

# Robust blind pairwise Kalman algorithms using QR decompositions

Valérien Nêmesin, and Stéphane Derrode

**Abstract**—The Pairwise Kalman Filter (PKF) [1] is an extension of the classical Kalman filter that keeps propagation equations explicit, *i.e.* it does not require time consuming simulations. The contribution of this note is twofold. First, new robust equations for filtering, smoothing and unsupervised off-lined parameters estimation based on QR decompositions are presented. Second, since the model is over-parametrized, we give a simple condition to uniquely characterize a filter of interest when the dimension of observations is equal to the dimension of states. Unsupervised experiments based on simulated data confirm the nice behavior of the robust PKF, even for a limited number of observations.

**Index Terms**—Kalman filter, Pairwise Kalman filter, Estimation-Maximization, QR decomposition.

## I. INTRODUCTION

LET us consider the following constant parameters linear Gaussian dynamical stochastic system

$$\underbrace{\begin{pmatrix} \mathbf{x}_{n+1} \\ \mathbf{y}_n \end{pmatrix}}_{\mathbf{t}_{n+1}} = \underbrace{\begin{pmatrix} \mathbf{F}^{x,x} & \mathbf{F}^{x,y} \\ \mathbf{F}^{y,x} & \mathbf{F}^{y,y} \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} \mathbf{x}_n \\ \mathbf{y}_{n-1} \end{pmatrix}}_{\mathbf{t}_n} + \underbrace{\begin{pmatrix} \boldsymbol{\omega}_{n+1}^x \\ \boldsymbol{\omega}_{n+1}^y \end{pmatrix}}_{\boldsymbol{\omega}_{n+1}} \quad (1)$$

in which  $\mathbf{x}_n \in \mathbb{R}^{n_x}$  are the states and  $\mathbf{y}_n \in \mathbb{R}^{n_y}$  are the observations. Let  $\mathbf{x}$  and  $\mathbf{y}$  denote the state and observation processes. The Gaussian process  $\boldsymbol{\omega} = \{\boldsymbol{\omega}_n\}_{n \in \mathbb{N}}$ , where  $\boldsymbol{\omega}_n \in \mathbb{R}^{n_t}$  are mutually independent and independent of  $\mathbf{t}_0 \sim \mathcal{N}(\hat{\mathbf{t}}_0, \mathbf{Q}_0)$  and

$$\boldsymbol{\omega}_n \sim \mathcal{N}\left(\mathbf{0}_{n_t}, \underbrace{\begin{pmatrix} \mathbf{Q}^{x,x} & \mathbf{Q}^{x,y} \\ [\mathbf{Q}^{x,y}]^T & \mathbf{Q}^{y,y} \end{pmatrix}}_{\mathbf{Q}}\right).$$

Let  $\Theta = \{\hat{\mathbf{t}}_0, \mathbf{Q}_0, \mathbf{F}, \mathbf{Q}\}$  denotes the set of parameters (independent of  $n$ ) that describes the model.

From (1), the distribution  $p(\mathbf{t}_{n+1} | \mathbf{t}_{0:n}) = p(\mathbf{t}_{n+1} | \mathbf{t}_n)$  is that of a Markov chain, so that the process  $\mathbf{t} = (\mathbf{x}, \mathbf{y})$  is called a pairwise Markov process [1], by analogy with the discrete-state pairwise Markov chain [2]. The Gaussian and linear system (1), called “Pairwise Kalman Model” (denoted by PKF), is strictly more general than the standard linear Gaussian system (denoted by KF), since the latter can be recovered from the former by setting  $\mathbf{F}^{x,y}$ ,  $\mathbf{F}^{y,x}$  and  $\mathbf{Q}^{x,y}$  to  $\mathbf{0}$ . Especially,  $\mathbf{x}$  is no more necessarily Markovian, but it is still possible to derive Kalman-like algorithms for filtering and smoothing [3, proposition 3], *i.e.* without having recourse to simulations used, for example, in the Unscented Kalman Filters (UKF) [4].

However, the PKF involves more parameters and its higher complexity can make it difficult to tune manually for a given application. To deal with this problem, an Expectation-Maximization (EM) algorithm has been designed for unsupervised off-lined parameters estimation in [5], by iterating and alternating filtering and smoothing. However, as for the KF, direct equations obtained from calculations lead to severe numerical instabilities, *i.e.* not positive semi-definite covariance matrices and so, to the divergence of the likelihood.

V. Nêmesin and S. Derrode are with the GSM team, Institut Fresnel (CNRS UMR 7249), Aix-Marseille University, Domaine universitaire de Saint Jérôme, avenue Escadrille Normandie-Niémen, 13397 Marseille Cedex 20, France. e-mail: {*firstname.lastname*}@fresnel.fr.

Therefore the first contribution of this note is to adapt the robust numerical methods inherited from the KF (known as factorization methods, see [6]–[9]) to the PKF, for (i) prediction and filtering, (ii) smoothing, and (iii) EM equations. The main idea is to replace the direct computation of covariance matrices by triangular square roots using QR decompositions, guaranteeing symmetry and positivity, and preventing error propagation.

Also, given a set of parameters  $\Theta$ , it is easy to find transformation matrices  $\mathbf{M}$  that define  $\mathbf{M}$ -equivalent parameter sets  $\Theta_M$ , with which neither observations nor likelihood are altered. As a consequence, EM algorithm can converge to one of the equivalent filters, so that restored states can be very different from one running of EM to another (and especially from one initialization of EM to another), and difficult to interpret. Hence, the second contribution is to force a simple constraint on the system (1) to get a unique solution for  $\Theta$  when  $n_x = n_y$ .

The remaining of the paper is organized as follows. Original PKF equations for filtering, smoothing and parameters estimation are briefly recalled in Section II. Then Section III presents robust equivalent equations based on QR decompositions, whereas Section IV deals with equivalent parameters solutions. Illustrative and comparative experiments on simulated data are reported in Section V. Finally, Section VI draws conclusions and presents further possible improvements.

## II. PAIRWISE KALMAN FILTER: ORIGINAL EQUATIONS

Here is a brief recall of work by B. Ait-el-Fquih *et al* on PKF [5] for filtering, smoothing and unsupervised parameters estimation. All this material will serve as a basis for discussion on numerical robustness in next Sections.

### A. Prediction, filtering and smoothing

Let us first define a few constants for latter use:

$$\begin{aligned} \mathbf{F}_2^{x,x} &= \mathbf{F}^{x,x} - \mathbf{Q}^{x,y} [\mathbf{Q}^{y,y}]^{-1} \mathbf{F}^{y,x}, \\ \mathbf{F}_2^{x,y} &= \mathbf{F}^{x,y} - \mathbf{Q}^{x,y} [\mathbf{Q}^{y,y}]^{-1} \mathbf{F}^{y,y}, \end{aligned} \quad (2)$$

$$\begin{aligned} \mathbf{F}^{y,t} &= (\mathbf{F}^{y,x}, \mathbf{F}^{y,y}), \quad \mathbf{F}_2^{t,t} = (\mathbf{F}_2^{x,x}, \mathbf{F}_2^{x,y}) \\ \mathbf{Q}_2^{x,x} &= \mathbf{Q}^{x,x} - \mathbf{Q}^{x,y} [\mathbf{Q}^{y,y}]^{-1} \mathbf{Q}^{y,x} \end{aligned} \quad (3)$$

Predicting state expectation  $\hat{\mathbf{x}}_{n+1|n}$  and covariance matrix  $\mathbf{P}_{n+1|n}$  are given by

$$\begin{aligned} \hat{\mathbf{x}}_{n+1|n} &= \mathbf{F}_2^{x,x} \hat{\mathbf{x}}_{n|n} + \mathbf{Q}^{x,y} [\mathbf{Q}^{y,y}]^{-1} \mathbf{y}_n + \mathbf{F}_2^{x,y} \mathbf{y}_{n-1} \\ \mathbf{P}_{n+1|n} &= \mathbf{Q}_2^{x,x} + \mathbf{F}_2^{x,x} \mathbf{P}_{n|n} [\mathbf{F}_2^{x,x}]^T \end{aligned} \quad (4)$$

and filtering state moments  $\hat{\mathbf{x}}_{n+1|n+1}$  and  $\mathbf{P}_{n+1|n+1}$  by

$$\begin{aligned} \tilde{\mathbf{y}}_{n+1} &= \mathbf{y}_{n+1} - \mathbf{F}^{y,x} \hat{\mathbf{x}}_{n+1|n} - \mathbf{F}^{y,y} \mathbf{y}_n \\ \mathbf{S}_{n+1} &= \mathbf{Q}^{y,y} + \mathbf{F}^{y,x} \mathbf{P}_{n+1|n} [\mathbf{F}^{y,x}]^T \end{aligned} \quad (5)$$

$$\mathbf{K}_{n+1|n+1} = \mathbf{P}_{n+1|n} [\mathbf{F}^{y,x}]^T [\mathbf{S}_{n+1}]^{-1} \quad (6)$$

$$\begin{aligned} \hat{\mathbf{x}}_{n+1|n+1} &= \hat{\mathbf{x}}_{n+1|n} + \mathbf{K}_{n+1|n+1} \tilde{\mathbf{y}}_{n+1} \\ \mathbf{P}_{n+1|n+1} &= \mathbf{P}_{n+1|n} - \mathbf{K}_{n+1|n+1} \mathbf{S}_{n+1} [\mathbf{K}_{n+1|n+1}]^T \end{aligned} \quad (7)$$

with  $\tilde{\mathbf{y}}_{n+1}$  the innovation,  $\mathbf{S}_{n+1}$  the innovative state covariance matrix, and  $\mathbf{K}_{n+1|n+1}$  the pairwise Kalman gain.

Using a backward pass, it is also possible to compute smoothing state moments according to

$$\begin{aligned} \mathbf{K}_{n|N} &= \mathbf{P}_{n|n} [\mathbf{F}_2^{x,x}]^T [\mathbf{P}_{n+1|n}]^{-1} \\ \hat{\mathbf{x}}_{n|N} &= \hat{\mathbf{x}}_{n|n} + \mathbf{K}_{n|N} (\hat{\mathbf{x}}_{n+1|N} - \hat{\mathbf{x}}_{n+1|n}) \\ \mathbf{P}_{n|N} &= \mathbf{P}_{n|n} + \mathbf{K}_{n|N} (\mathbf{P}_{n+1|N} - \mathbf{P}_{n+1|n}) [\mathbf{K}_{n|N}]^T \end{aligned} \quad (8)$$

### B. EM-based parameters estimation

Automatic off-line parameters estimation can be performed using EM algorithm [10], [11], which is able to find a local maximum of the likelihood  $\mathcal{L}(\mathbf{y}; \Theta)$ . At EM step  $i$ , parameters  $\Theta^{(i)}$  are obtained by maximizing an auxiliary function  $q(\Theta^{(i-1)}, \Theta)$ , which finally gives

$$\begin{aligned} \hat{\mathbf{t}}_0^{(i)} &= \hat{\mathbf{t}}_{0|N}, \mathbf{Q}_0^{(i)} = \mathbf{Q}_{0|N}, \mathbf{F}^{(i)} = \tilde{\mathbf{C}}_{1,0} [\tilde{\mathbf{C}}_{0,0}]^{-1} \\ \mathbf{Q}^{(i)} &= \frac{1}{N+1} \left( \tilde{\mathbf{C}}_{1,1} - \tilde{\mathbf{C}}_{1,0} [\tilde{\mathbf{C}}_{0,0}]^{-1} [\tilde{\mathbf{C}}_{1,0}]^T \right) \end{aligned}$$

where  $\hat{\mathbf{t}}_{n|N} = (\hat{\mathbf{x}}_{n|N}^T, \mathbf{y}_n^T)^T$  (see [5] for details) and

$$\begin{aligned} \tilde{\mathbf{C}}_{0,0} &= \sum_{n=0}^N \begin{pmatrix} \mathbf{P}_{n|N} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \hat{\mathbf{t}}_{n|N} \hat{\mathbf{t}}_{n|N}^T \\ \tilde{\mathbf{C}}_{1,0} &= \sum_{n=0}^N \begin{pmatrix} \mathbf{P}_{n+1|N} [\mathbf{K}_{n|N}]^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \hat{\mathbf{t}}_{n+1|N} \hat{\mathbf{t}}_{n|N}^T \\ \tilde{\mathbf{C}}_{1,1} &= \sum_{n=0}^N \begin{pmatrix} \mathbf{P}_{n+1|N} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \hat{\mathbf{t}}_{n+1|N} \hat{\mathbf{t}}_{n+1|N}^T \end{aligned}$$

Each iteration requires to filter and smooth data according to parameters from previous iteration.

## III. ROBUST PAIRWISE KALMAN FILTER

The direct equations defined in previous Section are subject to strong numerical instabilities, as for the classical KF. Indeed, when the process is well known, *i.e.* the process noise covariance is small, it is easy for rounding error to make the state covariance matrix  $\mathbf{P}$  invalid: negative diagonal entries or otherwise not positive semi-definite. Positive definite matrices have the property that they have upper triangular matrix square roots  $\mathbf{S}$  such that  $\mathbf{P} = \mathbf{S}^T \mathbf{S}$ , which can be computed efficiently using the QR decomposition algorithm. More importantly, if the covariance is kept in this form, it can never have a negative diagonal or become asymmetric. We now propose to adapt QR factorizations developed for KF [7] to the PKF case.

*Remark:* Notation  $\mathbf{A}^{\frac{1}{2}}$  represents an upper-triangular square root of  $\mathbf{A}$ .

### A. Covariance matrices

#### 1) Forward QR equations:

*Filtering:* According to Kaminski *et al* approach [7], filtering equations from (5) to (7) can be written

$$\begin{pmatrix} \mathbf{S}_{n+1} & \mathbf{S}_{n+1} [\mathbf{K}_{n+1|n+1}]^T \\ \mathbf{K}_{n+1|n+1} \mathbf{S}_{n+1} & \mathbf{P}_{n+1|n+1} + \mathbf{K}_{n+1|n+1} \mathbf{S}_{n+1} [\mathbf{K}_{n+1|n+1}]^T \\ \mathbf{Q}^{y,y} + \mathbf{F}^{y,x} \mathbf{P}_{n+1|n} [\mathbf{F}^{y,x}]^T & \mathbf{F}^{y,x} \mathbf{P}_{n+1|n} \\ \mathbf{P}_{n+1|n} [\mathbf{F}^{y,x}]^T & \mathbf{P}_{n+1|n} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{C}}_{0,0} & \tilde{\mathbf{C}}_{0,1} \\ [\tilde{\mathbf{C}}_{0,1}]^T & \tilde{\mathbf{C}}_{1,1} \end{pmatrix}, \quad (10)$$

which writes

in which the left side is concerned with filtering (output), whereas the right side is concerned with prediction (input). Matrices  $\mathbf{Q}^{y,y}$ ,  $\mathbf{S}_{n+1}$ ,  $\mathbf{P}_{n+1|n}$  and  $\mathbf{P}_{n+1|n+1}$  are symmetrical and positive semidefinite so

they admit square roots. Trivial, but not necessarily triangular, square roots of both sides of eq. (10) are respectively given by

$$\underbrace{\begin{pmatrix} [\mathbf{S}_{n+1}]^{\frac{1}{2}} & [\mathbf{S}_{n+1}]^{\frac{1}{2}} [\mathbf{K}_{n+1|n+1}]^T \\ \mathbf{0}_{n_x, n_y} & [\mathbf{P}_{n+1|n+1}]^{\frac{1}{2}} \end{pmatrix}}_{\mathbf{N}_{output}^f}, \quad (11)$$

$$\underbrace{\begin{pmatrix} [\mathbf{Q}^{y,y}]^{\frac{1}{2}} & \mathbf{0}_{n_y, n_x} \\ [\mathbf{P}_{n+1|n}]^{\frac{1}{2}} [\mathbf{F}^{y,x}]^T & [\mathbf{P}_{n+1|n}]^{\frac{1}{2}} \end{pmatrix}}_{\mathbf{N}_{input}^f}. \quad (12)$$

As the two triangular matrices are related by an orthogonal matrix  $\mathbf{Q}^*$ , the QR decomposition of  $\mathbf{N}_{input}^f$  gives  $\mathbf{N}_{output}^f$ , so that we get  $[\mathbf{S}_{n+1}]^{\frac{1}{2}}$  and  $[\mathbf{P}_{n+1|n+1}]^{\frac{1}{2}}$ . Computation of square roots of  $\mathbf{Q}^{y,y}$  and  $\mathbf{Q}_2^{x,x}$  are given below.

Using the same principle, we can now express the square roots of other covariance matrices. Please note that  $\mathbf{Q}^*$  is a common notation for different matrices.

*Prediction:*  $\mathbf{P}_{n+1|n}$ , see eq. (4)

$$\begin{pmatrix} [\mathbf{P}_{n+1|n}]^{\frac{1}{2}} \\ \mathbf{0}_{n_x, n_x} \end{pmatrix} = \mathbf{Q}^* \begin{pmatrix} [\mathbf{Q}_2^{x,x}]^{\frac{1}{2}} \\ [\mathbf{P}_{n|n}]^{\frac{1}{2}} [\mathbf{F}_2^{x,x}]^T \end{pmatrix} \quad (13)$$

#### 2) Backward QR equations:

*Smoothing:*  $\mathbf{P}_{n|N}$ , see eq. (8) and (9):

$$\begin{pmatrix} [\mathbf{P}_{n+1|n}]^{\frac{1}{2}} & [\mathbf{P}_{n+1|n}]^{\frac{1}{2}} [\mathbf{K}_{n|N}]^T \\ \mathbf{0}_{n_x, n_x} & [\mathbf{P}_{n|N}]^{\frac{1}{2}} \\ \mathbf{0}_{n_x, n_x} & \mathbf{0}_{n_x, n_x} \end{pmatrix} = \mathbf{Q}^* \begin{pmatrix} [\mathbf{Q}_2^{x,x}]^{\frac{1}{2}} & \mathbf{0}_{n_x, n_x} \\ [\mathbf{P}_{n|n}]^{\frac{1}{2}} [\mathbf{F}_2^{x,x}]^T & [\mathbf{P}_{n|n}]^{\frac{1}{2}} \\ \mathbf{0}_{n_x, n_x} & [\mathbf{P}_{n+1|N}]^{\frac{1}{2}} [\mathbf{K}_{n|N}]^T \end{pmatrix} \quad (14)$$

For latter use, let  $\mathbf{C}'_{n+1, n|N} = [\mathbf{P}_{n+1|N}]^{\frac{1}{2}} [\mathbf{K}_{n|N}]^T$ .

#### 3) Constants $\mathbf{Q}^{y,y}$ and $\mathbf{Q}_2^{x,x}$ :

$$\begin{pmatrix} [\mathbf{Q}^{y,y}]^{\frac{1}{2}} & [\mathbf{Q}^{y,y}]^{-\frac{1}{2}} \mathbf{Q}^{y,x} \\ \mathbf{0}_{n_x, n_y} & [\mathbf{Q}_2^{x,x}]^{\frac{1}{2}} \end{pmatrix} \quad (15)$$

$$= \mathbf{Q}^* \begin{pmatrix} \underbrace{[\mathbf{Q}^{y,y} - \mathbf{Q}^{y,x} [\mathbf{Q}^{x,x}]^{-1} \mathbf{Q}^{x,y}]^{\frac{1}{2}}}_{\mathbf{A}^{y,y}} & \mathbf{0}_{n_y, n_x} \\ \underbrace{[\mathbf{Q}^{x,x}]^{-\frac{1}{2}} \mathbf{Q}^{x,y}}_{\mathbf{A}^{x,y}} & \underbrace{[\mathbf{Q}^{x,x}]^{\frac{1}{2}}}_{\mathbf{A}^{x,x}} \end{pmatrix} \quad (16)$$

where  $\mathbf{A} = \begin{pmatrix} \mathbf{A}^{x,x} & \mathbf{A}^{x,y} \\ \mathbf{0}_{n_y, n_x} & \mathbf{A}^{y,y} \end{pmatrix}$  is a block upper-triangular square root of  $\mathbf{Q}$ .

### B. Robust EM algorithm

Robust estimation for  $\mathbf{Q}^{(i)}$  and  $\mathbf{F}^{(i)}$  matrices at iteration  $i$  can be obtained by computing the square root of

$$\mathbf{M}_{\text{sums}} = \begin{pmatrix} \tilde{\mathbf{C}}_{0,0} & \tilde{\mathbf{C}}_{0,1} \\ [\tilde{\mathbf{C}}_{0,1}]^T & \tilde{\mathbf{C}}_{1,1} \end{pmatrix}$$

$$[\mathbf{M}_{\text{sums}}]^{\frac{1}{2}} = \begin{pmatrix} [\tilde{\mathbf{C}}_{0,0}]^{\frac{1}{2}} & [\tilde{\mathbf{C}}_{0,0}]^{\frac{1}{2}} [\mathbf{F}^{(i)}]^T \\ \mathbf{0}_{n_t, n_t} & \sqrt{N+1} [\mathbf{Q}^{(i)}]^{\frac{1}{2}} \end{pmatrix} \quad (17)$$

To compute a square root of  $\mathbf{M}_{\text{sums}}$  we propose the following iterative algorithm which computes recursively the square root of  $\mathbf{W}_p = \sum_{n=p}^N \mathbf{\Gamma}_{n,n+1|N}^t$ , where  $\mathbf{\Gamma}_{n,n+1|N}^t$  is the auto-correlation of  $([\mathbf{t}_{n|N}]^T, [\mathbf{t}_{n+1|N}]^T)^T$ :

- For  $p = N$ ,  $[\mathbf{W}_N]^{\frac{1}{2}} = [\mathbf{\Gamma}_{N,N+1|N}^t]^{\frac{1}{2}}$
- For  $p = N - 1$  to 0, use the next QR decomposition

$$\begin{pmatrix} [\mathbf{W}_p]^{\frac{1}{2}} \\ \mathbf{0}_{2n_t, 2n_t} \end{pmatrix} = \mathbf{Q}^* \begin{pmatrix} [\mathbf{W}_{p+1}]^{\frac{1}{2}} \\ [\mathbf{\Gamma}_{p,p+1|N}^t]^{\frac{1}{2}} \end{pmatrix} \quad (18)$$

Hence  $[\mathbf{W}_p]^{\frac{1}{2}}$  is obtained from  $[\mathbf{W}_{p+1}]^{\frac{1}{2}}$  and  $[\mathbf{\Gamma}_{p,p+1|N}^t]^{\frac{1}{2}}$ . The latter is obtained using the following three-steps procedure:

- 1) Compute a square root of the covariance matrix  $\mathbf{C}_{n,n+1|N}^x$  of  $([\mathbf{x}_{n|N}]^T, [\mathbf{x}_{n+1|N}]^T)^T$  using the next QR decomposition:

$$\begin{pmatrix} [\mathbf{P}_{n+1|n}]^{\frac{1}{2}} & \begin{bmatrix} [\mathbf{P}_{n+1|n}]^{\frac{1}{2}} [\mathbf{K}_{n|N}]^T & \mathbf{0}_{n_x, n_x} \end{bmatrix} \\ \begin{bmatrix} \mathbf{0}_{n_x, n_x} \\ \mathbf{0}_{n_x, n_x} \end{bmatrix} & \begin{bmatrix} \mathbf{P}_{n|N} & \mathbf{K}_{n|N} \mathbf{P}_{n+1|N} \\ \mathbf{P}_{n+1|N} [\mathbf{K}_{n|N}]^T & \mathbf{P}_{n+1|N} \end{bmatrix}^{\frac{1}{2}} \end{pmatrix} \\ = \mathbf{Q}^* \begin{pmatrix} [\mathbf{Q}_2^{x,x}]^{\frac{1}{2}} & \mathbf{0}_{n_x, n_x} & \mathbf{0}_{n_x, n_x} \\ [\mathbf{P}_{n|n}]^{\frac{1}{2}} [\mathbf{F}_2^{x,x}]^T & [\mathbf{P}_{n|n}]^{\frac{1}{2}} & \mathbf{0}_{n_x, n_x} \\ \mathbf{0}_{n_x, n_x} & \mathbf{C}'_{n+1, n|N} & [\mathbf{P}_{n+1|N}]^{\frac{1}{2}} \end{pmatrix} \quad (19)$$

Let us define the following decomposition for latter use (each  $\mathbf{C}_{i,j}^x$  has dimension  $n_x \times n_x$ )

$$[\mathbf{C}_{n,n+1|N}^x]^{\frac{1}{2}} = \begin{pmatrix} \mathbf{C}_{0,0}^x & \mathbf{C}_{0,1}^x \\ \mathbf{0}_{n_x, n_x} & \mathbf{C}_{1,1}^x \end{pmatrix}$$

- 2) Compute a square root of  $\mathbf{C}_{n,n+1|N}^t$ , the covariance matrix of  $([\mathbf{t}_{n|N}]^T, [\mathbf{t}_{n+1|N}]^T)^T$ :

$$[\mathbf{C}_{n,n+1|N}^t]^{\frac{1}{2}} = \begin{pmatrix} \mathbf{C}_{0,0}^x & \mathbf{0}_{n_x, n_y} & \mathbf{C}_{0,1}^x & \mathbf{0}_{n_x, n_y} \\ \mathbf{0}_{n_y, n_x} & \mathbf{0}_{n_y, n_y} & \mathbf{0}_{n_y, n_x} & \mathbf{0}_{n_y, n_y} \\ \mathbf{0}_{n_x, n_x} & \mathbf{0}_{n_x, n_y} & \mathbf{C}_{1,1}^x & \mathbf{0}_{n_x, n_y} \\ \mathbf{0}_{n_y, n_x} & \mathbf{0}_{n_y, n_y} & \mathbf{0}_{n_y, n_x} & \mathbf{0}_{n_y, n_y} \end{pmatrix} \quad (20)$$

- 3) Compute a square root of the auto-correlation  $\mathbf{\Gamma}_{n,n+1|N}^t$  of  $([\mathbf{t}_{n|N}]^T, [\mathbf{t}_{n+1|N}]^T)^T$ , using the following QR decomposition:

$$\begin{pmatrix} [\mathbf{\Gamma}_{n,n+1|N}^t]^{\frac{1}{2}} \\ \mathbf{0}_{1, n_t} \end{pmatrix} = \mathbf{Q}^* \begin{pmatrix} [\mathbf{C}_{n,n+1|N}^t]^{\frac{1}{2}} \\ ([\hat{\mathbf{t}}_{n|N}]^T, [\hat{\mathbf{t}}_{n+1|N}]^T)^T \end{pmatrix} \quad (21)$$

All equations reported in this Section allow for a robust implementation of the PKF and the estimation of its parameters, according to the algorithm sketched in Fig. 1a. In order to illustrate the respective behavior of the original and robust EM algorithms, the following experiment was performed. A PKF signal with  $N = 100$  samples was simulated with parameters

$$\begin{aligned} \mathbf{Q}_0 &= \begin{pmatrix} \mathbf{I}_2 & \mathbf{0}_{2,2} \\ \mathbf{0}_{2,2} & \mathbf{0}_{2,2} \end{pmatrix}, \mathbf{Q} = \begin{pmatrix} 0.1\mathbf{I}_2 & \mathbf{0}_{2,2} \\ \mathbf{0}_{2,2} & \mathbf{I}_2 \end{pmatrix} \\ \mathbf{F}^{x,x} &= \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}, \mathbf{F}^{x,y} = \begin{pmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \end{pmatrix}, \\ \mathbf{F}^{y,x} &= \mathbf{I}_2, \mathbf{F}^{y,y} = \mathbf{0}_{2,2}, \end{aligned}$$

and  $\hat{\mathbf{t}}_0 = \mathbf{0}_{4,1}$ . Then parameters were learned from the sample using both original and robust EM algorithms, which were initialized with

$$\begin{aligned} \hat{\mathbf{t}}_0^{(0)} &= \hat{\mathbf{t}}_0, \mathbf{Q}_0^{(0)} = \mathbf{Q}_0, \\ \mathbf{F}^{(0)} &= \begin{pmatrix} \mathbf{I}_2 & \mathbf{0}_{2,2} \\ \mathbf{I}_2 & \mathbf{0}_{2,2} \end{pmatrix}, \mathbf{Q}^{(0)} = \begin{pmatrix} \mathbf{I}_2 & \mathbf{0}_{2,2} \\ \mathbf{0}_{2,2} & \mathbf{I}_2 \end{pmatrix} \end{aligned}$$

Fig. 1b shows the evolution of the smallest eigenvalue of matrix  $\mathbf{Q}$  according to iterations for the two EM algorithms. The original algorithm is destabilized from iteration 46, whereas the robust algorithm is not. We also observed, in experiments not reported here, a loss of positivity for  $\mathbf{Q}$ .

#### IV. EQUIVALENT PAIRWISE KALMAN FILTERS

Given series of observations, we first show that there exists several equivalent PKF, *i.e.* there exists several set of parameters  $\Theta$ , that give exactly the same likelihood for a given set of observations (*i.e.* the PKF is over-parametrized). To overcome this critical problem for applications, we next propose a simple constraint on the system in eq. (1) to fix a unique set of parameters when  $n_x = n_y$ .

##### A. Equivalent systems

From a PKF parameterized by  $\Theta = \{\hat{\mathbf{t}}_0, \mathbf{Q}_0, \mathbf{F}, \mathbf{Q}\}$ , it is possible to construct an  $M$ -equivalent filter parameterized by  $\Theta_M$  according to  $\hat{\mathbf{t}}_0^M = M\hat{\mathbf{t}}_0$ ,  $\mathbf{Q}_0^M = M\mathbf{Q}_0[M]^T$ ,  $\mathbf{F}^M = M\mathbf{F}M^{-1}$  and  $\mathbf{Q}^M = M\mathbf{Q}M^T$ , where matrix  $M$  is defined as

$$M = \begin{pmatrix} M^{x,x} & M^{x,y} \\ \mathbf{0}_{n_y, n_x} & \mathbf{I}_{n_y} \end{pmatrix}$$

with  $M^{x,x}$  invertible. In other words, the PKF is not identifiable.

The  $M$ -equivalent predicting, filtering and smoothing state expectations and covariance matrices are simply obtained from

$$\begin{aligned} \mathbf{x}_{n|p}^M &= M^{x,x} \mathbf{x}_{n|p} + M^{x,y} \mathbf{y}_{n-1} \\ \mathbf{P}_{n|p}^M &= M^{x,x} \mathbf{P}_{n|p} [M^{x,x}]^T \end{aligned} \quad (22)$$

with  $p > n - 2$ .

Whatever the equivalent PKF filter used, the likelihood whose logarithm writes  $\ln \mathcal{L}(\hat{\mathbf{x}}_0, \mathbf{P}_0, \mathbf{F}, \mathbf{Q}) \propto \sum_{n=0}^N \log(|\mathbf{S}_n|) + [\hat{\mathbf{y}}_n]^T [\mathbf{S}_n]^{-1} \tilde{\mathbf{y}}_n$  is not affected so that EM has strictly the same behavior.

Proof: denoting  $\mathbf{t}_n^M$  the  $M$ -equivalent vector of  $\mathbf{t}_n$ , we get

$$\begin{aligned} \begin{pmatrix} \mathbf{x}_n^M \\ \mathbf{y}_{n-1}^M \end{pmatrix} &= \underbrace{\begin{pmatrix} M^{x,x} & M^{x,y} \\ \mathbf{0}_{n_y, n_x} & \mathbf{I}_{n_y} \end{pmatrix}}_M \underbrace{\begin{pmatrix} \mathbf{x}_n \\ \mathbf{y}_{n-1} \end{pmatrix}}_{\mathbf{t}_n} \\ &= \begin{pmatrix} M^{x,x} \mathbf{x}_n + M^{x,y} \mathbf{y}_{n-1} \\ \mathbf{y}_{n-1} \end{pmatrix} \end{aligned}$$

The  $M$ -equivalent innovation expectation writes  $\tilde{\mathbf{y}}_n^M = \mathbf{y}_n^M - \hat{\mathbf{y}}_{n|n-1}^M = \mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1} = \tilde{\mathbf{y}}_n$ . The  $M$ -equivalent innovation covariance matrix writes  $\mathbf{S}_n^M = \mathbf{I}_{n_y} \mathbf{S}_n [\mathbf{I}_{n_y}]^T = \mathbf{S}_n$ . So, in conclusion, the  $M$ -equivalent system has the same likelihood than the original system and the parameters set  $\Theta_M^{(i)}$  that maximizes  $q(\Theta_M^{(i-1)}, \Theta)$  is exactly the  $M$ -equivalent of the parameters set  $\Theta^{(i)}$  that maximizes  $q(\Theta^{(i-1)}, \Theta)$ .

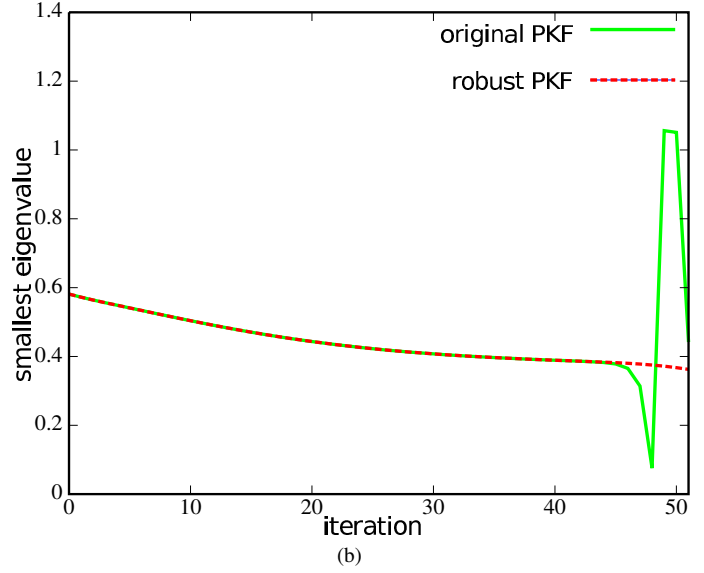
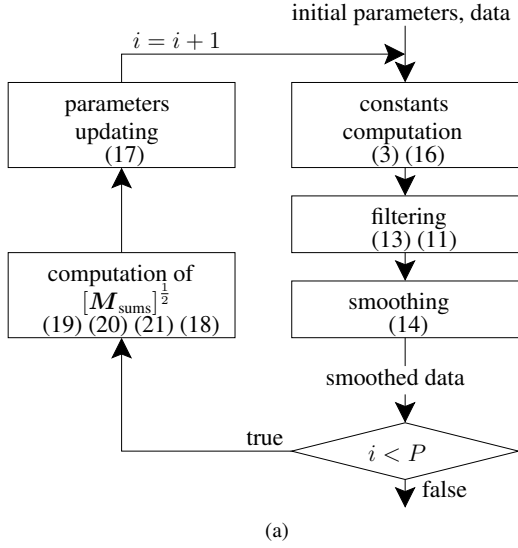


Fig. 1. (a) Flowchart of the robust blind PKF EM algorithm. (b) Evolution of the smallest eigenvalue of  $\mathbf{Q}$  for the original and robust PKF EM algorithms.

### B. Constraining the PKF

From an experimental point of view, equivalent filters can appear problematic since filtered and smoothed estimated states can be very different from one convergence of EM to another one and so from one initialization of EM to another one<sup>1</sup>. For example, the equivalent PKF defined by

$$\mathbf{M} = \begin{pmatrix} -\mathbf{I}_{n_x, n_x} & \mathbf{0}_{n_x, n_y} \\ \mathbf{0}_{n_x, n_x} & \mathbf{I}_{n_x} \end{pmatrix}$$

tends to restore a signal with opposite values with respect to the original signal.

For signal restoration, a simple hypothesis can be used to constraint parameter estimation by setting  $E(\mathbf{x}_n^M) = E(\mathbf{y}_n)$ . This hypothesis requires the matrix  $\mathbf{F}^{y,x}$  to be invertible. It can be easily shown that it constrains the matrix  $\mathbf{F}^M$  to have the form

$$\mathbf{F}^M = \begin{pmatrix} [\mathbf{F}^M]^{x,x} & [\mathbf{F}^M]^{x,y} \\ \mathbf{I}_{n_x, n_x} & \mathbf{0}_{n_x, n_x} \end{pmatrix}$$

The matrix  $\mathbf{M}$  which transforms  $\mathbf{F}$  to  $\mathbf{F}^M$  is given by

$$\mathbf{M} = \begin{pmatrix} \mathbf{F}^{y,x} & \mathbf{F}^{y,y} \\ \mathbf{0}_{n_y, n_x} & \mathbf{I}_{n_y} \end{pmatrix} \quad (23)$$

So, assuming that  $E(\mathbf{x}_n^M) = E(\mathbf{y}_n)$ , the PKF model writes

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{F}^{x,x} \mathbf{x}_n + \mathbf{F}^{x,y} \mathbf{y}_{n-1} + \boldsymbol{\omega}_{n+1}^x \\ \mathbf{y}_n &= \mathbf{x}_n + \boldsymbol{\omega}_{n+1}^y \end{aligned}$$

The dependency between the previous observation and the current one is dropped.

## V. EXPERIMENTS

In order to illustrate the nice behavior of the robust (Section III) and constrained (Section IV-B) PKF, we propose two series of experiments. The Mean Square Error (MSE) was used to compare the KF with the PKF. We did not compare the original PKF algorithm with respect to the robust one because the former produces wrong

<sup>1</sup>This remark does not refer to the well-known property of EM which searches for a local maximum.

estimate of parameters, resulting in highly degraded restored signals. Mean results are for 10,000 experiments of  $N = 100$  samples and with 100 EM iterations. The two experiments take about 15 minutes on a 2.3 GHz PC running Linux.

### A. Simulated PKF signals

All the 10,000 PKF signals were simulated according to parameters given in first row of Table I in left column. The parameters  $\hat{\mathbf{t}}_0$  and  $\mathbf{Q}_0$  were respectively set to  $(0, 0)^T$  and to  $\mathbf{I}$ . Then the simulations were smoothed, based on estimated parameters by EM, with both the KF and PKF robust and constrained algorithms. Initial parameters are given in row two, and final mean estimation in row three and four, for the PKF and KF respectively. An example of generated observations and hidden states are reported in Fig. 2a, and its restorations in Fig. 2b. Mean square restoration error for the 10,000 simulations are reported in Fig. 2c.

These experiments allow to draw these three main conclusions: (1) KF seems to over-smooth data, whereas PKF is more able to follow quick variations present in data; (2) as expected, the KF was not able to restore PKF simulated data as well as PKF (the MSE is approximately halved). (3) estimated parameters are very similar to the true parameters, even with initial values set far from the true ones.

### B. Sinusoidal signals

The same kind of experiments was conducted with simulated data that do not follow the PKF model (a sinusoid with a Gaussian noise with density  $\mathcal{N}(0, 0.1)$ ). The parameters  $\hat{\mathbf{t}}_0$  and  $\mathbf{Q}_0$  were respectively set to  $(0, 1.4)^T$  and to  $0.5 \mathbf{I}$ . Initial parameters for EM are given in row two of table I in right column, and final mean estimation in row three and four, for the PKF and KF respectively. An example of generated observations and hidden states are reported in Fig. 3a, and its filtered version in Fig. 3b. Mean square restoration error for the 10,000 simulations are reported in Fig. 3c.

These experiments allow to draw the following three main conclusions: (1) KF is not able at all to restore noisy data, as confirmed, in Table I, by the poor estimation of the measurement noise covariance; (2) Sinusoid shaped-signals are adapted to the PFK model, as

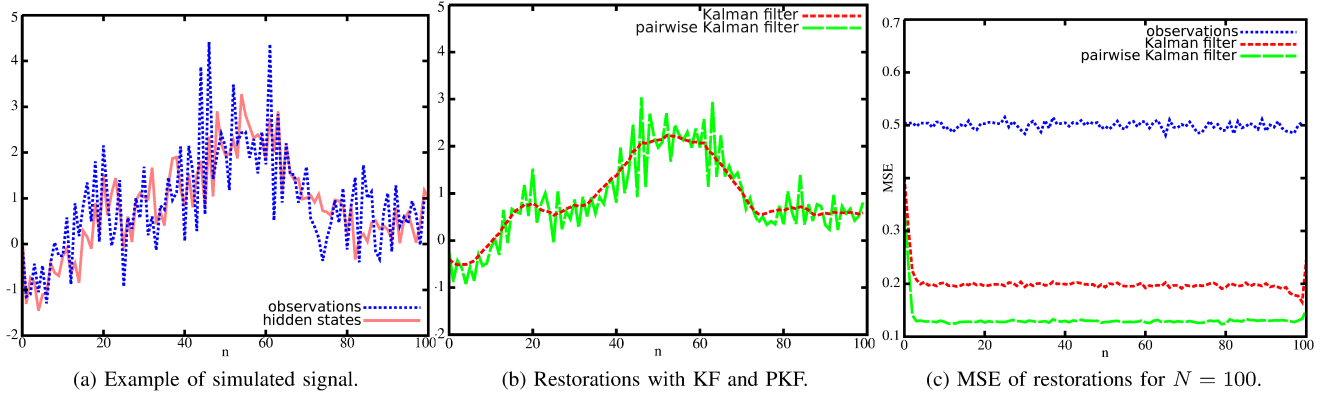


Fig. 2. Smoothed simulated signal (Section V-A).

TABLE I  
PARAMETERS USED IN SECTION V-A (SIMULATED SIGNAL) AND IN SECTION V-B (SINUSOIDAL SIGNAL). ROW 1: TRUE PARAMETERS, ROW 2: INITIAL VALUES FOR EM; ROWS 3 AND 4: ESTIMATED PARAMETERS.

		Simulated signal		Sinusoidal signal	
		$F$	$Q$	$F$	$Q$
1	$\Theta$	$\begin{pmatrix} 0.5 & 0.5 \\ 1.0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.1 & 0 \\ 0 & 0.5 \end{pmatrix}$	/	/
2	$\Theta^{(0)}$	$\begin{pmatrix} 1.0 & 0 \\ 1.0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.5 & 0 \\ 0 & var(y) \end{pmatrix}$	$\begin{pmatrix} 1.0 & 0 \\ 1.0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.1 & 0 \\ 0 & var(y) \end{pmatrix}$
3	$\Theta_{PKF}^{(100)}$	$\begin{pmatrix} 0.49 & 0.48 \\ 1.00 & 0.00 \end{pmatrix}$	$\begin{pmatrix} 0.18 & 0.05 \\ 0.05 & 0.39 \end{pmatrix}$	$\begin{pmatrix} 1.54 & -0.90 \\ 1.00 & 0.00 \end{pmatrix}$	$\begin{pmatrix} 0.05 & 0.09 \\ 0.09 & 0.16 \end{pmatrix}$
4	$\Theta_{KF}^{(100)}$	$\begin{pmatrix} 0.99 & 0 \\ 1.0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.06 & 0 \\ 0 & 0.50 \end{pmatrix}$	$\begin{pmatrix} 0.74 & 0 \\ 1.0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.5 & 0 \\ 0 & 0.01 \end{pmatrix}$

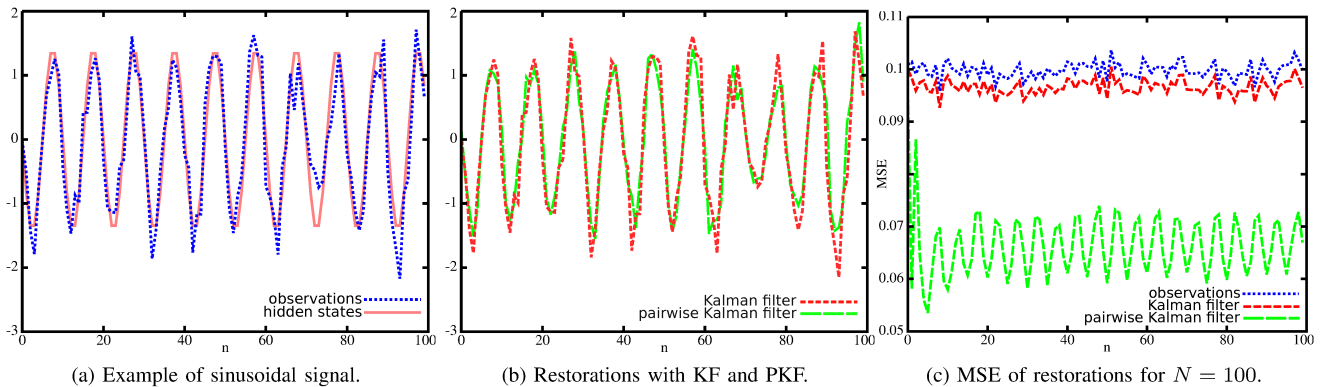


Fig. 3. Filtered sinusoidal signal (Section V-B).

confirmed by the low process noise covariance; (3) Once again, mean MSE was halved.

## VI. CONCLUSION

The main goal of this note was to propose factorization methods for making filtering, smoothing and parameters estimation robust for the pairwise linear system proposed in [3]. Then, among the equivalent systems we characterized, a simple condition was set to constrain the over-parametrized PKF to a filter which keeps the signal expectation.

Several experiments have been conducted on small-sized signals ( $N = 100$ ). Each time, EM algorithm has proved robust and efficient, even with initial parameters set far from the simulation ones. This nice behavior can be explained by the QR decompositions which guaranty the symmetry and positiveness of covariance matrices.

One interesting perspective for this work is to extend the robust EM algorithm to triplet Kalman filters [12], which are strict extensions

to the PKF. Most of equations remain the same except the constraint which should be rewritten.

## REFERENCES

- [1] W. Pieczynski and F. Desbouvries, "Kalman filtering using pairwise gaussian models," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'03)*, Hong-Kong, Apr. 2003.
- [2] S. Derrode and W. Pieczynski, "Unsupervised signal and image segmentation using pairwise Markov chains," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2477–2489, Sep. 2004.
- [3] B. Ait-el-Fquih and F. Desbouvries, "Bayesian smoothing algorithms in pairwise and triplet Markov chains," in *Proc. of the 2005 IEEE Workshop on Statistical Signal Processing (SSP'05)*, Bordeaux, France, Jul. 2005.
- [4] S. Julier, J. Uhlmann, and H. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *Proc. of the 1995 American Control Conf.*, Seattle, Washington, USA, Jun. 1995.

- [5] B. Ait-el-Fquih and F. Desbouvries, "Unsupervised signal restoration in partially observed Markov chains," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'06)*, Toulouse, France, May 2006.
- [6] J. Potter and R. Stem, "Statistical filtering of space navigation measurements," in *Proc. 1963 AIAA Guidance and Control Conf.*, Cambridge, Massachusetts, USA, Aug. 1963.
- [7] P. Kaminski, A. Bryson, and S. Schmidt, "Discrete square-root filtering: a survey of current techniques," *IEEE Trans. Autom. Control*, vol. 16, no. 6, pp. 727–736, Dec. 1971.
- [8] M. Morf and T. Kailath, "Square-root algorithms for least-squares estimation," *IEEE Trans. Autom. Control*, vol. 20, no. 4, pp. 487–497, Aug. 1975.
- [9] T. Kailath, A. Sayed, and B. Hassibi, *Linear estimation, Ch. 12: Array algorithms*. Prentice Hall, 2000.
- [10] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. of the Royal Statistical Society (B)*, vol. 39, no. 1, pp. 1–38, 1977.
- [11] R. Shumway and D. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *J. of Time Series Analysis*, vol. 3, no. 4, pp. 253–264, Jul. 1982.
- [12] B. Ait-el-Fquih and F. Desbouvries, "Kalman filtering in triplet Markov chains," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2957–2963, Aug. 2006.