Applied Mathematics & Optimization (2021) 84:399–423
https://doi.org/10.1007/s00245-019-09649-z

ORIGINAL PAPER

On the Anticipative Nonlinear Filtering Problem and Its Stability

Samy Tindel[1] · **Yanghui Liu**[1] · **Guang Lin**[1]

Published online: 8 January 2020

Abstract
In this paper, we consider an anticipative nonlinear filtering problem, in which the observation noise is correlated with the past of the signal. This new signal-observation model has its applications in both finance models with insider trading and in engineering. We derive a new equation for the filter in this context, analyzing both the nonlinear and the linear cases. We also handle the case of a finite filter with Volterra type observation. The performance of our algorithm is presented through numerical experiments.

Keywords Nonlinear filtering · Anticipative systems · Asymptotic stability · Volterra-type integral equations

1 Introduction

In its most classical setting, the filtering problem can be summarized as follows: let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, let $(\mathcal{F}_t,\ t \geq 0)$ be an increasing family of sub σ-fields of $\mathcal{F}$, and suppose that all $\mathbb{P}$-null sets belong to $\mathcal{F}_0$. We assume the existence of an underlying signal process $(X_t)_{t\geq 0}$ with values in $\mathbb{R}^m$ which can not be observed directly, and we are given an observation process $(Z_t)_{t\geq 0}$ with values in $\mathbb{R}^n$ which is related to $(X_t)_{t\geq 0}$ and disturbed by the noise process $(N_t)_{t\geq 0}$. A main task in the filtering theory is to estimate the signal process $(X_t)_{t\geq 0}$ based on the $(Z_t)_{t\geq 0}$. To give an example close enough to the situation which will be handled in the current

✉ Yanghui Liu
liu2048@purdue.edu

Samy Tindel
stindel@purdue.edu

Guang Lin
guanglin@purdue.edu

[1] Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, IN 47907, USA

paper, a model for the dynamics of both X and Z can be given as:

$$
\begin{aligned}
X_t &= X_0 + \int_0^t a(X_s)ds + W_t, \\
Z_t &= \int_0^t h(X_s)ds + N_t,
\end{aligned}
\tag{1.1}
$$

where X and Z are respectively the signal and the observation process, and W and N are standard Brownian motions (or martingales), both adapted to the filtration $\mathcal{F}_t$. In (1.1), we also assume that the noises W and N are decorrelated for convenience.

We refer to [8, Chapters 2, 3], [17, Chapter 8] and [22] for a detailed account on the classical model (1.1) in full generality. Although we cannot give a real overview of the vast literature on filtering in this short introduction, let us mention that a first basic step in order to solve the nonlinear filtering problem (1.1) is to obtain an equation for the unnormalized conditional density of the signal X_t given Z. This equation turns out to be a linear stochastic PDE called Zakai's equation (see [29]), which is one of the most classical objects studied in stochastic analysis. Then the numerical methods either rely on a particle representation of Zakai's equation (see [10,19]) or on a direct discretization of Zakai's equation [13,23]. The reader is sent to [8] for the abundant literature on this topic. We should also mention a different direction which aims at taking into account possible time correlations for both the signal and observation noises. This is mainly achieved by considering a fractional Brownian motion model for the noises W and N (see e.g [7,18]), for which a complete answer in the nonlinear case is still a challenging issue.

In the current contribution, we wish to go back to one of the fundamental assumptions in Eq. (1.1), namely the fact that the initial condition X_0 and the observation noise N are independent (which follows from the fact that N is $\mathcal{F}_t$-Brownian motion). While this assumption seems to be natural at first sight, one can argue that this hypothesis is violated in many interesting situations. Among the possible applications we have in mind, let us mention the following:

(i) Consider a classical target tracking problem (see e.g. [14, Chapter 5]). Suppose that the misspecification of the initial condition occurs due to temperature, wind conditions or other environmental variables. Since the observation noise of the system is usually influenced by the same factors, it is often an oversimplification to assume that the initial conditions and the observation noise are completely independent.

(ii) Another important context in which X_0 and N cannot be assumed to be independent concerns finance models with insider trading (see e.g. [2,6,21] for the insider problem and [27, Sect. 1.1.2 and Example 5.8] for applications of filtering to finance models), where some investors of a public company's stock have access to nonpublic information about the company. In such models the signal is the nonpublic information of that company, and the observation is stock price and other public information. In this situation, we expect the observation noise to be related with X_0, since the initial nonpublic information creates some fluctuations on the stock price.

With those potential applications in mind, in the sequel we will see how the classical filtering problem is modified when X_0 and N are correlated, by considering X_0 as an anticipative random variable (the reader might think of X_0 as a Wiener integral of the form $\int_0^\infty \phi_s \, dN_s$ for a deterministic function $\phi \in L^2(\mathbb{R}_+)$ in order to have a concrete example in mind). Notice that explicit real-valued optimal filters for this anticipative filtering problem have been obtained only recently; see [1], where the authors assume that the signal $X_t = X_0$ is constant in time and correlated to the observation noise, and where the observation is linearly dependent on X.

The aim of this paper is thus to consider the aforementioned anticipative filtering problem in the general setting given by (1.1), where we just assume the following general correlation property between the initial condition X_0 and the noise N.

Hypothesis 1.1 Let X_0, W and N be given as in Eq. (1.1). We assume that the family $(X_0, W_t, N_t, 0 \le t \le T)$ is Gaussian, such that the correlation

$$\rho_N(t) := \mathbb{E}(N_t X_0^T), \qquad 0 \le t \le T \tag{1.2}$$

is a function in $C^2([0, T], \mathbb{R}^{n\times m})$.

Within this general framework, we will focus on the following issue concerning the filtering problem:

(i) In the general nonlinear case given by Eq. (1.1), we derive a Zakai type equation for the unnormalized filter, as well as a Kushner-FKK type stochastic equation for the normalized filter. As the reader will see, the anticipative nature of our problem will affect all the coefficients of those equations.
(ii) Whenever a linear situation is considered in the system (1.1), we get a modified version of the Kalman-Bucy filter. As in the general case alluded to in (i), the coefficients of the linear filter are nontrivially affected by the correlation ρ_N. However, the Kalman-Bucy filter derived in Sect. 4 is very convenient to implement numerically.
(iii) If we further assume that the correlation ρ_N defined by (1.2) is compactly supported, then as expected, we will be able to give some stability results for the anticipative Kalman-Bucy filter. More specifically, we show that for large times the difference between the anticipative and non-anticipative Kalman-Bucy filters converges exponentially fast to 0.
(iv) Still in the linear case, we handle the case of a weighted Volterra type observation Z and get the corresponding expression for the anticipative finite dimensional optimal filter. This should be seen as an alternative point of view on [7], where a nonlinear Voterra type signal-observation system had been considered. Our method yields a straightforward implementation for the computation of the conditional mean and variance of the signal X.
(v) Our simulation section will be focused on a classical radar-tracking example, where the vehicle's initial location is correlated with the observation noise. The simulations show a significant improvement on the estimation accuracy whenever the anticipative filter is used.

Remark 1.2 Our anticipative filtering problem can be considered as a particular instance of filtering problems with path-space valued signal and observation processes. These path-space models allow to consider diffusion-type signal-observation processes that depend on the whole history of the process. For discrete time problems these path-space filtering models and their genealogical tree based particle approximations are rather well understood, see for instance [11]. A stochastic analysis for the continuous time version of path-space models have been developed in [12] for filtering purposes, and in [5,16,26] in a mathematical finance context. These papers are based on elegant functional approaches allowing to write the analog of the traditional Kallianpur-Striebel formula, as well as the Kushner-Stratonovich equations, with innovations processes and generators of processes evolving on path spaces.

However, we should mention that our methods yield an algorithm which does not require the introduction of genealogical trees. In particular, in case of a linear system our filtering equation will be finite dimensional (as opposed to the genealogical trees obtained in the path-space valued SPDEs context [3]). An interesting task is thus to evaluate the numerical performance of our algorithm in a linear situation (see Sect. 6 for an implementation). In the general nonlinear case, however, both our algorithm and the path-dependent algorithm are based on infinite dimensional filtering equations. Hence a careful numerical algorithm comparison is required. We will deal with this problem in some subsequent papers.

The basic technique we will resort to in order to deal with the anticipative filtering framework given by (1.1) can be summarized as follows.
(a) We first rely on a general result concerning enlargements of filtrations. Namely, we will show that one can modify N by a simple enough drift so that it becomes a $\mathcal{F}_t$-Brownian motion (let us insist again on the fact that the system filtration $\mathcal{F}$ includes the information on X_0). This general additional drift is then reflected in the coefficients of the filtering equations. Note that the general enlargement of filtration result we invoke is interesting in its own right. It can be seen as a multidimensional generalization of [15] and is detailed in Sect. 2.
(b) Once the enlargement of filtration result is obtained, another key ingredient in the construction of the filter is a convenient introduction of some auxiliary signal process. Therefore, at the price of increasing slightly the dimension of our system and changing its coefficients, we will be able to go back to a more standard filtering setting.

Notation 1.3 For any integrable process G_t, $t \geq 0$, we denote by $\mathcal{F}_t^G$, $t \geq 0$ the filtration $\sigma(G_s, s \in [0, t])$. For convenience, we also write $\hat{G}_t := \mathbb{E}[G_t | \mathcal{F}_t^Z]$ and $\tilde{G}_t := G_t - \hat{G}_t$, where Z is the observation process.

2 Enlargement of Filtration

When we consider an anticipative model like (1.1) under Hypothesis 1.1, one of the main problems is the following: while N can be seen as a standard Brownian motion in $\mathcal{F}^N$, it is no longer a Brownian motion with respect to the system filtration $\mathcal{F}$. Having this problem in mind, in this section we show that there is a simple transformation of N that makes it a Brownian motion in the enlarged filtration. This result will be first handled in a general framework. As mentioned in the introduction, it should be considered as a multidimensional generalization of [15].

Lemma 2.1 *Let $(B_t, 0 \le t \le T)$ be a standard n-dimensional Brownian Motion, and let $X = (X_1, X_2, \dots, X_m)^T$ be a centered (mean zero) Gaussian random vector in $\mathbb{R}^m$ with covariance $\Sigma \in \mathbb{R}^{m\times m}$. Suppose that $\mathbb{E}[B_t X^T] = \rho(t)$, where $\rho \in C^2([0,T];\mathbb{R}^{n\times m})$. In addition, we assume that the family $\{X, B_t; 0 \le t \le T\}$ is jointly Gaussian. We consider a function $g(t) \in C([0,T];\mathbb{R}^{n\times m})$ defined on $[0,T]$ (see Lemma 2.2 for the existence of such a function) such that*

$$g'(t)\left(\Sigma - \int_0^t \rho'(u)^T \rho'(u)du\right) = \rho'(t), \quad g(0) = 0. \tag{2.1}$$

Let also $\lambda = \{\lambda(t,s); 0 \le s \le t \le t\}$ be the two-parameter $\mathbb{R}^{n\times n}$-valued function defined by

$$\lambda(t,s) = g(t)p(s) + q(s), \tag{2.2}$$

where $p \in C([0,T];\mathbb{R}^{m\times n})$ and $q \in C([0,T];\mathbb{R}^{n\times n})$ are respectively given by $p(s) = \rho''(s)^T$ and $q(s) = -g(s)\rho''(s)^T - g'(s)\rho'(s)^T$. Then the process $\tilde{B}$ defined by

$$\tilde{B}_t = B_t - \int_0^t \lambda(t,u)B_u du - g(t)X \tag{2.3}$$

is a n-dimensional $(\mathcal{G}_t)$-Brownian motion, where $(\mathcal{G}_t)$ is the augmented filtration $\sigma(X, B_s; 0 \le s \le t)$, $0 \le t \le T$.

Proof We are going to show that $\tilde{B}_t$ is a $\mathcal{G}$-martingale. Then, taking into account Lévy's characterization and the continuity of $\tilde{B}$, we can conclude immediately that $\tilde{B}$ is a standard $\mathcal{G}$-Brownian motion.

On a Gaussian space, it is well known that decorrelation implies independence. Therefore, in order to show that $\tilde{B}$ is a $\mathcal{G}$-martingale, it suffices to show that

$$\mathbb{E}\{(\tilde{B}_t - \tilde{B}_s)X^T\} = 0 \quad \text{and} \quad \mathbb{E}\{(\tilde{B}_t - \tilde{B}_s)B_r^T\} = 0, \quad 0 \le r \le s \le t. \tag{2.4}$$

Note that by (2.3) and the identities $\mathbb{E}[XX^T] = \Sigma$ and $\mathbb{E}[B_t X^T] = \rho(t)$, the first equation of (2.4) is equivalent to

$$\begin{aligned} 0 = &\left(\rho(t) - \int_0^t \lambda(t,u)\rho(u)du - g(t)\Sigma\right) \\ &- \left(\rho(s) - \int_0^s \lambda(s,u)\rho(u)du - g(s)\Sigma\right). \end{aligned} \tag{2.5}$$

We will now focus on the proof of (2.5). To this aim observe that, due to the fact that $\rho(0) = 0$ and $g(0) = 0$ (see relation (2.1)), Eq. (2.5) is equivalent to

$$0 = \rho(t) - \textstyle\int_0^t \lambda(t,u)\rho(u)du - g(t)\Sigma\,. \tag{2.6}$$

We are now reduced to the proof of (2.6).

Denote by ∂_s the partial derivative with respect to the parameter s. Thanks to the definition (2.2) of $\lambda(t,s)$, the reader can easily check that λ satisfies the following relation:

$$\lambda(t,s) = \partial_s \left(g(t)\rho'(s)^T - g(s)\rho'(s)^T \right). \tag{2.7}$$

Substituting (2.7) into the right side of (2.6) and then by integration by parts (together with the fact that $\rho(0)=0$) we obtain

$$\begin{aligned} \rho(t) - \int_0^t \lambda(t,u)\rho(u)du - g(t)\Sigma &= \rho(t) \\ &+ \int_0^t [g(t)\rho'(u)^T - g(u)\rho'(u)^T]\rho'(u)du - g(t)\Sigma \\ &= \rho(t) + \int_0^t [g(t) - g(u)]\rho'(u)^T\rho'(u)du - g(t)\Sigma. \end{aligned} \tag{2.8}$$

We now apply integration by parts again to $U(u) = g(t) - g(u)$ and $dV(u) = \rho'(u)^T\rho'(u)du$, which yields

$$\int_0^t [g(t) - g(u)]\rho'(u)^T\rho'(u)du = \int_0^t g'(u)\int_0^u \rho'(s)^T\rho'(s)dsdu. \tag{2.9}$$

Substituting (2.9) into (2.8) and writing $g(t)\Sigma = \int_0^t g'(u)\Sigma du$ we obtain

$$\begin{aligned} \rho(t) \ - \int_0^t \lambda(t,u)\rho(u)du - g(t)\Sigma &= \rho(t) \\ &+ \int_0^t g'(u)\left(\int_0^u \rho'(s)^T\rho'(s)ds - \Sigma\right)du. \end{aligned} \tag{2.10}$$

Recalling that g satisfies (2.1), it is now readily checked that the right-hand side of (2.10) vanishes. Hence we have proved our claim (2.6), which in turn proves the first assertion of relation (2.4).

Let us turn to the second equation of (2.4). Similarly to (2.5) and recalling that λ is a $\mathbb{R}^{n\times n}$-valued function, the second equation of (2.4) can be written as

$$\begin{aligned} 0 &= -\int_0^t \lambda(t,u)\begin{bmatrix} u\wedge r & & 0 \\ & \ddots & \\ 0 & & u\wedge r \end{bmatrix}du + \int_0^s \lambda(s,u)\begin{bmatrix} u\wedge r & & 0 \\ & \ddots & \\ 0 & & u\wedge r \end{bmatrix}du \\ &\quad - g(t)\rho(r)^T + g(s)\rho(r)^T \\ &= -r\int_r^t \lambda(t,u)du - \int_0^r \lambda(t,u)udu + r\int_r^s \lambda(s,u)du + \int_0^r \lambda(s,u)udu \\ &\quad - g(t)\rho(r)^T + g(s)\rho(r)^T. \end{aligned} \tag{2.11}$$

In the following, we show that (2.11) holds. Note that, owing to the fact that $g(0) = 0$, it is clear that (2.11) holds for $r = 0$, so Eq. (2.11) is equivalent to

$$0 = -\int_r^t \lambda(t,u)du + \int_r^s \lambda(s,u)du - g(t)\rho'(r)^T + g(s)\rho'(r)^T, \tag{2.12}$$

which is obtained by differentiating both sides of (2.11) with respect to r. Furthermore, along the same lines as for (2.6) and invoking the fact that $\int_r^s \lambda(s,u)du = 0$ when $s = r$, it is easy to see that Eq. (2.12) is equivalent to

$$0 = -\int_r^t \lambda(t,u)du - g(t)\rho'(r)^T + g(r)\rho'(r)^T. \tag{2.13}$$

Now relation (2.13) follows immediately from (2.7). Plugging this information in our previous considerations, we get that (2.12) and therefore (2.11) hold true. This proves that the second equation of (2.4) is satisfied. The proof is complete. □

With Lemma 2.1 in hand, notice that one can derive an explicit formula for $g(t)$ from (2.1) if the symmetric matrix $\Sigma - \int_0^t \rho'(s)^T\rho'(s)ds$ is non-singular. The following result provides an equivalent condition to the non-singularity of $\Sigma - \int_0^t \rho'(s)^T\rho'(s)ds$.

Lemma 2.2 *Let the assumptions be as in Lemma 2.1. Then*

(a) *The following identity holds for all $t \geq 0$:*

$$\Sigma - \int_0^t \rho'(s)^T\rho'(s)ds = \mathbb{E}\left(\left(X - \mathbb{E}(X|\mathcal{F}_t^B)\right)\left(X - \mathbb{E}(X|\mathcal{F}_t^B)\right)^T\right).$$

(b) *The matrix $\Sigma - \int_0^t \rho'(s)^T\rho'(s)ds$ is non-singular for all $t \in [0,T]$ whenever $X \notin \mathcal{F}_T^B$.*

(c) *If $\rho'(t_0) \neq 0$ for some $t_0 > 0$, then $\Sigma - \int_0^t \rho'(s)^T\rho'(s)ds$ is non-singular for all $t \in [0, t_0]$.*

Proof Note first that the $\mathcal{F}^B$-Gaussian martingale $\mathbb{E}(X|\mathcal{F}_t^B)$ can be represented as a Wiener integral $\int_0^t f_s dB_s$, where $f_s \in L_2([0,T], \mathbb{R}^{m\times n})$. By the definition of $\rho(t)$ it is easy to show that $f_t = \rho'(t)^T$. Therefore, we have

$$\begin{aligned}\mathbb{E}\left(\left(X - \mathbb{E}(X|\mathcal{F}_t^B)\right)\left(X - \mathbb{E}(X|\mathcal{F}_t^B)\right)^T\right) &= \Sigma - \mathbb{E}\left(\mathbb{E}(X|\mathcal{F}_t^B)\mathbb{E}(X|\mathcal{F}_t^B)^T\right)\\ &= \Sigma - \int_0^t \rho'(s)^T\rho'(s)ds,\end{aligned}$$

which finishes the proof of our assertion (a).

We turn to the proof of (b). Note that the matrix $\mathbb{E}\{[X - \mathbb{E}(X|\mathcal{F}_t^B)][X - \mathbf{E}(X|\mathcal{F}_t^B)]^T\}$ is singular if and only if there exists a constant vector $(k_1, k_2, \ldots, k_m) \neq 0$ such that $\Sigma_{i=1}^m k_i X_i \in \mathcal{F}_t^B$. In other words, $\mathbb{E}\{[X - \mathbb{E}(X|\mathcal{F}_t^B)][X - \mathbf{E}(X|\mathcal{F}_t^B)]^T\}$ is nonsingular if and only if $\Sigma_{i=1}^m k_i X_i \notin \mathcal{F}_t^B$ for any $(k_1, k_2, \dots, k_m)$. But this holds whenever $X \notin \mathcal{F}_t^B$. This completes the proof of (b).

In order to prove (c), suppose now that $\rho'(t_0) \neq 0$ for some $t_0 > 0$. Then invoking the continuity of ρ' we can find a positive number $\varepsilon > 0$ such that $\rho'(s) \neq 0$ for $s \in [t_0, t_0 + \varepsilon]$. On the other hand, we have shown at the beginning of the proof that

$$\mathbb{E}(X|\mathcal{F}_{t_0+\varepsilon}^B) = \int_0^{t_0+\varepsilon} \rho'(s) dB_s = \int_0^{t_0} \rho'(s) dB_s + \int_{t_0}^{t_0+\varepsilon} \rho'(s) dB_s.$$

Consider now a given $t \in [0, t_0]$ and note that since $\rho'(s) \neq 0$ for $s \in [t_0, t_0 + \varepsilon]$, we have $\mathbb{E}(X|\mathcal{F}_{t_0+\varepsilon}^B) \notin \mathcal{F}_t^B$. Therefore it is readily checked that $X = \mathbb{E}(X|\mathcal{F}_{t_0+\varepsilon}^B) + (X - \mathbb{E}(X|\mathcal{F}_{t_0+\varepsilon}^B)) \notin \mathcal{F}_t^B$. We can now apply directly item (b) and we conclude that $\Sigma - \int_0^t \rho'(s)^T \rho'(s) ds$ is non-singular for all $t \in [0, t_0]$. The proof is complete. $\square$

We can now give an explicit version for the relation (2.1) of g under non-degeneracy assumptions in terms of ρ. This follows immediately from Lemma 2.2 (c).

Corollary 2.3 *Let the assumptions be as in Lemma 2.1. Let $T_0 = \sup\{t \geq 0 : \rho'(t) \neq 0\} \in [0, \infty]$. For $t < T_0$ we consider a function g defined by $g(0) = 0$ and*

$$g'(t) = \rho'(t) \left(\Sigma - \int_0^t \rho'(u)^T \rho'(u) du \right)^{-1}.$$

We also set $g'(t) = 0$ for $t \geq T_0$. Then the function g satisfies Eq. (2.1). In particular, it is always possible to find a function g such that (2.1) is satisfied.

3 Anticipative Filtering Equation: Nonlinear Case

In this section, we go back to the anticipative filtering problem (1.1), which is recalled here for the reader's convenience:

$$\begin{aligned} X_t &= X_0 + \int_0^t a(X_s) ds + W_t, \\ Z_t &= \int_0^t h(X_s) ds + N_t, \end{aligned} \tag{3.1}$$

where the family $(X_0, W_t, N_t; 0 \leq t \leq T)$ satisfies Hypothesis 1.1. In particular, we assume that $\rho_N(t) := \mathbb{E}(N_t X_0^T)$, $0 \leq t \leq T$ is a function in $C^2([0, T], \mathbb{R}^{n \times m})$.

In order to derive an equation for the optimal filter, let us first see how Lemma 2.1 allows us to reduce our computations to an adaptive signal-observation system with modified coefficients.

Lemma 3.1 *Let (X, Z) be the solution of (3.1) and assume that Hypothesis 1.1 is satisfied. Let p, q, λ, and g be functions defined as in Lemma 2.1 with ρ replaced by*

ρ_N. Set

$$\tilde{N}_t = N_t - \int_0^t \lambda(t,u) N_u du - g(t) X_0. \tag{3.2}$$

The following statements holds:

(a) *$\tilde{N}$ is a $\mathbb{R}^n$-valued $\mathcal{F}$-Brownian motion.*
(b) *Consider the $\mathbb{R}^m$-valued process $(\bar{X}_t)_{t\leq T}$ defined by:*

$$\bar{X}_t = X_0 + \int_0^t p(s) N_s ds. \tag{3.3}$$

In addition, denote

$$r(t) = \lambda(t,t) = -g'(t)\rho'(t)^T, \tag{3.4}$$

and observe that $r(t) \in \mathbb{R}^{n\times n}$ for all $t \leq T$. We define a process $(U_t)_{t\leq T}$ such that $U_t \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n$ and some coefficients b, σ, c as follows:

$$U = \begin{bmatrix} X \\ \bar{X} \\ N \end{bmatrix}, \qquad b(U_t) = \begin{bmatrix} a(X_t) \\ p(t)N_t \\ g'(t)\bar{X}_t + r(t)N_t \end{bmatrix},$$
$$\sigma = \begin{bmatrix} I_m \\ 0 \\ 0 \end{bmatrix}, \qquad c = \begin{bmatrix} 0 \\ 0 \\ I_n \end{bmatrix}. \tag{3.5}$$

Eventually, define a $\mathbb{R}^m$-valued coefficient k by:

$$k(U_t) = h(X_t) + g'(t)\bar{X}_t + r(t)N_t\,. \tag{3.6}$$

Then (X, Z) satisfies a signal-observation system expressed in terms of (U, Z):

$$dU_t = b(U_t)dt + c\,d\tilde{N}_t + \sigma dW_t \tag{3.7}$$
$$dZ_t = k(U_t)dt + d\tilde{N}_t. \tag{3.8}$$

(c) *The augmented system (3.7)–(3.8) is now governed by $(W, \tilde{N})$, which is a $\mathcal{F}$-Brownian motion.*

Proof Item (a) follows from a direct application of Lemma 2.1 with $B = N$, $X = X_0$ and $\rho = \rho_N$.

We turn to the proof of (b). First, plugging (3.2) into (3.1) we can write

$$Z_t = \int_0^t h(X_s)ds + \int_0^t \lambda(t,s)N_s ds + g(t)X_0 + \tilde{N}_t\,. \tag{3.9}$$

Recall that in (2.2) we have defined λ as $\lambda(t,s) = g(t)p(s) + q(s)$. Therefore, we can write the second and third terms of the right side of (3.9) as

$$\begin{aligned}\int_0^t \lambda(t,s)N_s ds + g(t)X_0 &= g(t)\left(\int_0^t p(s)N_s ds + X_0\right) + \int_0^t q(s)N_s ds \\ &= g(t)\bar{X}_t + \int_0^t q(s)N_s ds,\end{aligned}$$

where the second relation stems from (3.3). We now apply the elementary relation $\alpha_t\beta_t = \alpha_0\beta_0 + \int_0^t \alpha_s'\beta_s ds + \int_0^t \alpha_s\beta_s' ds$ to the C^1-functions $\alpha_t = g(t)$ and $\beta_t = \bar{X}_t$, which yields

$$\begin{aligned}\int_0^t \lambda(t,s)N_s ds + g(t)X_0 &= \int_0^t g'(s)\bar{X}_s ds + \int_0^t g(s)p(s)N_s ds + \int_0^t q(s)N_s ds \\ &= \int_0^t g'(s)\bar{X}_s ds + \int_0^t r(s)N_s ds,\end{aligned} \tag{3.10}$$

where the last equality is due to the definition (3.4) of r. Reporting (3.10) into (3.9), and taking the definition (3.6) of k into account,

$$\begin{aligned}Z_t &= \int_0^t h(X_s)ds + \int_0^t g'(s)\bar{X}_s ds + \int_0^t r(s)N_s ds + \tilde{N}_t \\ &= \int_0^t k(U_s)ds + \tilde{N}_t,\end{aligned}$$

which is equation in (3.8).

In the following, we derive the Eq. (3.7) for U. Note again that by (3.2) and taking into account (3.10) we obtain

$$\begin{aligned}N_t &= \int_0^t \lambda(t,s)N_s ds + g(t)X_0 + \tilde{N}_t \\ &= \int_0^t g'(s)\bar{X}_s ds + \int_0^t r(s)N_s ds + \tilde{N}_t.\end{aligned} \tag{3.11}$$

In order to get the equation for the process U given by (3.5), it is thus sufficient to combine Eq. (3.1) for X, Eq. (3.3) for $\bar{X}$ and relation (3.11) for N. $\square$

In Lemma 3.1, let us highlight again the fact that the new system (3.7)–(3.8) is governed by a $\mathcal{F}$-Brownian motion $(W, \tilde{N})$. Hence we have reduced our anticipative problem to a classical filtering equation with modified coefficients. In order to give specific statements in this context, we now recall some basic notation.

Notation 3.2 In the context of Lemma 3.1, set

$$M_t = \exp\left(\int_0^t k(U_s)^T dZ_s - \frac{1}{2}\int_0^t |k(U_s)|^2 ds\right).$$

We also denote by $\tilde{\mathbb{P}}$ the measure on Ω that is absolutely continuous with respect to $\mathbb{P}$ with a Radon-Nickodym derivative on $(\Omega, \mathcal{F}_t)$ given by:

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}\Big|_{\mathcal{F}_t} = M_t^{-1},$$

where recall that $\mathcal{F}_t$ is the system filtration.

With our modified setting in hand, the classical filtering results contained e.g. in [27] yield the following result.

Theorem 3.3 *Let (X, Z) be the solution of (3.1) and let U be as in Lemma 3.1. Then*
(a) *The optimal filter $\pi_t : \langle \pi_t, f\rangle = \mathbb{E}(f(U_t)|\mathcal{F}_t^Z)$ satisfies the following nonlinear stochastic differential equation on $[0, T]$:*

$$\begin{aligned}\langle \pi_t, f\rangle &= \langle \pi_0, f\rangle + \int_0^t \langle \pi_s, Lf\rangle ds \\ &\quad + \int_0^t (\langle \pi_s, \nabla f c + f k^T\rangle - \langle \pi_s, f\rangle\langle \pi_s, k^T\rangle) d\nu_s, \end{aligned} \tag{3.12}$$

for all $f \in C_b^2(\mathbb{R}^{2m+n})$, where ν is the innovation process defined by $\nu_t = Z_t - \int_0^t \langle \pi_s, k\rangle ds$ and where ∇f stands for the vector $\nabla f = (\partial_1 f, \dots, \partial_d f)$.
(b) *Let V_t be the unnormalized filter, defined by:*

$$\langle V_t, f\rangle = \tilde{\mathbb{E}}(M_t f(U_t)|\mathcal{F}_t^Z),$$

where $\tilde{\mathbb{E}}$ refers to the expectation with respect to the measure $\tilde{\mathbb{P}}$. Then V satisfies the following linear stochastic differential equation (usually called Zakai's equation):

$$\langle V_t, f\rangle = \langle V_0, f\rangle + \int_0^t \langle V_s, Lf\rangle ds + \int_0^t \langle V_s, \nabla f c + f k^T\rangle dZ_s,$$

where the second order differential operator L is defined by

$$Lf = \frac{1}{2}\sum_{i,j=1}^d A_{ij}\partial_{ij}^2 f + \sum_{i=1}^d b_i \partial_i f. \tag{3.13}$$

In (3.13), we have also set $A = cc^T + \sigma\sigma^T$, and d is the dimension of U.

Remark 3.4 In (3.12) we have obtained an equation for the augmented signal U_t, while we are originally interested in the optimal filter $\tilde{\pi}_t(\cdot) := \mathbb{P}(X_t \in \cdot|\mathcal{F}_t^Z)$. However, one can easily derive an equation for $\tilde{\pi}$ by taking $f \in C_b^2(\mathbb{R}^m)$ in Eq. (3.12) (that is freezing the $\bar{X}$ and N components in (3.12)). Therefore the equation for $\tilde{\pi}_t$ is in the same form as (3.12), and it still depends on $\bar{X}$ and N.

Remark 3.5 The uniqueness of the solution for the filtering Eq. (3.12) can be obtained by applying some classical uniqueness results for nonlinear SPDEs (see e.g. [19]). For sake of conciseness, we leave to the patient reader the task of checking that the usual conditions of [19] are satisfied in our case.

4 Anticipative Filtering Equation: Linear Case

This section is devoted to a particularization of problem (3.1) to a linear context. As usual in filtering theory, we will see that more explicit solutions to the filtering problem can be computed in this case. We also study the asymptotic stability of the filter in this framework.

4.1 Filter Equations

Let us specify the filtering system we consider in this linear case. Namely, the couple (X, Z) is assumed to satisfy the following system:

$$X_t = X_0 + \int_0^t a(s) X_s ds + \sigma_0 W_t, \tag{4.1}$$

$$Z_t = \int_0^t h(s) X_s ds + N_t\,, \tag{4.2}$$

where $X_t \in \mathbb{R}^m$, $Z_t \in \mathbb{R}^n$, $W_t \in \mathbb{R}^l$, $a(s) \in \mathbb{R}^{m\times m}$, $\sigma_0 \in \mathbb{R}^{m\times l}$, and $h(s) \in \mathbb{R}^{n\times m}$.

In our linear context, we still define $\bar{X}$ by (3.3). We also define an $\mathbb{R}^{2m+n}$-valued augmented signal U and some augmented coefficients b, σ, c and k by

$$U = \begin{bmatrix} X \\ \bar{X} \\ N \end{bmatrix}, \qquad b(t) = \begin{bmatrix} a(t) & 0 & 0 \\ 0 & 0 & p(t) \\ 0 & g'(t) & r(t) \end{bmatrix}, \qquad \sigma = \begin{bmatrix} \sigma_0 \\ 0 \\ 0 \end{bmatrix}, \qquad c = \begin{bmatrix} 0 \\ 0 \\ I_n \end{bmatrix},$$

and

$$k(t) = (h(t), g'(t), r(t)),$$

where a is defined by (4.1), p and g are introduced in Lemma 2.1 and r is given in Lemma 3.1. It then follows from Lemma 3.1 that the linear system (4.2) is equivalent to the following regular Kalman-Bucy signal-observation system:

$$\begin{aligned} dU_t &= b(t) U_t dt + c d\tilde{N}_t + \sigma dW_t \\ dZ_t &= k(t) U_t dt + d\tilde{N}_t. \end{aligned} \tag{4.3}$$

Let us also recall that for the linear filtering problem (4.3), the optimal filter π_t is obtained as the following regular conditional law:

$$\pi_t = \mathcal{N}(\hat{U}_t, P_t), \tag{4.4}$$

where $\hat{U}_t = \mathbb{E}[U_t|\mathcal{F}_t^Z]$ and $P_t = \mathbb{E}((U_t - \hat{U}_t)(U_t - \hat{U}_t)^T|\mathcal{F}_t^Z)$ designates the conditional variance of U_t given $\mathcal{F}_t^Z$. The following theorem specifies the expressions of $\hat{U}$ and P:

Theorem 4.1 *Let $\hat{U}$ and P be the conditional mean and covariance of U given by* (4.4). *Then:*

(i) $\hat{U}_t$ *solves the equation*

$$\hat{U}_t = \hat{U}_0 + \int_0^t b(s)\hat{U}_s ds + \int_0^t (c + P_s k(s)^T) d\nu_s, \tag{4.5}$$

where the innovation process ν is given by $\nu_t = Z_t - \int_0^t k(s)\hat{U}_s ds$.

(ii) *The $\mathbb{R}^{2m+n,2m+n}$-valued conditional variance P satisfies a Riccati equation of the form:*

$$P_t' = P_t b(t)^T + b(t)P_t + A - (c + P_t k(t)^T)(c + P_t k(t)^T)^T, \tag{4.6}$$

where $A = cc^T + \sigma\sigma^T$ as in Eq. (3.13)*.*

Proof Once expression (4.3) is given for the augmented Kalman filter, our result is obtained as in the standard case, see e.g. [27, Chapter 9]. □

4.2 Asymptotic Stability

We now particularize our situation to a linear context with constant coefficients. That is, we consider the following signal-observation system:

$$\begin{aligned} X_t &= X_0 + \int_0^t a X_s ds + W_t, \\ Z_t &= \int_0^t h X_s ds + N_t , \end{aligned} \tag{4.7}$$

where $a \in \mathbb{R}^{m\times m}$ and $h \in \mathbb{R}^{n\times m}$ are constant matrices. In order to state and prove our asymptotic stability result, we first need to recall some classical notions which can be found in [20, Theorem 4.11] or [27, Chapter 9].

Definition 4.2 In the following, A stands for a $m \times m$ matrix, while $D \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times l}$ for given integers l, m, n.
(i) We define *the stable subspace of matrix* A as the direct sum of the (right) kernels of $(\lambda_i I - A)^{m_i}$, where λ_i are negative eigenvalues of A and m_i is the multiplicity of

λ_i. We define *the unstable subspace of* A as the orthogonal of the stable subspace of A.
(ii) We call the couple of matrices (A, D) *detectable* if the (right) kernel of

$$\begin{bmatrix} D \\ DA \\ \vdots \\ DA^{m-1} \end{bmatrix}$$

is contained in the stable subspace of A.
(iii) We call the couple of matrices (A, B) *stabilizable* if the unstable subspace of A is contained in the linear space spanned by the columns of $(B, AB, \ldots, A^{m-1}B)$.

With the preliminary notions we have just introduced, we can state a result about existence of solutions for Riccati equations.

Lemma 4.3 *Let a and h be the coefficients given in* (4.7). *We assume that* (a, h) *is detectable and* (a, I) *is stabilizable. Then*
(i) *The algebraic Riccati equation*

$$\gamma_\infty a^T + a\gamma_\infty + I - \gamma_\infty h^T h\gamma_\infty = 0 \tag{4.8}$$

admits a unique solution γ_∞ in $\mathbb{R}^{m\times m}$.
(ii) *We have*

$$\lambda_0 \equiv \inf\{-Re\lambda : \lambda \textit{ is an eigenvalue of the matrix } a - \gamma_\infty h^T h\} > 0. \tag{4.9}$$

In the classical situation (i.e. for X_0 independent of N) and for a system like (4.7), it is well-known that the covariance of the optimal filter $\hat{X}_t = \mathbb{E}(X_t|\mathcal{F}_t^Z)$ converges exponentially fast to the solution of the algebraic Riccati equation (4.8); see [25]. The stability problem has been further studied in the recent works [3,4], where the stability of the optimal filter $\hat{X}_t$ with respect to (X_0, P) and the corresponding convergence rate is obtained.

Our aim now is to prove that this convergence still holds true, and recover the rate of convergence exhibited in [3,4], when the covariance between X_0 and N vanishes in finite time.

Theorem 4.4 *Consider the signal-observation system given by* (4.7) *under the same conditions as for Lemma* 4.3. *We also assume that there exists $T_0 > 0$ such that $\rho'(t) = 0$ for all $t > T_0$. Then*

(a) *Let P_t^{11} be the conditional variance of X_t given $\mathcal{F}_t^Z$. For all $\lambda < \lambda_0$, where λ_0 is defined by* (4.9)*, we have*

$$\lim_{t\to\infty} e^{\lambda t}(P_t^{11} - \gamma_\infty) = 0. \tag{4.10}$$

(b) *Let* $(\hat{X}_t^0)$ *be the optimal conditional expectation of* X_t *given the observation obtained when* X_0 *and* N *are independent (that is,* $\rho_N = 0$*). Then for all* $\lambda < \lambda_0$ *we have the convergence* $\lim_{t\to\infty} e^{\lambda t}(\hat{X}_t - \hat{X}_t^0) = 0$ *almost surely.*

(c) *Let* π_t *be the conditional Gaussian probability measure with mean* $\hat{X}_t$ *and covariance matrix* P_t*. In the same way, we define a conditional Gaussian measure* π_t^0 *given as* $\pi_t^0 = \mathcal{N}(\hat{X}_t^0, P_t^0)$*, where* $\hat{X}_t^0$ *is defined in item (b) above and* P_t^0 *is the conditional covariance of the usual Kalman filter(see e.g.* [27, Eq. (9.8)]*). Then we have*

$$\lim d_W(\pi_t, \pi_t^0) \to 0, \text{ almost surely,}$$

where d_W *denotes the Wasserstein metric in the space of probability measures.*

Proof It follows from Theorem 4.1 that the filter equations for the augmented version of Eq. (4.7) is:

$$\begin{aligned} d\hat{U}_t &= (b(t) - ck(t) - P_t k(t)^T k(t))\hat{U}_t dt + (c + P_t k(t)^T) dZ_t, &(4.11)\\ \dot{P}_t &= P_t(b(t) - ck(t))^T + (b(t) - ck(t))P_t + \sigma\sigma^T - P_t k(t)^T k(t) P_t, &(4.12) \end{aligned}$$

where U, b, σ and c are respectively defined by:

$$U = \begin{bmatrix} X \\ \bar{X} \\ N \end{bmatrix}, \quad b(t) = \begin{bmatrix} a & 0 & 0 \\ 0 & 0 & p(t) \\ 0 & g'(t) & r(t) \end{bmatrix}, \quad \sigma = \begin{bmatrix} I_m \\ 0 \\ 0 \end{bmatrix}, \quad c = \begin{bmatrix} 0 \\ 0 \\ I_n \end{bmatrix},$$

and $k(t)$ is the matrix given by

$$k(t) = [h \quad g'(t) \quad r(t)].$$

Observe that the following elementary identities hold true:

$$ck(t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ h & g'(t) & r(t) \end{bmatrix}, \quad b(t) - ck(t) = \begin{bmatrix} a & 0 & 0 \\ 0 & 0 & p(t) \\ -h & 0 & 0 \end{bmatrix}. \tag{4.13}$$

Moreover, recalling that P_t is a $\mathbb{R}^{(2m+n)\times(2m+n)}$ matrix, we decompose P_t into blocks of size $k \times l$ with $k, l \in \{m, n\}$ according to the 3 components of U. Hence, projecting Eq. (4.11) on the X component and recalling (4.13), it is readily checked that $\hat{X}_t$ satisfies

$$\begin{aligned} d\hat{X}_t &= a\hat{X}_t dt - (P^{11}h^T + P^{12}g'(t)^T + P^{13}r(t)^T)(h\hat{X}_t + g'(t)\hat{\bar{X}}_t + r(t)\hat{N})dt \\ &\quad + (P^{11}h^T + P^{12}g'(t)^T + P^{13}r(t)^T)dZ_t. \end{aligned} \tag{4.14}$$

In the same way, projecting relation (4.12) on the first component of U, we obtain that $P^{11} = \mathbb{E}((X_t - \hat{X}_t)(X_t - \hat{X}_t)^T)$ is

$$\begin{aligned} \dot{P}^{11} = P^{11}a^T + aP^{11} + I \\ -(P^{11}h^T + P^{12}g'(t)^T \\ +P^{13}r(t)^T)(hP^{11} + g'(t)P^{21} + r(t)P^{31}). \end{aligned} \tag{4.15}$$

Let us recall that the expression for g' is obtained in Corollary 2.3 and r is defined by (3.4). Therefore, since $\rho'(t) = 0$ for $t \geq T_0$, we easily get that for $t > T_0$ we also have $g'(t) = 0$ and $r(t) = 0$. Plugging this information into (4.14) and (4.15), the equations for $\hat{X}$ and P^{11} becomes:

$$\begin{aligned} d\hat{X}_t &= a\hat{X}_t dt - P_{11}h^T h\hat{X}_t dt + P_{11}h^T dZ_t\,, \\ \dot{P}^{11} &= P^{11}a^T + aP^{11} + I - P^{11}h^T hP^{11}. \end{aligned}$$

According to Lemma 4.3, if (a, h) is detectable and (a, I) is stabilizable, then Eq. (4.8) has a unique solution. Furthermore, it is shown in [20, Theorem 4.11] that under the same conditions we have

$$\lim_{t\to\infty} e^{\lambda t}(P_t^{11} - \gamma_\infty) = 0,$$

which is our claim (4.10). Observe also that Lemma 4.3 implies that the matrix $a - \gamma_\infty h^T h$ is asymptotically stable. Items (b) and (c) in our Theorem thus follow from the results in [25, Sect. 2] (see also [27, Sect. 9.5]). □

5 A Finite Filter

In this section, we consider another application of the methods used for the anticipative filter (1.1). Namely, inspired by e.g. [7,18], we wish to handle the case of a weighted Volterra type observation.

To be more specific, we are now considering a signal $(X_t)_{t\leq T}$ and an observation $(Z_t)_{t\leq T}$ governed by the stochastic differential equations

$$X_t = X_0 + \int_0^t a(s)X_s ds + W_t, \tag{5.1}$$

$$Z_t = \int_0^t H(t,s)X_s ds + N_s, \tag{5.2}$$

where (X_0, W_t, N_t) is a Gaussian family and the three terms are mutually independent. As in the previous sections, we assume that (W_t, N_t) is a standard Brownian motion, and $a : [0, T] \to \mathbb{R}^{m\times m}$, $H : [0, T]^2 \to \mathbb{R}$ are continuous functions. The observation information is given by the filtration of the observation process: $\mathcal{F}_t^Z = \sigma(Z_s; s \leq t)$, $t \in [0, T]$. The initial condition X_0 is assumed to be independent of N in this section. However, the fact that Z_t is governed by a Volterra type dynamics will force us to resort to the same augmented filtering equation as in the anticipative case. Observe that an anticipative initial condition in (5.1) could also be treated with our methods. We have refrained from going in this direction for sake of conciseness.

In order to ease our computations, we assume that the function H satisfies the following conditions:

Hypothesis 5.1 Let H be the kernel appearing in the definition (5.2) of Z. We assume the following holds true:

(i) H is a continuous function on $[0, T]^2$.
(ii) H admits the following decompositions:

$$H(t,s) = \sum_{i=1}^{\infty} p_i(t) q_i(s), \tag{5.3}$$

where p_i, q_i in (5.3) are such that $p_i \in C^1([0,T])$ and $q_i \in C([0,T])$, and where the convergence in (5.3) occurs in $L_1([0,T]^2)$.
(iii) For $n \geq 1$, set

$$H_n(t,s) = \sum_{i=1}^{n} p_i(t) q_i(s), \quad L_n(t,s) = \frac{d}{dt} H_n(t,s) = \sum_{i=1}^{n} p_i'(t) q_i(s). \tag{5.4}$$

Then L_n converges in $L_1([0,T]^2)$ to a continuous function $L(t,s)$.

Following is the main result of this section:

Theorem 5.2 *Consider the signal-observation Eq. (5.2). Suppose that H satisfies Hypothesis 5.1. For $0 \leq t \leq r$ we define the following augmented signal:*

$$V_{r,t} = \begin{bmatrix} X_t \\ \int_0^t L(r,s) X_s ds \end{bmatrix},$$

as well as the augmented coefficients

$$B_r(s) = \begin{bmatrix} a(s) \\ L(r,s) \end{bmatrix} [I_m \quad 0], \qquad \Sigma = \begin{bmatrix} I_m \\ 0 \end{bmatrix},$$

and an initial condition

$$P_0 = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}.$$

Then the conditional mean $\hat{V}_{r,t} = \mathbb{E}[V_{r,t} | \mathcal{F}_t^Z]$ satisfies the equation

$$\hat{V}_{r,t} = \hat{V}_{r,0} + \int_0^t \begin{bmatrix} a(s) \\ L(r,s) \end{bmatrix} \hat{X}_s ds + \int_0^t \mathcal{P}_{r,s} [H(s,s) \quad I]^T d\nu, \tag{5.5}$$

where $\nu_t = Z_t - \int_0^t [H(s,s) \quad I]\hat{V}_{r,s} ds$, *and where the conditional covariance* $\mathcal{P}_{r,t} = \mathbb{E}((V_{r,t} - \hat{V}_{r,t})(V_{t,t} - \hat{V}_{t,t})^T | \mathcal{F}_t^Z)$ *verifies the Riccati type equation*

$$\begin{aligned}\mathcal{P}_{r,t} - P_0 = \int_0^t \Big(& \mathcal{P}_{r,s}^T B_t(s)^T + B_r(s)\mathcal{P}_{r,s} + \Sigma\Sigma^T \\ & - \mathcal{P}_{r,s} [H(s,s) \quad I]^T [H(s,s) \quad I] \mathcal{P}_{t,s}^T \Big) ds. \end{aligned} \tag{5.6}$$

Proof We proceed according to the approximation given in Hypothesis 5.1. This will be done in two steps.

Step 1: High-dimensional augmented signal. Consider the signal X given by (5.1), as well as the following approximation of the observation:

$$\begin{aligned} dX_t &= a(t)X_t dt + dW_t, \\ Z_t^n &= \int_0^t H_n(t,s)X_s ds + N_s = \sum_{i=1}^n p_i(t)X_t^i + N_s, \end{aligned}$$

where we have set

$$X_t^i = \int_0^t q_i(s)X_s ds, \quad i = 1, \ldots, n.$$

Then an elementary product rule allows to write

$$Z_t^n = \int_0^t H_n(s,s)X_s ds + \sum_{i=1}^n \int_0^t p_i'(s)X_s^i ds + N_s. \tag{5.7}$$

We now consider an augmented signal and some augmented coefficients as follows:

$$\begin{aligned} \bar{U}^n &= \begin{bmatrix} X^1 \\ \vdots \\ X^n \end{bmatrix}, \quad \bar{b}^n(t) = \begin{bmatrix} q_1(t) \\ \vdots \\ q_n(t) \end{bmatrix}, \quad \sigma = \begin{bmatrix} I_m \\ 0 \end{bmatrix}, \\ \bar{h}^n(t) &= \left[p_1'(t) \cdots p_n'(t) \right], \end{aligned} \tag{5.8}$$

and we set

$$U^n = \begin{bmatrix} X \\ \bar{U}^n \end{bmatrix}, \quad b^n(t) = \begin{bmatrix} a(t) \\ \bar{b}^n(t) \end{bmatrix} \left[I_m \ 0 \right], \quad h^n(t) = \left[H_n(t,t) \ \bar{h}^n(t) \right]. \tag{5.9}$$

Then we obtain the following linear system for the observation U^n:

$$dU_t^n = b^n(t)U_t^n dt + \sigma dW_t. \tag{5.10}$$

In addition, it is easily seen that the process Z^n defined by (5.7) verifies

$$Z_t^n = \int_0^t h^n(s)U_s^n ds + N_t. \tag{5.11}$$

As in Theorem 4.1, Eqs. (5.10) and (5.11) can now be seen as a classical Kalman-Bucy filtering system. Hence we can invoke [27, Chapter 9] again, which yields the following equation for $\hat{U}_t^n = \mathbb{E}(U^n | \mathcal{F}_t^{Z^n})$:

$$\hat{U}_t^n = \hat{U}_0^n + \int_0^t b^n(s)\hat{U}_s^n ds + \int_0^t P_s^n h^n(s)^T d\nu^n, \tag{5.12}$$

where $\nu_t^n = Z_t^n - \int_0^t h^n(s)\hat{U}_s^n ds$ is the corresponding innovation process. As far as the covariance function $P_t^n = \mathbb{E}((U_t^n - \hat{U}_t^n)(U_t^n - \hat{U}_t^n)^T)$ is concerned, it satisfies the following Riccati equation:

$$P_t^n - P_0^n = \int_0^t \left(P_s^n b^n(s)^T + b^n(s)P_s^n + \sigma\sigma^T - P_s^n h^n(s)^T h^n(s) P_s^n \right) ds. \tag{5.13}$$

Step 2: Dimension reduction. In Step 1, the dimension of the augmented signal U^n grows with n. In order to go back to a low-dimensional signal, let us first compute the quantity $h^n(t)U_t^n$ in (5.11). Thanks to the definition (5.9) of h^n and U^n we have

$$\begin{aligned} h^n(t)U_t^n &= H_n(t,t)X_t + \bar{h}^n(t)\bar{U}_t^n = H_n(t,t)X_t + \sum_{i=1}^n p_i'(t) \int_0^t q_i(s)X_s ds \\ &= H_n(t,t)X_t + \int_0^t L_n(t,s)X_s ds, \end{aligned} \tag{5.14}$$

where the second equality is due to the definition (5.8) of $\bar{h}^n$ and the last equality stems from (5.4). Interestingly enough, Eq. (5.14) suggests to consider the filtering for the signal

$$R_t := \sum_{i=1}^\infty p_i'(t) \int_0^t q_i(s)X_s ds = \int_0^t L(t,s)X_s ds.$$

To this aim, we consider a new process $R_{r,t}^n$ defined for $0 \le t \le r$ by

$$R_{r,t}^n = \sum_{i=1}^n p_i'(r) \int_0^t q_i(s)X_s ds = \bar{h}^n(r)\bar{U}_t^n,$$

and we consider the following augmented signal (notice that the argument of $\bar{h}^n$ is frozen to r in the equation below; see Remark):

$$V_{r,t}^n := \begin{bmatrix} X_t \\ R_{r,t}^n \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & \bar{h}^n(r) \end{bmatrix} U_t^n. \tag{5.15}$$

We will now get the filtering equations for the augmented signal $V_{r,t}^n$. In order to get the equation for the conditional variance, we set

$$\mathcal{P}_{r,t}^n = \mathbb{E}((V_{r,t}^n - \hat{V}_{r,t}^n)(V_{t,t}^n - \hat{V}_{t,t}^n)^T | \mathcal{F}_t^{Z^n}) = \begin{bmatrix} I & 0 \\ 0 & \bar{h}^n(r) \end{bmatrix} P_t^n \begin{bmatrix} I & 0 \\ 0 & \bar{h}^n(t) \end{bmatrix}^T,$$

where the second relation is obtained thanks to the definition (5.15) of V^n and the fact that $P_t^n = \mathbb{E}[(U_t^n - \hat{U}_t^n)(U_t^n - \hat{U}_t^n)^T]$.

Hence multiplying relation (5.13) by $\begin{bmatrix} I & 0 \\ 0 & \bar{h}^n(r) \end{bmatrix}$ on the left and by $\begin{bmatrix} I & 0 \\ 0 & \bar{h}^n(t) \end{bmatrix}^T$ on the right, we get

$$\begin{aligned} &\mathcal{P}_{r,t}^n - P_0 \\ &= \int_0^t \Big((\mathcal{P}_{r,s}^n)^T B_t^n(s)^T + B_r^n(s)\mathcal{P}_{r,s}^n + \Sigma\Sigma^T \\ &\quad - \mathcal{P}_{r,s}^n \, [H_n(s,s) \quad I]^T \, [H_n(s,s) \quad I] \, (\mathcal{P}_{t,s}^n)^T \Big) \, ds, \end{aligned}$$

where the coefficients B^n and Σ are defined by

$$B_r^n(s) = \begin{bmatrix} a(s) \\ L_n(r,s) \end{bmatrix} [I_m \quad 0], \quad \Sigma = \begin{bmatrix} I_m \\ 0 \end{bmatrix}.$$

Sending $n \to \infty$ on both sides of the equation and applying Lemma 5.3 we obtain Eq. (5.6) for $\mathcal{P}_{r,t}$. The Eq. (5.5) for $\hat{V}_{r,t}$ can be derived in a similar way by multiplying (5.12) by the proper factor given by (5.15). This completes the proof. □

We now give some details about our auxiliary result needed in the proof of Theorem 5.2. Let us now consider the following auxiliary result:

Lemma 5.3 *Let (x, y), $(\tilde{x}, y)$, (x, y_n) and $(\tilde{x}, y_n)$ be joint Gaussian random vectors, and suppose that y_n converges to y in $L^2(\Omega)$. Then*

(i) $\mathbb{E}[x|y_n] \to \mathbb{E}[x|y]$ *in* $L^2(\Omega)$;
(ii) $\mathbb{E}[\mathbb{E}(x|y_n)\mathbb{E}(\tilde{x}|y_n)] \to \mathbb{E}[\mathbb{E}(x|y)\mathbb{E}(\tilde{x}|y)]$.

Proof Denote $(x, y) \sim N(\mu, \Sigma)$, $(\tilde{x}, y) \sim N(\tilde{\mu}, \tilde{\Sigma})$, $(x, y_n) \sim N(\mu^n, \Sigma^n)$, $(\tilde{x}, y) \sim N(\tilde{\mu}^n, \tilde{\Sigma}^n)$. Since y_n converges to y in $L^2(\Omega)$, we have as $n \to +\infty$,

$$\mathbb{E}(y_n) \to \mathbb{E}(y) \quad \text{and} \quad \mathbb{E}\left\{ \begin{pmatrix} x \\ y_n \end{pmatrix} \begin{pmatrix} x & y_n \end{pmatrix} \right\} \to \mathbb{E}\left\{ \begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} x & y \end{pmatrix} \right\},$$

which implies $(\mu^n, \Sigma^n) \to (\mu, \Sigma)$. On the other hand, writing

$$\mathbb{E}\left[\begin{pmatrix} x \\ y \end{pmatrix}\right] = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \text{and} \quad \mathrm{Cov}\left[\begin{pmatrix} x \\ y \end{pmatrix}\right] = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

it is well known that

$$\mathbb{E}(x|y) = \mathbb{E}[x] + \Sigma_{12}\Sigma_{22}^{-1}(y - \mathbb{E}[y]).$$

Taking into account the above two points, we have (i) and (ii). □

6 Application to Radar Tracking

In this section we will apply the anticipative linear filter described in Sect. 4 to a standard practical problem considered in the literature. Specifically, we consider an anticipative version of the radar tracking system given in [14, Chapter 5]. We shall observe how the algorithm induced by Theorem 4.1 improves the estimation, versus a method using the classical Kalman filter and ignoring the anticipative problem.

In the radar tracking situation taken from [14] the signal X is governed by Eq. (4.1), where $a(s)$ is a constant matrix. Namely, $X_t = [\, r_t\ \dot{r}_t\ u_t^1\ \theta_t\ \dot{\theta}_t\ u_t^2\,]$ is a 6-dimensional process, where (r_t, θ_t) describes the position of the tracked vehicle expressed in polar coordinates in $\mathbb{R}^2$ (r is called range and θ is called bearing in [14]). The coordinates (u_t^1, u_t^2) also stand for a maneuvering-correlated state noise, while $\dot{r}_t$ and $\dot{\theta}_t$ respectively represent the time derivatives for the range and the bearing.

In [14] the dynamics for the process X is supposed to be governed by the following equation:

$$dX_t = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \kappa - 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & \kappa - 1 \end{bmatrix} X_t dt + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \sigma_1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \sigma_2 \end{bmatrix} dW_t, \tag{6.1}$$

where κ is a real valued constant, $\sigma_1, \sigma_2 > 0$ and W is a 2-dimensional Brownian motion. Equation (6.1) can be interpreted in the following way: we write that $r_t = r_0 + \int_0^t \dot{r}_s ds$, where the velocity $\dot{r}_t$ is equal to u_t^1 and u_t^1 is an Ornstein-Uhlenbeck process driven by W_t^1. Similar assumptions are also in order for the bearing θ.

As far as the observation process is concerned, we write Eq. (4.2) under the following form:

$$dZ_t = \begin{bmatrix} \sigma_\theta^{-1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_\theta^{-1} & 0 & 0 \end{bmatrix} X_t dt + dN_t, \tag{6.2}$$

where σ_θ is a positive constant and N is a 2-dimensional Brownian motion independent of W. Note that according to (6.2), Z_t^1 (resp. Z_t^2) is a linear function of r_t (resp. θ_t) plus a noisy perturbation:

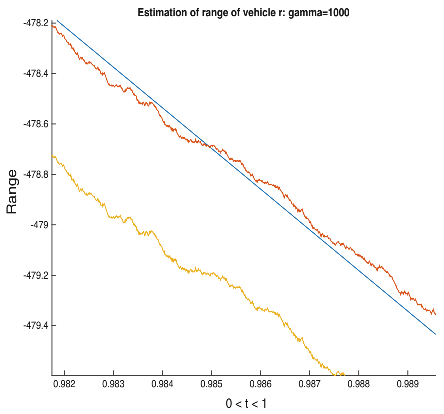
$$dZ_t^1 = \sigma_\theta^{-1} r_t dt + dN_t^1, \quad \text{and} \quad dZ_t^2 = \sigma_\theta^{-1} \theta_t dt + dN_t^2.$$

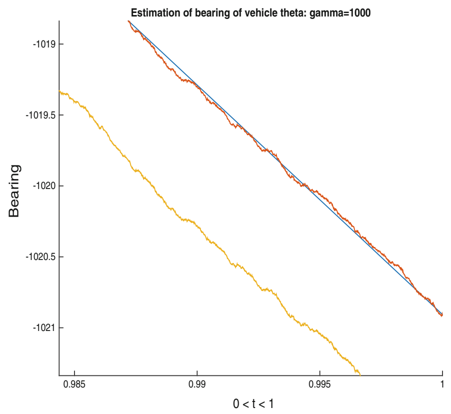
The anticipative nature of our system is enclosed in the following assumption: We assume that the initial location X_0 is influenced by the temperature, wind conditions and other environmental variables, while at the same time these factors have an impact on the observation. More precisely, we assume that

$$X_0 := \begin{bmatrix} r_0 \\ \dot{r}_0 \\ u_0^1 \\ \theta_0 \\ \dot{\theta}_0 \\ u_0^2 \end{bmatrix} = \xi + \eta, \quad \text{where} \quad \eta = \gamma \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} N_1 = \gamma \begin{bmatrix} N_1^1 \\ 0 \\ N_1^1 \\ N_1^2 \\ 0 \\ N_1^2 \end{bmatrix}, \tag{6.3}$$

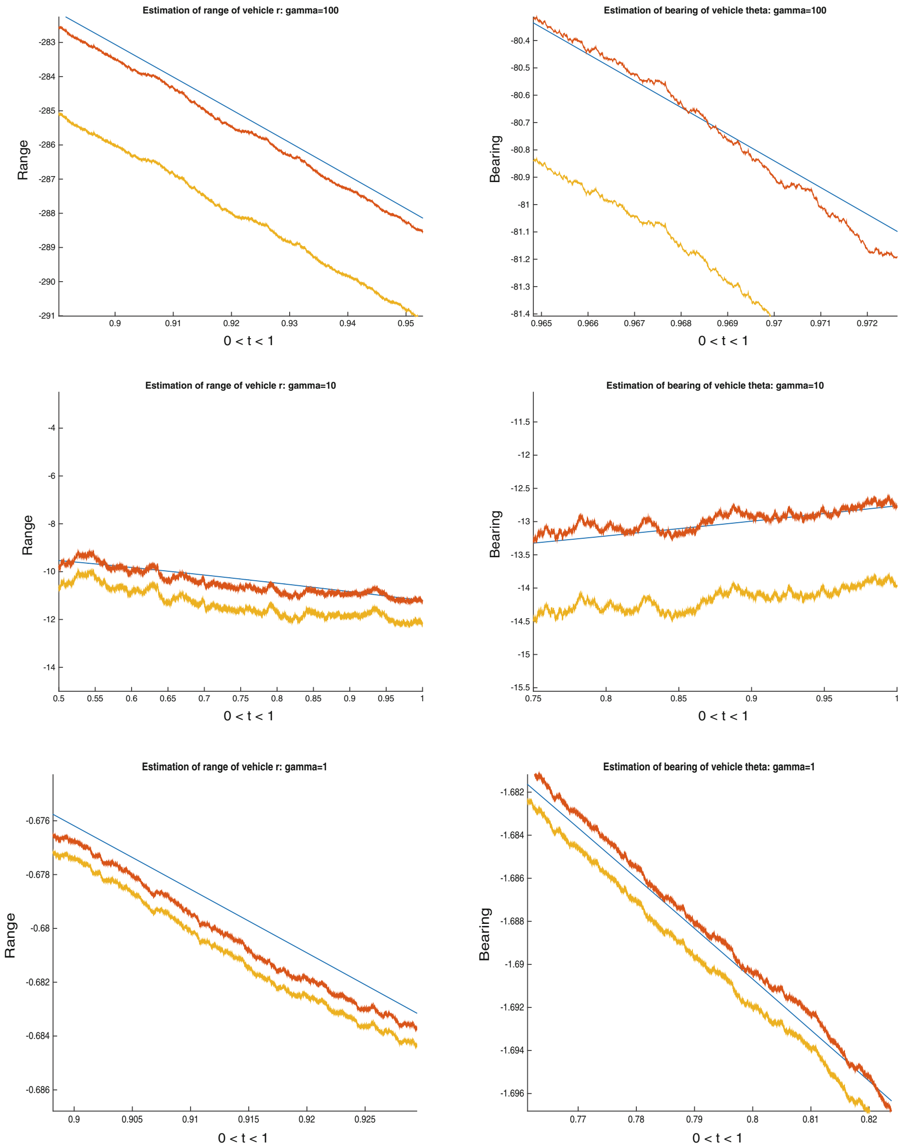
where we recall that N_1^j stands for the j-th component of N evaluated at $t = 1$. In Eq. (6.3) the vector ξ is a standard $\mathbb{R}^6$-valued Gaussian random variable independent of (W, N), and γ is a positive constant measuring the anticipation strength. Notice that according to Eq. (6.3) the anticipation of X_0 is only felt on the components r_0, θ_0 and u_0 of X_0. Moreover in (6.3) we assume that r_0 and u_0^1 depend on N_1^1, while θ_0 and u_0^2 depend on N_1^2, which is natural in our context. For our numerical simulations we take $\kappa = 0.5$, $\sigma_\theta = 0.017$ rad, $\sigma_1 = 103/3$ and $\sigma_2 = 1.3$.

As the reader might expect, our new filter (4.5)–(4.6) provides a much better estimation for the anticipative signal-observation system (6.1)–(6.2). This is attested by the following simulation. Namely, in the figures below the blue curve represents the signal path, the yellow curve is drawn according to the classical Kalman filter, and the orange curve is drawn thanks to our new filter. We have zoomed in the picture for comparison purposes, so that the signal curve appears to be linear.





In this set of figures we successively take $\gamma = 1000, 100, 10, 1$, in order to observe the effect of the anticipation strength on the filter performance.



Remark 6.1 As our simulation shows, the improvement of the new filter from the classical Kalman filter becomes more significant as the anticipation gets stronger.

Denote by $\hat{X}_t^i$, $i = 1, \ldots, 6$ the optimal filter, i.e. $\hat{X}_t^i = \mathbb{E}(X_t^i | \mathcal{F}_t^Z)$, and denote by $\bar{X}_t^i$ the estimate obtained while the anticipation is ignored. The following tables present the ratio R_i between the error deviations of these two estimates at time t, that is,

$$R_i(t) = \left(\frac{\mathbb{E}\left[|X_t^i - \hat{X}_t^i|^2\right]}{\mathbb{E}\left[|X_t^i - \bar{X}_t^i|^2\right]} \right)^{1/2}$$

For $t = 1$, we consider the anticipation strength $\gamma = 1, 10, 100$.

t=1	R_1	R_2	R_3	R_4	R_5	R_6
$\gamma = 1$	0.0100	0.0361	0.0608	0.0141	0.0400	0.0616
$\gamma = 10$	0.0100	0.0265	0.0574	0.0100	0.0316	0.0574
$\gamma = 100$	0.0100	0.0265	0.0574	0.0100	0.0316	0.0574

For $t = 3/4$, we consider the anticipation strength $\gamma = 1, 10, 100, 1000$.

t=3/4	R_1	R_2	R_3	R_4	R_5	R_6
$\gamma = 1$	0.3670	0.3684	0.3688	0.3670	0.3686	0.3689
$\gamma = 10$	0.4603	0.4609	0.4615	0.4574	0.4610	0.4615
$\gamma = 100$	0.4965	0.4969	0.4981	0.4953	0.4970	0.4981
$\gamma = 1000$	0.5007	0.5011	0.5023	0.4997	0.5012	0.5023

As the reader can see, our ratios are small regardless of the values of t and γ. This indicates that our filter performs well compared with a filter ignoring the anticipative nature of the signal.

We end this section with a remark on the stability of our new filter (4.5) and (4.6):

Remark 6.2 In our numerical experiments, we find that the stability of the original system (4.1) and (4.2) and the new system (4.3) can be quite different. As η gets smaller the stability of the new system usually decreases, and therefore finer mesh is needed in the simulations in order to capture the accuracy improvement achieved by our new filter.

Acknowledgements Y. Liu wishes to thank Professor Yaozhong Hu and Professor David Nualart for helpful discussions. G. Lin gratefully acknowledges the support from National Science Foundation (DMS-1555072, DMS-1736364, and DMS-1821233). S. Tindel gratefully acknowledges the support from National Science Foundation DMS-1613163.

Compliance with Ethical Standards

Conflicts of interest The authors declare that they have no conflict of interest.

References

1. Aase, K., Bjuland, T., Øksendal, B.: An anticipative linear filtering equation. Syst. Control Lett. **60**(7), 468–471 (2011)

2. Back, K.: Insider trading in continuous time. Rev. Financ. Stud. **5**, 387–409 (1992)
3. Bishop, A.N., Del Moral, P.: On the stability of Kalman-Bucy diffusion processes. SIAM J. Control Optim. **55**(6), 4015–4047 (2017)
4. Bishop, A.N., Del Moral, P.: An explicit Floquet-type representation of Riccati aperiodic exponential semigroups. Int. J. Control. (2019)
5. Cont, R., Fournié, D.A.: Functional Itô calculus and stochastic integral representation of martingales. Ann. Probab. **41**(1), 109–133 (2013)
6. Corcuera, J., Imkeller, P., Kohatsu-Higa, A., Nualart, D.: Additional utility of insiders with imperfect dynamical information. Financ. Stoch. **8**(3), 437–450 (2004)
7. Coutin, L., Decreusefond, L.: Abstract nonlinear filtering theory in the presence of fractional Brownian motion. Ann. App. Probab. **9**(4), 1058–1090 (1999)
8. Crisan, D., Rozovskii, B. (eds.): The Oxford Handbook of Nonlinear Filtering. Oxford University Press, Oxford (2011)
9. Davis, M.H.A.: Linear Estimation and Stochastic Control. Halsted Press, New York (1977)
10. Del Moral, P.: Non-linear filtering using random particles. Theory Probab. Appl. **40**, 690–701 (1995)
11. Del Moral, P., Miclo, L.: Genealogies and increasing propagations of chaos for Feynman-Kac and genetic models. Ann. App. Probab. **11**(4), 1166–1198 (2001)
12. Del Moral, P., Miclo, L.: Strong propagations of chaos in Moran's type particle interpretations of Feynman-Kac measures. Stoch. Anal. Appl. **25**(3), 519–575 (2007)
13. Florchinger, P., Le Gland, F.: Time-discretization of the Zakai equation for diffusion processes observed in correlated noise. Stoch. Stoch. Rep. **35**, 233–256 (1991)
14. Grewal, M., Andrews, A.: Kalman Filtering. Theory and Practice Using MATLAB, 4th edn. Wiley, Hoboken (2015)
15. Hu, Y.: An enlargement of filtration for Brownian motion. Acta Math. Sci. Ser. B (Engl. Ed.) **31**(5), 1671–1678 (2011)
16. Jazaerli, S., Saporito, Y.F.: Functional Itô calculus, path-dependence and the computation of Greeks. Stoch. Process. Appl. **127**(12), 3997–4028 (2017)
17. Kallianpur, G.: Stochastic Filtering Theory. Springer, New York (1980)
18. Kleptsyna, M., Le Breton, A., Roubaud, M.: General approach to filtering with fractional Brownian noises—application to linear systems. Stochastics **71**(1–2), 119–140 (2000)
19. Kurtz, T., Xiong, J.: Particle representations for a class of nonlinear SPDEs. Stoch. Process. Appl. **83**, 103–126 (1999)
20. Kwakernaak, H., Sivan, R.: Linear Optimal Control Systems. Wiley, New York (1972)
21. Kyle, A.S.: Continuous auctions and insider trading. Econometrica **53**, 1315–1335 (1985)
22. Liptser, R., Shiryaev, A.: Statistics of Random Processes. I–II. Springer-Verlag, Berlin (2001)
23. Lototsky, S.V.: Wiener chaos and nonlinear filtering. Appl. Math. Optim. **54**(3), 265–291 (2006)
24. Norros, I., Valkeila, E., Virtamo, J.: An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions. Bernoulli **5**(4), 571–587 (1999)
25. Ocone, D., Pardoux, E.: Asymptotic stability of the optimal filter with respect to its initial condition. SIAM J. Control Optim. **34**(1), 226–243 (1996)
26. Saporito, Y.F.: Topics on Functional Itô Calculus and Multiscale Stochastic Volatility Modeling. PhD. thesis, Santa Barbara (2014)
27. Xiong, J.: An Introduction to Stochastic Filtering Theory. Oxford University Press Inc., New York (2008)
28. Xiong, J., Zhao, X.: Nonlinear filtering with fractional Brownian motion noise. Stoch. Anal. Appl. **23**(1), 55–67 (2005)
29. Zakai, M.: On the optimal filtering of diffusion processes. Z. Wahrsch. Verw. Gebiete **11**, 230–243 (1969)