method for computing derivatives in the matrix orthogonal LD transformation [Tsyganova and Tsyganov, 2018], which is based on the forward MWGS orthogonalization procedure.

Denote $A \triangleq A(\theta)$ and $D_w \triangleq D_w(\theta)$. As mentioned above, the meaning of the LD decomposition is to use the forward modified weighted Gram–Schmidt orthogonalization to compute the pair of matrices $\{\bar{L}, D_\beta\}$ from a given pair $\{A, D_w\}$ such that

$$A^T = \bar{L} B^T \quad \text{and} \quad A^T D_w A = \bar{L} D_\beta \bar{L}^T \quad (38)$$

where $A \in \mathbb{R}^{r \times s}$, $B \in \mathbb{R}^{r \times s}$ is the MWGS transformation matrix, $\bar{L} \in \mathbb{R}^{s \times s}$ is a lower triangular matrix with unit diagonal ($r \geq s$). The diagonal matrices $D_w \in \mathbb{R}^{r \times r}$ and $D_\beta \in \mathbb{R}^{s \times s}$ satisfy $D_w > 0$ and $B^T D_w B = D_\beta$.

Lemma 1. [Tsyganova and Tsyganov, 2018] Suppose the entries of matrices $\{A, D_w\}$ are differentiable functions of a scalar parameter θ . Consider the transformation (38). Given the derivative matrices A'_θ and $(D_w)'_\theta$, the derivatives of $\{\bar{L}, D_\beta\}$ at $\theta = \hat{\theta}$ can be computed as

$$\bar{L}'_\theta = \bar{L} (\bar{\mathcal{L}}_0 + \bar{\mathcal{L}}_2 + \bar{U}_0^T) D_\beta^{-1}, (D_\beta)'_\theta = 2D_0 + D_2 \quad (39)$$

where $\bar{\mathcal{L}}_0, D_0, \bar{U}_0$ are, respectively, the strictly lower triangular, diagonal and strictly upper triangular parts of the matrix product $B^T D_w A'_\theta \bar{L}^{-T}$. The matrices D_2 and $\bar{\mathcal{L}}_2$ are, respectively, the diagonal and strictly lower triangular parts of the matrix product $B^T (D_w)'_\theta B$.

The algorithm MWGS_DIFF_LD for computing derivatives in the matrix LD transformation (38) is also presented in [Tsyganova and Tsyganov, 2018]:

$$\begin{aligned} & \left[\mathcal{L}, \mathcal{D}_\mathcal{L}, \frac{d\mathcal{L}}{d\theta}, \frac{d\mathcal{D}_\mathcal{L}}{d\theta} \right] \\ &= \text{MWGS_DIFF_LD} \left(\mathcal{A}, \frac{d\mathcal{A}}{d\theta}, \mathcal{D}_\mathcal{A}, \frac{d\mathcal{D}_\mathcal{A}}{d\theta} \right) \end{aligned}$$

where its input and output parameters are:

$\mathcal{A}$ is the pre-array (block matrix),

$\frac{d\mathcal{A}}{d\theta}$ is the pre-array of derivative values (block matrix),

$\mathcal{D}_\mathcal{A}$ is the weight pre-array (block diagonal matrix),

$\frac{d\mathcal{D}_\mathcal{A}}{d\theta}$ is the weight pre-array of derivative values (block diagonal matrix),

$\mathcal{L}$ is the post-array (block lower triangular matrix with unit diagonal),

$\mathcal{D}_\mathcal{L}$ is the weight post-array (block diagonal matrix),

$\frac{d\mathcal{L}}{d\theta}$ is the post-array of derivative values (block lower triangular matrix),

$\frac{d\mathcal{D}_\mathcal{L}}{d\theta}$ is the weight post-array of derivative values (block diagonal matrix).

An undoubted advantage of the algorithm is its homogeneous structure, which consists in applying the generalized operation MWGS_DIFF_LD to pairs of pre-arrays (block matrices). Consequently, its software implementation is very simple and does not depend on the particular matrices defining the pairwise Markov model (1), (2). A second advantage is that it simultaneously computes both the LD factors of the block matrices and the derivatives of their entries at a given point θ . It should also be noted that the computation of derivatives in this case is performed algorithmically and does not require numerical differentiation, thus avoiding all the drawbacks inherent in that class of methods. In the case of a vector parameter $\theta = [\theta_1, \dots, \theta_p]^T$, the algorithm MWGS_DIFF_LD allows computing the derivatives for each $\theta_i, i = 1, \dots, p$.

The computational complexity depends on the order of the model (1), (2) and on the size p of the vector parameter θ . An analysis and complexity estimates of the MWGS-based LD decomposition and the algorithms based on it can be found in [Tsyganova and Tsyganov,

2018; Tsyganova et al., 2020]. In particular, [Tsyganova and Tsyganov, 2018] provides experimental execution time estimates of the MWGS_DIFF_LD algorithm in MATLAB.

It is also worth noting that the algorithm requires only one matrix MWGS decomposition, which bears the main computational load. The remaining part consists of addition and multiplication operations with diagonal and triangular matrices. Only one trivial inversion of a diagonal matrix and a simple inversion of a triangular matrix are needed.

In the next section, the proposed algorithm will be used to compute the gradient of the new identification criterion (31).

3.3 Computing the gradient of the identification criterion in terms of the “differentiated” pairwise LD filter

Here we propose a new method for computing the gradient of the identification criterion (31) in terms of the LD-PKF algorithm. Let us describe the obtained result.

Theorem 1. Let the entries of the matrices defining the pairwise Markov model (1), (2) depend on a parameter $\theta \in D(\theta)$, where $\theta \in \mathbb{R}^p$ and the entries are differentiable functions of θ . Then the gradient of the identification criterion (31) expressed in terms of the pairwise LD filter has the form

$$\begin{aligned} \nabla_\theta J_{LD-PKF}(\theta, Y_0^N) &= [\partial J_{LD-PKF}(\theta, Y_0^N) / \partial \theta_1 \mid \\ &\dots \mid \partial J_{LD-PKF}(\theta, Y_0^N) / \partial \theta_p]^T \quad (40) \end{aligned}$$

where for $i = 1, \dots, p$

$$\begin{aligned} \frac{\partial}{\partial \theta_i} J_{LD-PKF}(\theta, Y_0^N) &= \frac{1}{2} \sum_{k=1}^N \left\{ \frac{\partial}{\partial \theta_i} (\ln \det D_{R_{e,k}}(\theta)) \right. \\ &\quad \left. + \frac{\partial}{\partial \theta_i} \left(\|\bar{e}_k(\theta)\|_{D_{R_{e,k}}^{-1}(\theta)}^2 \right) \right\}, \quad (41) \end{aligned}$$

$$\frac{\partial}{\partial \theta_i} (\ln \det D_{R_{e,k}}(\theta)) = \text{tr} \left(D_{R_{e,k}}^{-1}(\theta) \frac{\partial D_{R_{e,k}}(\theta)}{\partial \theta_i} \right), \quad (42)$$

$$\begin{aligned} \frac{\partial}{\partial \theta_i} \|\bar{e}_k(\theta)\|_{D_{R_{e,k}}^{-1}(\theta)}^2 &= 2\bar{e}_k^T(\theta) D_{R_{e,k}}^{-1}(\theta) \frac{\partial \bar{e}_k(\theta)}{\partial \theta_i} \\ &\quad - \bar{e}_k^T(\theta) D_{R_{e,k}}^{-2}(\theta) \left(\frac{\partial D_{R_{e,k}}(\theta)}{\partial \theta_i} \right) \bar{e}_k(\theta). \quad (43) \end{aligned}$$

Proof. Consider the criterion (31). Applying the differentiation operator with respect to θ_i yields the general expression (41).

Expression (42) follows from the rule for differentiating a matrix expression

$$\frac{\partial}{\partial \theta} \ln A(\theta) = \text{tr} \left(A^{-1}(\theta) \frac{\partial A(\theta)}{\partial \theta} \right),$$

where $A(\theta)$ is a square nonsingular matrix, and $\text{tr}(\cdot)$ denotes the trace of a matrix.

Expression (43) is obtained using the rules of vector-matrix differentiation, taking into account that the matrices $D_{R_{e,k}}^{-1}(\theta)$ and $\frac{\partial D_{R_{e,k}}(\theta)}{\partial \theta_i}$ are diagonal.

The values of $\frac{\partial \bar{e}_k(\theta)}{\partial \theta_i}$ and $\frac{\partial D_{R_{e,k}}(\theta)}{\partial \theta_i}$ appearing in (42) and (43) are computed by direct differentiation of (28) and according to Lemma 1 for the current parameter value $\theta = \theta_i, i = 1, \dots, p$.

Thus, theorem 1 is proved. $\square$

Based on these theoretical results, we develop an algorithm for the simultaneous computation of the identification criterion (31) and its gradient (40) in terms of the pairwise LD filter. For clarity, we assume $\theta \in \mathbb{R}^1$. All computations easily scale to the vector case $\theta \in \mathbb{R}^p$.

We present the obtained practical result as Algorithm 4.

Algorithm 4. Simultaneous computation of $J_{\text{LD-PKF}}(\theta, Y_0^N)$ and $\frac{dJ_{\text{LD-PKF}}(\theta, Y_0^N)}{d\theta}$ using LD pairwise filter.

Input: $\mathcal{F}(\theta), \mathcal{Q}(\theta), \bar{x}_0(\theta), P_0(\theta)$ and their derivatives $\frac{d\mathcal{F}}{d\theta}, \frac{d\mathcal{Q}}{d\theta}, \frac{d\bar{x}_0}{d\theta}, \frac{dP_0}{d\theta}, \{y_0, \dots, y_N\}$.

Output: $J_{\text{LD-PKF}}(\theta, Y_0^N), \frac{dJ_{\text{LD-PKF}}(\theta, Y_0^N)}{d\theta}$.

1. Initialization: Compute $\hat{F}_{x,x}, \hat{F}_{x,y}, \hat{Q}_{x,x}$ and their derivatives using (5)–(7). Set $S = Q_{x,y}, R = Q_{y,y}, \hat{S} = SR^{-1}, \frac{d\hat{S}}{d\theta} = \frac{dS}{d\theta}R^{-1} - SR^{-1}\frac{dR}{d\theta}R^{-1}$. Using modified Cholesky decomposition to factorize: $\hat{Q}_{x,x} = L_Q D_Q L_Q^T, R = L_R D_R L_R^T, P_0 = L_P D_P L_P^T$, and similarly for derivatives: $\frac{d\hat{Q}_{x,x}}{d\theta}, \frac{dR}{d\theta}, \frac{dP_0}{d\theta}$. Set $\hat{x}_{0|0} = \bar{x}_0, \frac{d\hat{x}_{0|0}}{d\theta} = \frac{d\bar{x}_0}{d\theta}, J = 0, \frac{dJ}{d\theta} = 0$.

2. Main loop for $k = 0, \dots, N - 1$

2.1. Time update

Compute the predicted state estimate and its derivative:

$$\hat{x}_{k+1|k} = \hat{F}_{x,x}\hat{x}_{k|k} + \hat{S}y_k + \hat{F}_{x,y}y_{k-1}, \quad (44)$$

$$\begin{aligned} \frac{d\hat{x}_{k+1|k}}{d\theta} &= \frac{d\hat{F}_{x,x}}{d\theta}\hat{x}_{k|k} + \hat{F}_{x,x}\frac{d\hat{x}_{k|k}}{d\theta} \\ &\quad + \frac{d\hat{S}}{d\theta}y_k + \frac{d\hat{F}_{x,y}}{d\theta}y_{k-1}. \end{aligned} \quad (45)$$

Construct the pre-arrays and their derivatives:

$$\mathcal{A}_{\text{pred}} = \begin{bmatrix} L_{P_{k|k}}^T & \hat{F}_{x,x}^T & L_Q^T \end{bmatrix}^T, \quad (46)$$

$$\frac{d\mathcal{A}_{\text{pred}}}{d\theta} = \begin{bmatrix} \frac{dL_{P_{k|k}}^T}{d\theta}\hat{F}_{x,x}^T + L_{P_{k|k}}^T \frac{d\hat{F}_{x,x}^T}{d\theta} & \frac{dL_Q^T}{d\theta} \end{bmatrix}^T, \quad (47)$$

$$\mathcal{D}_{\text{pred}} = \begin{bmatrix} D_{P_{k|k}} & 0 \\ 0 & D_Q \end{bmatrix}, \quad \frac{d\mathcal{D}_{\text{pred}}}{d\theta} = \begin{bmatrix} \frac{dD_{P_{k|k}}}{d\theta} & 0 \\ 0 & \frac{dD_Q}{d\theta} \end{bmatrix}. \quad (48)$$

Compute the LD factors and their derivatives:

$$\begin{aligned} &\left[L_{P_{k+1|k}}, D_{P_{k+1|k}}, \frac{dL_{P_{k+1|k}}}{d\theta}, \frac{dD_{P_{k+1|k}}}{d\theta} \right] \\ &= \text{MWGS_DIFF_LD}\left(\mathcal{A}_{\text{pred}}, \frac{d\mathcal{A}_{\text{pred}}}{d\theta}, \mathcal{D}_{\text{pred}}, \frac{d\mathcal{D}_{\text{pred}}}{d\theta}\right). \end{aligned}$$

2.2. Measurement update

Compute innovation and its derivative:

$$e_{k+1} = y_{k+1} - F_{y,x}\hat{x}_{k+1|k} - F_{y,y}y_k, \quad (49)$$

$$\begin{aligned} \frac{de_{k+1}}{d\theta} &= -\frac{dF_{y,x}}{d\theta}\hat{x}_{k+1|k} - F_{y,x}\frac{d\hat{x}_{k+1|k}}{d\theta} \\ &\quad - \frac{dF_{y,y}}{d\theta}y_k. \end{aligned} \quad (50)$$

Construct the pre-arrays and their derivatives:

$$\begin{aligned} \mathcal{A}_{\text{upd}} &= \begin{bmatrix} L_R F_{y,x} L_{P_{k+1|k}} \\ 0 & L_{P_{k+1|k}} \end{bmatrix}, \\ \frac{d\mathcal{A}_{\text{upd}}}{d\theta} &= \begin{bmatrix} \frac{dL_R}{d\theta} \frac{dF_{y,x}}{d\theta} L_{P_{k+1|k}} + F_{y,x} \frac{dL_{P_{k+1|k}}}{d\theta} \\ 0 & \frac{dL_{P_{k+1|k}}}{d\theta} \end{bmatrix}, \\ \mathcal{D}_{\text{upd}} &= \begin{bmatrix} D_R & 0 \\ 0 & D_{P_{k+1|k}} \end{bmatrix}, \quad \frac{d\mathcal{D}_{\text{upd}}}{d\theta} = \begin{bmatrix} \frac{dD_R}{d\theta} & 0 \\ 0 & \frac{dD_{P_{k+1|k}}}{d\theta} \end{bmatrix}. \end{aligned} \quad (51)$$

Compute the LD factors and their derivatives:

$$\begin{aligned} &\left[\mathcal{L}, \mathcal{D}, \frac{d\mathcal{L}}{d\theta}, \frac{d\mathcal{D}}{d\theta} \right] \\ &= \text{MWGS_DIFF_LD}\left(\mathcal{A}_{\text{upd}}, \frac{d\mathcal{A}_{\text{upd}}}{d\theta}, \mathcal{D}_{\text{upd}}, \frac{d\mathcal{D}_{\text{upd}}}{d\theta}\right). \end{aligned}$$

Extract

$$L_{R_{e,k+1}} = \mathcal{L}(1 : n_y, 1 : n_y), \quad (52)$$

$$D_{R_{e,k+1}} = \mathcal{D}(1 : n_y, 1 : n_y), \quad (53)$$

$$\frac{dL_{R_{e,k+1}}}{d\theta} = \frac{d\mathcal{L}}{d\theta}(1 : n_y, 1 : n_y), \quad (54)$$

$$\frac{dD_{R_{e,k+1}}}{d\theta} = \frac{d\mathcal{D}}{d\theta}(1 : n_y, 1 : n_y), \quad (55)$$

$$L_{P_{k+1|k+1}} = \mathcal{L}(n_y + 1 : \text{end}, n_y + 1 : \text{end}), \quad (56)$$

$$D_{P_{k+1|k+1}} = \mathcal{D}(n_y + 1 : \text{end}, n_y + 1 : \text{end}), \quad (57)$$

$$\frac{dL_{P_{k+1|k+1}}}{d\theta} = \frac{d\mathcal{L}}{d\theta}(n_y + 1 : \text{end}, n_y + 1 : \text{end}), \quad (58)$$

$$\frac{dD_{P_{k+1|k+1}}}{d\theta} = \frac{d\mathcal{D}}{d\theta}(n_y + 1 : \text{end}, n_y + 1 : \text{end}). \quad (59)$$

Compute the normalized innovation and its derivative:

$$\bar{e}_{k+1} = L_{R_{e,k+1}}^{-1} e_{k+1}, \quad (60)$$

$$\frac{d\bar{e}_{k+1}}{d\theta} = L_{R_{e,k+1}}^{-1} \left(\frac{de_{k+1}}{d\theta} - \frac{dL_{R_{e,k+1}}}{d\theta} \bar{e}_{k+1} \right). \quad (61)$$

Compute the updated state estimate and its derivative

$$(K_{k+1}L_{R_{e,k+1}}) = \mathcal{L}(n_y + 1 : \text{end}, 1 : n_y), \quad (62)$$

$$\frac{d(K_{k+1}L_{R_{e,k+1}})}{d\theta} = \frac{d\mathcal{L}}{d\theta}(n_y + 1 : \text{end}, 1 : n_y), \quad (63)$$

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + (K_{k+1}L_{R_{e,k+1}})\bar{e}_{k+1}, \quad (64)$$

$$\begin{aligned} \frac{d\hat{x}_{k+1|k+1}}{d\theta} &= \frac{d\hat{x}_{k+1|k}}{d\theta} + \frac{d(K_{k+1}L_{R_{e,k+1}})}{d\theta} \bar{e}_{k+1} \\ &\quad + (K_{k+1}L_{R_{e,k+1}}) \frac{d\bar{e}_{k+1}}{d\theta}. \end{aligned} \quad (65)$$

2.3. Contribution to J and $\frac{dJ}{d\theta}$

$$\Delta J = \sum_{i=1}^{n_y} \left(\ln(D_{R_{e,k+1}})_{ii} + \frac{(\bar{e}_{k+1})_i^2}{(D_{R_{e,k+1}})_{ii}} \right), \quad (66)$$

$$\begin{aligned} \frac{d\Delta J}{d\theta} &= \text{tr} \left(D_{R_{e,k+1}}^{-1} \frac{dD_{R_{e,k+1}}}{d\theta} \right) \\ &+ 2\bar{e}_{k+1}^T D_{R_{e,k+1}}^{-1} \frac{d\bar{e}_{k+1}}{d\theta} - \bar{e}_{k+1}^T D_{R_{e,k+1}}^{-2} \frac{dD_{R_{e,k+1}}}{d\theta} \bar{e}_{k+1}, \end{aligned} \quad (67)$$

$$J = J + \Delta J, \quad \frac{dJ}{d\theta} = \frac{dJ}{d\theta} + \frac{d\Delta J}{d\theta}. \quad (68)$$

3. End

$$J_{\text{LD-PKF}}(\theta, Y_0^N) = \frac{Nn_y}{2} \ln(2\pi) + J, \quad (69)$$

$$\frac{dJ_{\text{LD-PKF}}(\theta, Y_0^N)}{d\theta} = \frac{dJ}{d\theta}. \quad (70)$$

Thus, in a single pass through the data Y_0^N for a given parameter value θ , we obtain both the value of the identification criterion $J_{\text{LD-PKF}}(\theta, Y_0^N)$ and its gradient $\nabla_{\theta} J_{\text{LD-PKF}}(\theta, Y_0^N)$. These quantities are then used in a gradient-based parameter identification method (e.g., steepest descent or Newton's method) to find the optimal parameter θ that minimizes $J_{\text{LD-PKF}}(\theta, Y_0^N)$.

4 Numerical experiments

In this section, we demonstrate the performance of the proposed LD-based gradient-type identification algorithm on two problems. The first test problem is a canonical pairwise Markov model with a scalar parameter θ , which allows direct comparison of the conventional and LD implementations. The second problem is the identification of the wave speed c in the one-dimensional wave equation, which serves as a representative example of a distributed parameter system. All numerical experiments were carried out in MATLAB R2017a on a computer with the following hardware and software platform: operating system — Windows 11, processor — 11th Gen Intel Core i3-1115G4 @ 3.00 GHz, 8 GB of RAM. The random number generator was seeded with a fixed value to ensure reproducibility of the results.

For the experiments, we implemented the following methods:

1. `j_lr` (M1) — conventional implementation of $J(\theta)$ using the conventional PKF;
2. `j_lr_grad` (M2) — conventional implementation with analytical gradient $\frac{dJ}{d\theta}$;
3. `j_lr_ld` (M3) — numerically stable LD implementation of $J(\theta)$;
4. `j_lr_ld_grad` (M4) — numerically stable LD implementation with analytical gradient $\frac{dJ}{d\theta}$.

4.1 Test problem

We consider the canonical pairwise Markov model with dimensions $n_x = 2, n_y = 2$:

$$\mathcal{F} = \begin{bmatrix} \theta & 0.10 & 0.11 & \theta \\ 0.11 & 0.10 & \theta & 0.10 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathcal{Q} = \begin{bmatrix} 0.18 & 0.15 & 0.16 & 0.14 \\ 0.15 & 0.18 & 0.14 & 0.16 \\ 0.16 & 0.14 & 0.18 & 0.14 \\ 0.14 & 0.16 & 0.14 & 0.18 \end{bmatrix}.$$

Initial conditions: $\bar{x}_0 = [0.5, 0.5]^T$, $P_0 = 2.5I_2$. The true parameter value is $\theta_{\text{true}} = 0.12$. The measurement sequence length is $N = 200$ (samples $y_0, \dots, y_N$).

The experiment consisted of two stages:

1. **Averaged plots of $J(\theta)$ and $\frac{dJ}{d\theta}$.** Using the Monte Carlo method (50 runs) on a uniform grid $\theta \in [0.05; 0.20]$ of 200 points, we computed the average value of the identification criterion (using M1 and M3) and the average gradient (using M2 and M4).
2. **Parameter identification.** For the four methods (M1–M4) we performed 50 optimization runs using the `fminunc` solver (an unconstrained optimization routine from the MATLAB Optimization Toolbox). The search was conducted without explicit constraints; the initial points were chosen uniformly at random in the interval $[0.05; 0.20]$, and the optimizer consistently converged to solutions within this range. The estimate $\hat{\theta}$ and the computation time were recorded.

Figure 1 shows the averaged value $\overline{J(\theta)}$ computed by the conventional method M1 and the stable LD implementation M3. The two curves coincide, confirming the correctness of the LD version. Figure 2 displays the averaged gradients $\overline{dJ/d\theta}$ obtained by M2 and M4. At the true point $\theta_{\text{true}} = 0.12$, the average gradient equals 1.852388 for both gradient-based methods (M2 and M4).

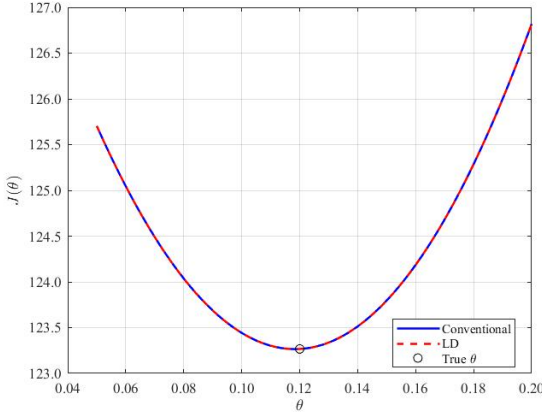


Figure 1. Averaged values $\overline{J(\theta)}$ of the identification criterion (50 runs, $N = 200$).

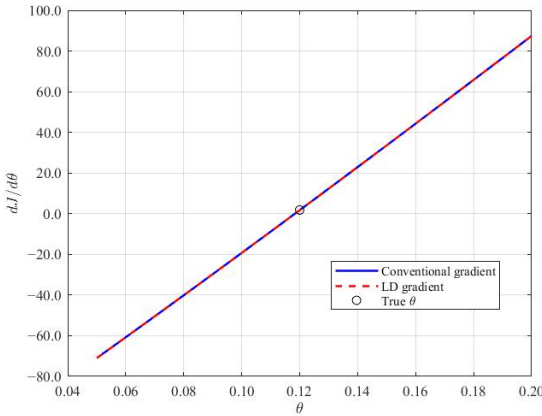


Figure 2. Averaged values $\overline{dJ/d\theta}$ of the gradient (50 runs, $N = 200$).

The identification results for the four methods are summarized in Table 1. All methods (M1–M4) yield identical estimates with mean $\hat{\theta} = 0.117085$, standard deviation $\sigma_{\hat{\theta}} = 0.027897$, MAE = 0.021508 and RMSE = 0.027770. This indicates a small bias relative to the true value 0.12, which is due to the finite sample size ($N = 200$) and the stochastic noise. Using the analytical gradient does not improve the accuracy, but increases the computation time. The slowest method is

M4 (0.3400 s), due to the additional cost of both factorization and derivative computations. In contrast, M3 is faster (0.2717 s) while still providing high numerical stability.

Table 1. Statistics of parameter estimates θ ($N = 200$, 50 runs)

Metric	Method			
	M1	M2	M3	M4
Mean $\hat{\theta}$	0.117085	0.117085	0.117085	0.117085
Std $\sigma_{\hat{\theta}}$	0.027897	0.027897	0.027897	0.027897
MAE	0.021508	0.021508	0.021508	0.021508
RMSE	0.027770	0.027770	0.027770	0.027770
Time (s)	0.1618	0.1633	0.2717	0.3400

The performed experiments confirm the reliability of the proposed algorithms for computing $J(\theta)$ and $\frac{dJ}{d\theta}$. The averaged plots demonstrate that the LD implementation yields the same values as the conventional one while being more stable. Parameter identification shows that using the analytical gradient does not improve the accuracy for this particular problem, but it also does not introduce any additional error. The obtained parameter estimates ($\hat{\theta} \approx 0.1171$) are biased with respect to the true value 0.12, which is due to the limited sample size ($N = 200$) and the stochastic noise. Improving the identification accuracy may require longer observation sequences or the use of regularization techniques. Overall, the developed Algorithm 4 provides a correct basis for gradient-based parameter identification within the class of pairwise Markov models.

4.2 Identification of the wave speed in the one-dimensional wave equation

To test the proposed approach on a distributed-parameter system, we consider the one-dimensional wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \quad x \in (0, 1), \quad t > 0,$$

with homogeneous Dirichlet boundary conditions $u(0, t) = u(1, t) = 0$ and initial conditions $u(x, 0) = \sin(\pi x)$, $\frac{\partial u}{\partial t}(x, 0) = 0$. The unknown parameter $c > 0$ is the wave speed.

Using the explicit “leap-frog” finite-difference scheme [Samarskii, 1977] on a uniform grid with steps Δx and Δt , the equation is approximated to second order:

$$u_i^{j+1} = ru_{i-1}^j + 2(1-r)u_i^j + ru_{i+1}^j - u_i^{j-1}, \quad r = \left(\frac{c\Delta t}{\Delta x}\right)^2.$$

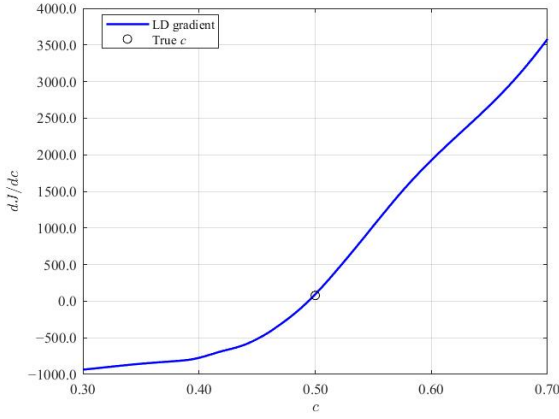


Figure 4. Averaged gradient $\overline{dJ/dc}$ for the wave equation (50 runs).

Introducing the state vector $u^j = (u_1^j, \dots, u_M^j)^T$ (M is the number of interior nodes), we obtain the non-Markovian dynamics $u^{j+1} = Au^j - u^{j-1}$, where A is a tridiagonal matrix. To convert it into a pairwise Markov model we use the augmented state and assume that all interior nodes are measured ($H = I_M$ which means that measurements are taken at M internal points). The model then takes the form (1), (2) with matrices

$$F_{x,x} = A, F_{x,y} = -I_M, F_{y,x} = I_M, F_{y,y} = 0, \\ Q = \begin{pmatrix} Q_1 + R & 0 \\ 0 & R \end{pmatrix},$$

where Q_1 and R are the process noise and measurement noise covariance matrices.

The numerical experiment uses the following parameters: $M = 5$ ($\Delta x = 1/6$), $\Delta t = 0.3$ (Courant number $r = 0.81 < 1$), $K = 100$ time steps. The true value is $c_{\text{true}} = 0.5$. Process noise $Q_1 = 10^{-4}I_5$ and measurement noise $R = 0.01$. Figure 3 shows the averaged values $J(c)$ over 50 Monte Carlo runs, computed with the stable LD implementation M3. A clear minimum near $c = 0.5$ is observed. The averaged gradient $\overline{dJ/dc}$, computed by M4, is shown in Figure 4.

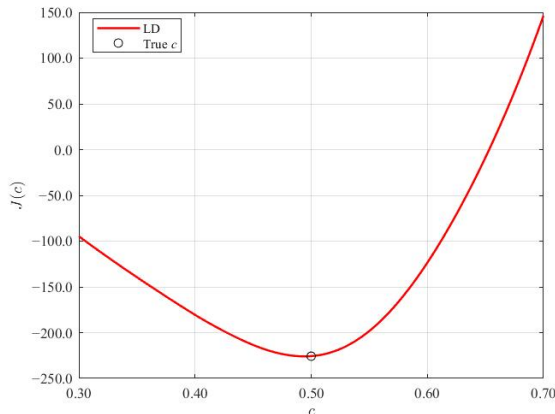


Figure 3. Averaged criterion $\overline{J(c)}$ for the wave equation (50 runs).

Parameter identification for c was then performed by Monte Carlo (50 runs) using the `fmincon` solver (a constrained optimization routine from the MATLAB Optimization Toolbox) with criteria M3 (without gradient) and M4 (with gradient). Since standard methods often diverge for this problem, we used only M3 and M4 here. The search interval for c was set to $[0.3; 0.7]$, and the initial points were chosen uniformly at random within this interval. The resulting statistics are presented in Table 2. The mean estimate $\bar{c} = 0.493727$ is close to the true value 0.5, and the small MAE and RMSE (0.006478 and 0.007906, respectively) confirm high identification accuracy. The standard deviation 0.004862 indicates the robustness of the procedure. Using the gradient (M4) increases the computation time by about a factor of 1.6 compared to the gradient-free method (M3).

Table 2. Identification results for the wave speed c (50 runs)

Metric	Method	
	M3	M4
Mean $\bar{c}$	0.493727	0.493727
Std σ_c	0.004862	0.004862
MAE	0.006478	0.006478
RMSE	0.007906	0.007906
Time (s)	0.2494	0.4101

Thus, the proposed method (in particular, the LD implementation of $J(\theta)$ and its gradient) successfully solves the problem of identifying unknown parameter of a distributed system described by the wave equation. This opens prospects for its application to more complex problems, including parameter identification in two-dimensional and nonlinear equations.

5 Conclusion

In this paper, we propose a new approach to constructing gradient-type methods for parameter identification of pairwise Markov models. The approach is based on the modified weighted Gram–Schmidt orthogonalization and the matrix LD decomposition. It is intended for solving parameter identification problems in the class of pairwise Markov models. Within this framework, the following theoretical results have been obtained:

- 1) a new criterion for identifying the parameters of a pairwise Markov model is formulated in terms of the pairwise LD filter;

- 2) a method for computing the derivatives with respect to the model parameter θ in the pairwise LD filter is proposed, based on original techniques for differentiating parameterized matrices;
- 3) a new algorithm for computing the gradient of the identification criterion using the pairwise LD filter is developed.

The advantage of the proposed approach lies in the use of the numerically stable modified weighted Gram–Schmidt orthogonalization procedure, which underlies all developed methods and algorithms. The pairwise LD filter maintains the required symmetry and positive definiteness of the error covariance matrices. The computation of derivatives in Algorithm 4 is performed algorithmically and does not require numerical differentiation, thus avoiding all the drawbacks inherent in that class of methods. The developed algorithms have a block-matrix structure convenient for software implementation. The practical applicability of the approach is demonstrated on the example of identifying the parameter of a wave equation model from noisy measurements.

Thus, the proposed algorithms can be used in constructing numerical methods for solving parameter identification problems of pairwise Markov models. The prospects of applying the results obtained lie in the development of new decentralized parameter identification methods for the pairwise Markov models [Tsyganov and Tsyganova, 2025a].

Acknowledgements

This work was carried out under Additional Agreement No. 073-03-2026-035/1 dated February 21, 2026, to the Agreement on the provision of a subsidy to a federal budgetary or autonomous institution for the financial support of the fulfillment of a state assignment for the provision of public services (performance of work) No. 073-03-2026-035 dated January 23, 2026, concluded between Ulyanovsk State Pedagogical University and the Ministry of Education of the Russian Federation.

References

- Bierman, G. J. (1977). *Factorization Methods for Discrete Sequential Estimation*. Academic Press, New York.
- Derrode, S. and Pieczynski, W. (2004). Signal and image segmentation using pairwise Markov chains. *IEEE Transactions on Signal Processing*, **52** (9), pp. 2477–2489.
- Gibbs, B. P. (2011). *Advanced Kalman filtering, least-squares and modeling : a practical handbook*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Golub, G. H. and Van Loan, C. F. (2013). *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 4th edition.
- Gorynin, I., Monfrini, E., and Pieczynski, W. (2017). Pairwise Markov models for stock index forecasting. In *Proceedings of the 25th European Signal Processing Conference (EUSIPCO 2017)*, Kos Island, Greece, pp. 2095–2099.
- Grewal, M. S. and Andrews, A. P. (2015). *Kalman Filtering: Theory and Practice with MATLAB*. John Wiley & Sons, New York, 4th edition.
- Ivanov, D., Zhdanov, A., and Sandler, I. (2024). Identification of dynamical systems with time-varying disturbances and fractional errors-in-variables. application example to a drone model. *Cybernetics and Physics*, **13** (1), pp. 49–56.
- Kailath, T., Sayed, A. H., and Hassibi, B. (2000). *Linear estimation*. Prentice Hall, New Jersey.
- Kulikova, M. V. (2017). Gradient-based parameter estimation in pairwise linear gaussian system. *IEEE Transactions on Automatic Control*, **62** (3), pp. 1511–1517.
- Kulikova, M. V., Tsyganova, J. V., and Kulikov, G. Y. (2020). UD-based pairwise and MIMO Kalman-like filtering for estimation of econometric model structures. *IEEE Transactions on Automatic Control*, **65** (10), pp. 4472–4479.
- Kulikova, M. V. and Tsyganova, Y. V. (2017). Numerically stable Kalman filter implementations for estimating linear pairwise Markov models in the presence of Gaussian noise. *Computational technologies*, **22** (3), pp. 45–60. (In Russian).
- Ljung, L. (1978). Convergence analysis of parametric identification methods. *IEEE Transactions on Automatic Control*, **23** (5), pp. 770–783.
- Ljung, L. (1999). *System Identification: Theory for the User*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 2th edition.
- Nemesin, V. and Derrode, S. (2012). Robust blind pairwise Kalman algorithms using QR decompositions. *IEEE Transactions on Signal Processing*, **61** (1), pp. 5–9.
- Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization. Springer Series in Operations Research and Financial Engineering*. Springer Nature, Berlin/Heidelberg, Germany, 2nd edition.
- Pham, C. T., Tran, T. T. T., Huynh, D. A. B., Nguyen, Q. C., and Nguyen, T. H. (2024). Computed tomography image reconstruction: Integrating iterative methods with maximum likelihood expectation maximization and deep learning models. *Cybernetics and Physics*, **13** (2), pp. 130–141.
- Pieczynski, W. (2003). Pairwise Markov chains. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25** (5), pp. 634–639.
- Pieczynski, W. (2007). Multisensor triplet Markov chains and theory of evidence. *International Journal of Approximate Reasoning*, **45**, pp. 1–16.
- Pieczynski, W. and Desbouvries, F. (2003). Kalman filtering using pairwise Gaussian models. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'03)*, vol. 6, Hong Kong, China, pp. 57–60.

- Rafi, S., Castella, M., and Pieczynski, W. (2011). Pairwise Markov model applied to unsupervised image separation. In *Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA 2011)*, Innsbruck, Austria.
- Samarskii, A. A. (1977). *Theory of Difference Schemes*. Nauka, Moscow. (In Russian).
- Thornton, C. L. (1976). *Triangular Covariance Factorizations for Kalman Filtering*. PhD thesis, School of Engineering, University of California at Los Angeles.
- Tsyganov, A. and Tsyganova, J. (2025a). Parameter identification of discrete-time linear stochastic systems based on decentralized square-root information filtering. *Cybernetics and Physics*, **14**(2), pp. 191–199.
- Tsyganov, A. V. (2026). Numerically stable implementation of the pairwise Kalman filter based on LD decomposition of covariance matrices. In *Proceedings of the XXIX International Conference on Soft Computing and Measurements (SCM'2026)*, Saint Petersburg, Russia, Saint Petersburg Electrotechnical University “LETI”, May, pp. 36–38. (In Russian, Preprint).
- Tsyganov, A. V. and Tsyganova, J. V. (2025b). A general approach to constructing gradient methods for parameter identification based on modified weighted Gram–Schmidt orthogonalization and information-type discrete filtering algorithms. *Computer Research and Modeling*, **17**(5), pp. 761–782. (In Russian).
- Tsyganov, A. V., Tsyganova, Y. V., Kuvshinova, A. N., and Golubkov, A. V. (2023). Discrete-time filtering algorithms based on the modified weighted Gram – Schmidt orthogonalization for discrete-time stochastic systems with multiplicative and additive noises. *Computational Technologies*, **28**(5), pp. 70–86. (In Russian).
- Tsyganova, J. V., Kulikova, M. V., and Tsyganov, A. V. (2020). A general approach for designing the MWGS-based information-form Kalman filtering methods. *European Journal of Control*, **56**, pp. 86–97.
- Tsyganova, J. V. and Tsyganov, A. V. (2018). On the computation of derivatives within LD factorization of parametrized matrices. *Bulletin of Irkutsk State University. Series Mathematics*, **23**, pp. 64–79. (In Russian).