Vector-valued generalised Ornstein-Uhlenbeck processes: properties and parameter estimation

Marko Voutilainen[1,2] | Lauri Viitasaari[3] | Pauliina Ilmonen[2] | Soledad Torres[4] | Ciprian Tudor[5]

[1]Department of Accounting and Finance, Turku School of Economics, Finland
[2]Department of Mathematics and Systems Analysis, Aalto University School of Science, Finland
[3]Department of Information and Service Management, Aalto University School of Business, Finland
[4]CIMFAV - Facultad de Ingeniería, Universidad de Valparaíso, Chile
[5]UFR Mathématiques, Université de Lille 1, France

Correspondence
Marko Voutilainen, Turku School of Economics, Department of Accounting and Finance, FI-20014 University of Turku, Finland
Email: mtvout@utu.fi

Funding information
Väisälä foundation; Magnus Ehrnrooth foundation

Generalisations of the Ornstein-Uhlenbeck process defined through Langevin equations, such as fractional Ornstein-Uhlenbeck processes, have recently received a lot of attention. However, most of the literature focuses on the one-dimensional case with Gaussian noise. In particular, estimation of the unknown parameter is widely studied under Gaussian stationary increment noise. In this article, we consider estimation of the unknown model parameter in the multidimensional version of the Langevin equation, where the parameter is a matrix and the noise is a general, not necessarily Gaussian, vector-valued process with stationary increments. Based on algebraic Riccati equations, we construct an estimator for the parameter matrix. Moreover, we prove the consistency of the estimator and derive its limiting distribution under natural assumptions. In addition, to motivate our work, we prove that the Langevin equation characterises essentially all multidimensional stationary processes.

KEYWORDS
algebraic Riccati equations, consistency, Langevin equation, multivariate Ornstein-Uhlenbeck process, nonparametric estimation, stationary processes

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/sjos.12552

1 | INTRODUCTION

In this article, we study statistical problems related to the multidimensional version of generalised Langevin equation

$$dU_t = -\Theta U_t \, dt + dG_t, \qquad t \in \mathbb{R}, \tag{1}$$

with some stationary increment noise G. Here Θ is a positive definite matrix, and the noise G and the solution U are understood as random vectors.

A classical Ornstein-Uhlenbeck process can be defined through the Langevin equation

$$dU_t = -\theta U_t \, dt + dW_t, \qquad t \in \mathbb{R}, \tag{2}$$

where $\theta > 0$ is a parameter and W is a Brownian motion. The stationary Ornstein-Uhlenbeck process U can be obtained by a suitable choice of the initial condition U_0. Such equations have connections especially to physics, and this is part of the reason why Langevin equations of the form (1) have received a lot of attention in the literature. In addition to the connections to physics, it was recently proven in Viitasaari (2016) that, in one dimensional case, Langevin equations characterise essentially all stationary processes (for analogous results in discrete time, we refer to Voutilainen et al. (2017, 2019)). This highlights the importance of (1) even further.

Equation (1) is well-motivated, and there is a vast array of literature studying it with varying driving force G. One natural generalisation is to replace the Brownian motion with a Lévy process. In the infinite-dimensional case, Equation (2) driven by a Lévy-process has connections to different branches of probability theory such as stochastic partial differential equations, branching processes, generalised Mehler semigroups, and self-decomposable distributions. For a recent survey on the topic, we refer to Applebaum (2015). In addition, Ornstein-Uhlenbeck processes have been generalised in connection with random recurrence equations. On the related work, we mention Behme and Lindner (2012) (and references therein), where the authors characterise multivariate generalised Ornstein-Uhlenbeck processes through (2) by replacing $(\theta t, W_t)$ with some other appropriate multivariate Lévy processes.

Another natural generalisation is to replace the Brownian motion W in (2) with a fractional Brownian motion B^H.

(For details on fractional Brownian motion, we refer e.g. to Mishura (2008).) The solution U, called the *fractional Ornstein-Uhlenbeck process*, was first introduced by Cheridito et al. (2003) (see also Kaarakka and Salminen (2011)). Statistical analysis related to the fractional Ornstein-Uhlenbeck model was initiated in Hu and Nualart (2010) and Kleptsyna and Breton (2002), and it has been a very active research topic ever since. Of the studies on parameter estimation in such models, we mention Azmoodeh and Viitasaari (2015); Bajja et al. (2017); Balde et al. (2018); Brouste and Iacus (2013); Dehling et al. (2017); Douissi et al. (2020); Es-Sebaiy and Ndiaye (2014); Es-Sebaiy and Tudor (2015); Hu et al. (2019); Kozachenko et al. (2015); Kubilius et al. (2015); Maslowski and Pospíšil (2008); Shen et al. (2016); Shen and Xu (2014); Sottinen and Viitasaari (2018); Sun and Guo (2015); Tanaka (2015) to name a few recent ones. Finally, we mention Magdziarz (2008), that considers fractional extensions of the Levy-driven Ornstein-Uhlenbeck processes.

Despite the vast amount of literature related to (1), to the best of our knowledge most of it focuses on one-dimensional case, and with a specific driver G. In particular, even if the problem is studied in a greater generality to some extent, usually the assumptions are somehow related to the one-dimensional case, or to a specific driver. For example, in Balde et al. (2018); Maslowski and Pospíšil (2008) the authors studied infinite dimensional fractional Ornstein-Uhlenbeck processes, but there was only one unknown parameter $\Theta \in \mathbb{R}$ to estimate and the driver was of a specific form. Similarly, for example in Bajja et al. (2017); Dehling et al. (2017) there were many parameters to estimate, but again the driver (and the model) was of a specific form. Finally, while in Sottinen and Viitasaari (2018) the authors studied a more general noise G that is not related to the fractional Brownian motion, G was still assumed to be Gaussian and the equation was considered only in one dimension with one parameter to estimate. Similarly, in Douissi et al. (2020) and Nourdin and Tran (2019), the authors considered non-Gaussian case, but only with one parameter and a specific, though non-Gaussian, driver.

The aim of this article is to study general multidimensional Langevin equations with arbitrary stationary increment noise. That is, we study (1) with an unknown positive definite matrix Θ. We prove that (1) characterises (essentially) all stationary processes, thus giving a natural multidimensional extension of the results presented in Viitasaari (2016). Moreover, given that the underlying processes are square integrable, we provide representations of the cross-covariance matrix $\gamma(t)$ of the stationary solution U_t, and show that the unknown Θ solves a certain continuous-time algebraic Riccati equation (CARE). Initiated by the seminal paper Kalman (1960), CAREs arise naturally in optimal control and filtering theory. As such, we relate the Langevin equation to these fields as well.

We also consider statistical estimation of the unknown matrix Θ. Motivated by the relation to CARE, we define the estimator as the solution to a perturbed CARE, in which the coefficient matrices are replaced by estimated ones, and where the cross-covariances $\gamma(t)$ are replaced by their estimators $\hat{\gamma}(t)$. We prove that our estimator is consistent whenever the cross-covariance estimators are consistent. We also study how the rate of convergence and the limiting distribution of our estimator are related to the convergence rate and the limiting distribution of $\hat{\gamma}(t)$.

The rest of the paper is organised as follows. In Section 2 we present and discuss our main results. In particular, we state the characterisation of stationary processes through (1) and we provide the connection to CARE. We also define our estimator for the unknown Θ and provide results on its asymptotic properties. In Subsection 2.2 we illustrate the applicability of our results to the Gaussian case. All the proofs are postponed to Appendix A.

2 | MULTIDIMENSIONAL GENERALISED LANGEVIN EQUATIONS

We consider the n-dimensional Langevin equation

$$dU_t = -\Theta U_t dt + dG_t, \qquad t \in \mathbb{R} \tag{3}$$

driven by $G = (G_t)_{t\in\mathbb{R}}$, with a positive definite coefficient matrix Θ. Note that G and the solution U are n-dimensional vector-valued processes. Here we understand the solution in the strong sense, with a given initial condition to be specified later. The components of the vectors are denoted by super indices, e.g. $U_t^{(k)}$ denotes the kth component of the vector U_t, and is a real-valued random process. We denote by S^n the set of symmetric $n \times n$-matrices, and with the notation $\Theta > 0$ we mean that the matrix Θ is positive definite. Similarly, by writing $\Theta \geq 0$ we mean that Θ is positive semidefinite. If not stated otherwise, we use $\|\cdot\|$ to denote the standard L^2 norm and the corresponding induced matrix norm. If two processes $(X_t)_{t\in\mathbb{R}}$ and $(Y_t)_{t\in\mathbb{R}}$ have equal finite dimensional distributions, we write $(X_t)_{t\in\mathbb{R}} \overset{law}{=} (Y_t)_{t\in\mathbb{R}}$. Furthermore, with stationary processes we mean n-dimensional strictly stationary processes, i.e. processes for which $(X_{t+s})_{t\in\mathbb{R}} \overset{law}{=} (X_t)_{t\in\mathbb{R}}$ for every $s \in \mathbb{R}$. In addition, we assume that the driver G (and consequently, the solution U), have continuous paths almost surely. This guarantees that integrals of type

$$\int_s^t e^{\Theta u} dX_u$$

can be understood componentwise as pathwise Riemann-Stieltjes integrals via integration by parts

$$\int_s^t e^{\Theta u} dX_u = e^{\Theta t} X_t - e^{\Theta s} X_s - \Theta \int_s^t e^{\Theta u} X_u du.$$

Indefinite integrals over an interval $[-\infty, t]$ are defined similarly as

$$\int_{-\infty}^t e^{\Theta u} dX_u = e^{\Theta t} X_t - \Theta \lim_{s \to -\infty} \int_s^t e^{\Theta u} X_u du \tag{4}$$

provided that the limit exists almost surely.

Our first main theorem below shows that the characterisation of stationary processes through (3) in dimension one, provided in Viitasaari (2016), can be generalised naturally to the multidimensional setting, and motivates the statistical analysis of Equation (3). For this, we present the following definition for the class $\mathcal{G}_\Theta$ of possible drivers G.

Definition 1 *Let $\Theta > 0$ be fixed. Let $G = (G_t)_{t \in \mathbb{R}}$ be an n-dimensional stochastic process with stationary increments and $G_0 = 0$. We denote $G \in \mathcal{G}_\Theta$ if*

$$\lim_{u \to \infty} \int_{-u}^0 e^{\Theta s} dG_s$$

defines an almost surely finite random vector.

Remark 1 *In the one-dimensional setting, existence of certain logarithmic moments are sufficient to ensure $G \in \mathcal{G}_\Theta$. This result can be extended to the multidimensional setting in a straightforward manner. Consequently, our estimation procedure, that does rely on the existence of the second moments, always guarantees $G \in \mathcal{G}_\Theta$.*

Theorem 1 *Let $\Theta > 0$ be fixed. A continuous time n-dimensional stochastic process $U = (U_t)_{t \in \mathbb{R}}$ is stationary if and only* *if it is the unique solution of the Langevin equation* (3) *for some* $G \in \mathcal{G}_\Theta$ *and the initial value*

$$U_0 = \int_{-\infty}^{0} e^{\Theta s} dG_s. \tag{5}$$

That is

$$U_t = e^{-\Theta t} \int_{-\infty}^{t} e^{\Theta s} dG_s. \tag{6}$$

Moreover, the process G *is unique.*

Motivated by this result, let us now turn our attention to the statistical analysis of (3). That is, we suppose that the solution U is observed, and our aim is to define an estimator for the unknown parameter Θ. Our approach is based on utilizing the cross-covariance matrices, and for this reason we require some moment assumptions. In the sequel, we assume that the components $G^{(i)}$ of the driver G satisfy, for all $i = 1, \ldots, n$,

$$\sup_{s \in [0,1]} \mathbb{E}\left[G_s^{(i)}\right]^2 < \infty. \tag{7}$$

This assumption ensures that G is square-integrable, and consequently so is the solution U. We remark that this assumption ensures that $G \in \mathcal{G}_\Theta$ (cf. Lemma 16). Also, without loss of generality, we assume that G is centred, i.e. $\mathbb{E}(G_t) = 0$ for every $t \in \mathbb{R}$. This gives that also $\mathbb{E}(U_t) = 0$ for every $t \in \mathbb{R}$.

Let us now introduce some notation. With $\gamma(t)$ we denote the cross-covariance matrix of the stationary solution U, namely

$$\gamma(t) = \mathbb{E}(U_t U_0^\top) = \begin{bmatrix} \mathbb{E}(U_t^{(1)} U_0^{(1)}) & \mathbb{E}(U_t^{(1)} U_0^{(2)}) & \dots & \mathbb{E}(U_t^{(1)} U_0^{(n)}) \\ \mathbb{E}(U_t^{(2)} U_0^{(1)}) & \mathbb{E}(U_t^{(2)} U_0^{(2)}) & \dots & \mathbb{E}(U_t^{(2)} U_0^{(n)}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}(U_t^{(n)} U_0^{(1)}) & \mathbb{E}(U_t^{(n)} U_0^{(2)}) & \dots & \mathbb{E}(U_t^{(n)} U_0^{(n)}) \end{bmatrix}. \tag{8}$$

Notice that $\gamma(-t) = \gamma(t)^\top$. In addition, we denote a single element $\mathbb{E}(U_t^{(i)} U_0^{(j)})$ by $\gamma_{i,j}(t)$. We also define the following matrix coefficients for every $t \geq 0$.

$$B_t = \int_0^t \left(\gamma(s) - \gamma(s)^\top\right) ds \tag{9}$$

$$C_t = \int_0^t \int_0^t \gamma(s-u) du ds \tag{10}$$

$$D_t = \mathrm{cov}(G_t) - \mathrm{cov}(U_t - U_0). \tag{11}$$

Remark 2 *The cross-covariance $\gamma(t)$ can be computed explicitly from* (6). *For representations in the case when G has independent components, see Lemma 21 and Lemma 22.*

With the help of the above notation, we are able to write the parameter matrix Θ as a solution to the so-called continuous-time algebraic Riccati equation (CARE), with matrices B_t, C_t, and D_t as coefficients. This will lead to a natural estimator for Θ.

Theorem 2 *Let U be the solution of the Langevin equation* (3) *with $\Theta > 0$ and initial* (5). *Then, for every $t \geq 0$, the CARE*

$$B_t^\top \Theta + \Theta B_t - \Theta C_t \Theta + D_t = 0 \tag{12}$$

is satisfied.

Remark 3 *In the one-dimensional case $B_t \equiv 0$. After a change of variable* (12) *transforms into*

$$2\Theta^2 \int_0^t \gamma(z)(t-z)dz = v(t) + 2\gamma(t) - 2\gamma(0),$$

where $v(t)$ is the variance function of G. From this, we can compute a solution $\Theta > 0$ easily whenever $\int_0^t \gamma(z)(t-z)dz \neq 0$ and $v(t)$ is known. More generally, the coefficients B_t, C_t, and D_t can be computed from the observed process U if one value of the covariance matrix function $t \mapsto \operatorname{cov}(G_t)$ of the noise is known. In the literature, it is a typical assumption that the variance function of the noise is known completely (up to scaling).

From practical point of view, it is desirable that (12) admits a unique positive definite solution. Indeed, then the solution is automatically the correct parameter matrix Θ. Moreover, uniqueness of the solution is also a wanted feature for numerical methods. If the coefficient matrices C_t and D_t are both positive definite, then the solution is unique (in the set of positive semidefinite matrices). In our model, it turns out that this is usually the case if one chooses t appropriately. A detailed discussion on the matter is postponed to Subsection 2.1 (See also Remark 6 below on how t can be chosen in practice).

Remark 4 *Even if the solution is unique,* (12) *is rarely solvable in a closed form. Thus, in practice or for simulations, one has to apply some numerical method. On the other hand, even in the one-dimensional general Gaussian setup one may need to rely on numerical approximations. For example, the ergodicity estimator studied in Sottinen and Viitasaari (2018) is based on a function ψ^{-1} that can be computed explicitly only in some particular cases. Actually, applying our method to the one-dimensional case we observe a closed form expression for the solution (cf. Remark 3). For numerical methods associated to* (12)*, see e.g. Byers (1987); Laub (1979) and the monograph Bini et al. (2011).*

In the sequel, we assume that t is chosen such that $C_t, D_t > 0$, guaranteeing that Θ is the unique solution to (12). For notational simplicity, we will omit the subindex t and simply write

$$B^\top\Theta + \Theta B - \Theta C\Theta + D = 0 \tag{13}$$

whenever confusion cannot arise.

Suppose now that we have an observation window $[0, T]$. We define estimators $\hat{B}_T$, $\hat{C}_T$, and $\hat{D}_T$ for the coefficient matrices by replacing $\gamma_{i,j}(s)$ with any cross-covariance estimator $\hat{\gamma}_{T,i,j}(s)$ in the defining equations (9)-(11) (see Section 2.2 for an example of covariance estimator). We also write $\Delta_T B = \hat{B}_T - B$, $\Delta_T C = \hat{C}_T - C$, and $\Delta_T D = \hat{D}_T - D$. This leads to a perturbed CARE that gives us an estimator for Θ.

Definition 2 *The estimator $\hat{\Theta}_T$ is defined as the positive semidefinite solution to the perturbed CARE*

$$\hat{B}_T^\top \hat{\Theta}_T + \hat{\Theta}_T \hat{B}_T - \hat{\Theta}_T \hat{C}_T \hat{\Theta}_T + \hat{D}_T = 0 \tag{14}$$

whenever there exists a unique solution in the class of positive semidefinite matrices. If the solution does not exists, we set $\hat{\Theta}_T = 0$.

The idea of our estimator is that if the estimators $\hat{\gamma}_{T,i,j}(s)$ are consistent and $C, D > 0$ in the original CARE (12), then the perturbed version (14) automatically has a unique solution $\hat{\Theta}_T$ (with probability increasing to one as T grows), that converges strongly to Θ.

Theorem 3 *Suppose $C, D > 0$ and assume that*

$$\sup_{s\in[0,t]} \|\hat{\gamma}_T(s) - \gamma(s)\| \xrightarrow{\mathbb{P}} 0. \tag{15}$$

Then for $\hat{\Theta}_T$, given by Definition 2, we have

$$\|\hat{\Theta}_T - \Theta\| \xrightarrow{\mathbb{P}} 0. \tag{16}$$

Remark 5 *If the convergence in (15) holds almost surely, we obtain a strong consistent estimator, i.e.*

$$\|\hat{\Theta}_T - \Theta\| \xrightarrow{a.s.} 0.$$

Moreover, by our proof (cf. Lemma 18) we obtain that instead of (15)*, weaker conditions*

$$\int_0^t \|\hat{\gamma}_T(s) - \gamma(s)\| ds \xrightarrow{\mathbb{P}} 0$$

and

$$\|\hat{\gamma}_T(\tau) - \gamma(\tau)\| \xrightarrow{\mathbb{P}} 0, \tau \in \{0, t\}$$

are sufficient. These conditions are usually easier to verify in practice.

Remark 6 *In practice, one does not know the underlying exact model, and thus one cannot determine whether for given t we have $C_t, D_t > 0$. However, one can always pre-check whether, for a given t, matrices $\hat{C}_T, \hat{D}_T$ that are computed from the observations are positive definite. This together with* (15) *indicates $C_t, D_t > 0$ implying that the original CARE* (13) *has a unique positive semidefinite solution Θ (cf. Theorem 17). Now Theorem 3 applies, and consequently one can estimate Θ from the observations by applying any numerical method for CARE, without pre-knowledge on positive definitiness of C_t and D_t. This practical approach can also be used for simulations.*

By Theorem 3, the consistency of $\hat{\Theta}_T$ is inherited from the consistency of $\hat{\gamma}_T$. Similarly, the rate of convergence and the limiting distribution for $\hat{\Theta}_T$ follow from the convergence rate and the limiting distribution of $\hat{\gamma}_T$, respectively.

Theorem 4 *Let $X = (X_s)_{s\in[0,t]}$ be an n^2-dimensional stochastic process with continuous paths almost surely and let $l(T)$ be an arbitrary rate function. If*

$$l(T)\,\mathrm{vec}(\hat{\gamma}_T(s) - \gamma(s)) \xrightarrow{law} X_s \tag{17}$$

in the uniform topology of continuous functions, then:

(1) If $\tilde{X}_s$ is the permutation of elements of X_s that corresponds to the order of elements of $\operatorname{vec}(\gamma(s)^\top)$, we have

$$I(T)\operatorname{vec}(\Delta_T C, \Delta_T B, \Delta_T D) \xrightarrow{law} \begin{bmatrix} \int_0^t (t-s)(X_s + \tilde{X}_s)\,ds \\ \int_0^t \left(X_s - \tilde{X}_s\right) ds \\ -2X_0 + X_t + \tilde{X}_t \end{bmatrix} =: L_1(X).$$

(2) If $C, D > 0$ and $\hat{\Theta}_T$ is given by Definition 2, we have

$$I(T)\operatorname{vec}(\hat{\Theta}_T - \Theta) \xrightarrow{law} L_2(L_1(X)),$$

where $L_2 : \mathbb{R}^{3n^2} \to \mathbb{R}^{n^2}$ is a linear operator depending only on Θ, t and the cross-covariance of G.

Remark 7 *The operator L_2 is given explicitly in the proof, see page 25.*

2.1 | On the uniqueness of the solution to (12)

The uniqueness of the solution to (12) is crucially important, as otherwise we cannot guarantee that a convergent numerical scheme (which we have to apply in practice) converges to the true parameter Θ. In our case, it turns out that one can usually choose t such that $C_t, D_t > 0$ giving us uniqueness. We next address the uniqueness issue particularly in our case. For the general theory of algebraic Riccati equations, see e.g. Lancaster and Rodman (1995).

We begin with some definitions.

Definition 3 *A square matrix A is stable if all its eigenvalues are in the open left half-plane.*

Definition 4 *A matrix pair (A, B) is stabilizable if there exists a matrix K such that $A + BK$ is stable.*

Definition 5 *A real matrix pair (A, B) is detectable if $(B^\top, A^\top)$ is stabilizable.*

We utilise the following uniqueness result (for more details on the topic, see e.g. Kucera (1972) or Wonham (1968)) to our case under the assumption that $C_t, D_t \geq 0$.

Lemma 5 *Let $C_t, D_t \geq 0$. If (B_t, C_t) is stabilizable and (D_t, B_t) is detectable, then the continuous time algebraic Riccati equation* (12) *has a unique positive semidefinite solution Θ. Furthermore, the matrix $B_t - C_t\Theta$ is stable.*

With this we obtain the following useful corollary.

Corollary 6 *Let $C_t, D_t > 0$. Then* (12) *has a unique positive definite solution Θ.*

Proof *Let S be any stable matrix and set $K_1 = C_t^{-1}(S - B_t)$ and $K_2 = (D_t^\top)^{-1}(S - B_t^\top)$. Then $B_t + C_t K_1 = S = B_t^\top + D_t^\top K_2$, and the conditions of Lemma 5 are satisfied. Thus CARE* (12) *has a unique solution $\Theta \geq 0$, which is then automatically the true parameter matrix $\Theta > 0$.*

Let us now address when one can choose t such that $C_t, D_t > 0$. For this recall that

$$C_t = \mathbb{E}\left[\int_0^t U_s ds \left(\int_0^t U_s ds\right)^\top\right] = \operatorname{cov}\left(\int_0^t U_s ds\right)$$

and

$$D_t = \operatorname{cov}(G_t) - \operatorname{cov}(U_t - U_0).$$

Thus $C_t \geq 0$ for every t. Consider now the matrix D_t. By stationarity of U the elements of $\operatorname{cov}(U_t - U_0)$ are uniformly bounded, implying $a^\top \operatorname{cov}(U_t - U_0)a < C\|a\|^2$ for some constant C. On the other hand, we have $a^\top \operatorname{cov}(G_t)a \geq \lambda_{min}\|a\|^2$, where λ_{min} is the smallest eigenvalue of $\operatorname{cov}(G_t)$. Thus

$$a^\top D_t a \geq (\lambda_{min} - C)\|a\|^2,$$

implying $D_t > 0$ provided that λ_{min} grows sufficiently. This happens, for example, when G has independent components with growing variances.

Consider next the matrix C_t. Since $C_t \geq 0$ always, it suffices to find one t such that $C_t > 0$. Let us, for a moment, suppose that this is not possible. Then $\operatorname{rank}(C_t) \leq n - 1$ implying that there exists a (vector-valued) function $a(t)$

such that, almost surely and for all t,

$$a(t)^\top \int_0^t U_s ds = 0.$$

Without loss of generality we can assume that $a(t)$ is normalised and oriented consistently. Furthermore, it follows from the continuity of $\int_0^t U_s ds$ that $a(t)^\top$ is also continuous. This further implies that $a(t)^\top \int_0^t U_s ds$ is indistinguishable from the zero process meaning that there exists $B \subset \Omega$ such that $\mathbb{P}(B) = 1$ and

$$\int_0^t U_s(\omega) ds \in M_t^{n-1} \quad \text{for every } \omega \in B \text{ and } t \in \mathbb{R}, \tag{18}$$

where M_t^{n-1} is a n-1-dimensional subspace of $\mathbb{R}^n$. We claim that this implies also degeneracy of the process U itself. We, again, proceed by contradiction and assume that there exists $\omega_i \in B$, $i = 1, 2, \ldots, n$ such that the vectors $U_0(\omega_i)$ are linearly independent. Then the matrix

$$[U_0(\omega_1), \cdots, U_0(\omega_n)] \tag{19}$$

is invertible. On the other hand, for any $\varepsilon > 0$ we can apply the mean value theorem to find $\delta > 0$ such that

$$\int_0^\delta U_s(\omega_i) ds = (U_0(\omega_i) + \varepsilon_{\omega_i,\delta})\delta, \quad \text{with } \|\varepsilon_{\omega_i,\delta}\| < \varepsilon.$$

Thus, by continuity of the eigenvalues and invertibility of the matrix (19), the matrix

$$\left[U_0(\omega_1) + \varepsilon_{\omega_1,\delta} \cdots U_0(\omega_n) + \varepsilon_{\omega_n,\delta}\right]$$

is invertible as well provided that ε is chosen small enough. This contradicts (18), meaning that if $\text{rank}(C_t) \leq n-1$, then we have (18) and $U_0(\omega) \in \tilde{M}_0^{n-1}$ for all $\omega \in B$ as well. Now stationarity of U implies that $\mathbb{P}(U_t \in \tilde{M}_0^{n-1}) = 1$ for all $t \in \mathbb{R}$, meaning that U is a degenerate process. In particular, then

$$b^\top U_0 = \int_{-\infty}^0 b^\top e^{\Theta s} dG_s = \int_{-\infty}^0 \sum_{i=1}^n \left(b^\top e^{\Theta s}\right)^{(i)} dG_s^{(i)} = \sum_{i=1}^n \int_{-\infty}^0 \left(b^\top e^{\Theta s}\right)^{(i)} dG_s^{(i)} \overset{\text{a.s.}}{=} 0$$

for some non-zero vector b. If now G has independent components, then we would also get

$$\int_{-\infty}^0 \left(b^\top e^{\Theta s}\right)^{(i)} dG_s^{(i)} \overset{\text{a.s.}}{=} 0 \quad \text{for all } i.$$

For many interesting processes $G^{(i)}$ this would further imply $\left(b^\top e^{\Theta s}\right)^{(i)} \equiv 0$ leading to a contradiction since $e^{\Theta s}$ is of full-rank. In particular, this is the case whenever $G^{(i)}$ is a Gaussian process for which Wiener integral is injective (for details on Wiener integrals, see e.g. Janson (1997)). Such Gaussian processes include, among others, Brownian motions and fractional Brownian motions. Finally, we note that in general, if we have a set of observations $\{U_t(\omega)\}_{t\in I}$ (with a fixed ω) and $\text{span}\{U_t(\omega)\}_{t\in I} = \mathbb{R}^n$, then one can always find t such that $C_t > 0$.

2.2 | Application to Gaussian processes

In this subsection, we illustrate the applicability of our results to the Gaussian case. That is, we suppose that the components G are independent Gaussian processes $G^{(i)}$. We state the results under conditions on the cross-covariance $\gamma(t)$. In practice, one can verify the assumptions for a given model by computing $\gamma(t)$ from the variance matrix $v(t) = \mathbb{E}[G_t G_t^\top]$. In particular, different representations for $\gamma(t)$ are given in Subsection A.4. We apply these representations to prove that all our results are applicable, whenever the components $G^{(i)}$ are independent fractional Brownian motions with Hurst indices $H_i < 3/4$ (cf. Corollary 9 below).

We first note that, by assumption, the components $G^{(i)}$ have continuous paths almost surely. By Gaussianity, this implies L^2 continuity as well, and hence (7) is valid, giving $G \in \mathcal{G}_\Theta$. For the cross-covariance estimator $\hat{\gamma}$, we use standard

$$\hat{\gamma}_T(\tau) = \frac{1}{T}\int_0^T U_{s+\tau}U_s^\top ds.$$

The following result gives us the consistency immediately, and covers all ergodic systems.

Proposition 7 *Let G be a vector of Gaussian processes. If $\lim_{t\to\infty}\|\gamma(t)\| = 0$, then $\|\hat{\Theta}_T - \hat{\Theta}\| \xrightarrow{\mathbb{P}} 0$.*

Proposition 7 guarantees that we can apply Theorem 3 if the cross-covariance $\gamma(t)$ vanishes at infinity. Similarly, we may apply Theorem 4 if $\gamma(t)$ decays rapidly enough.

Theorem 8 *Suppose that $\gamma(r)$ is locally absolutely continuous and*

$$\max\left(\|\gamma'(r)\|, \|\gamma(r)\|\right) \le h(r)$$

for some non-increasing function $h(r)$ such that, for some $K > 0$, we have $h(r) \in L^1([0,K])$ and $h(r) \in L^2([K,\infty))$. Then

$$\sqrt{T}\,\mathrm{vec}(\hat{\gamma}_T(s) - \gamma(s)) \xrightarrow{law} X_s \tag{20}$$

in the uniform topology of continuous functions, where X is an n^2-dimensional centered Gaussian process. In particular, Theorem 4 is applicable.

Remark 8 *The cross-covariance $\mathbb{E}[X_\tau X_\eta^\top]$ of the process X can be computed explicitly, and it consists of elements*

$$\int_0^\infty \gamma_{i,j}(r+\tau)\gamma_{p,q}(r+\eta)dr, \quad i,j,p,q \in \{1,2,\ldots,n\} \tag{21}$$

in the order corresponding to $\mathrm{vec}(\hat{\gamma}_T(s) - \gamma(s))$. We also note that assumptions on the function h ensures that the terms (21) *are finite.*

Remark 9 *By representation (42), the differentiability of $\gamma(r)$ follows provided that the variance functions $v_i(r)$ of the components $G^{(i)}$ are differentiable.*

Remark 10 *Convergence of finite dimensional distributions in the above result follows from some well-known facts. However, to the best of our knowledge, tightness of $\hat{\gamma}_T(t)$ with lag t as a free parameter, has not previously been acknowledged in the literature making it the most significant point of our example.*

To end this section we apply Theorem 8 to the case of *multidimensional fractional Ornstein-Uhlenbeck processes*. Recall that a fractional Brownian motion B^H with Hurst index $H \in (0,1)$ is a centered Gaussian process with covariance

$$R_{B^H}(t,s) = \frac{1}{2}\left[t^{2H} + s^{2H} - |t-s|^{2H}\right].$$

Corollary 9 *Let G be a vector of independent fractional Brownian motions B^{H_i} with Hurst indices $H_i < \frac{3}{4}$. Then*

$$\sqrt{T}\,\mathrm{vec}(\hat{\gamma}_T(s) - \gamma(s)) \xrightarrow{law} X_s \tag{22}$$

in the uniform topology of continuous functions, where X is an n^2-dimensional centered Gaussian process. In particular, Theorem 4 is applicable.

Remark 11 *By carefully examining our proof we actually observe that the tightness holds for arbitrary values of the Hurst indices $H_i \in (0,1)$. Indeed, this follows since $\gamma(t) \sim t^{2H_{max}-2}$ at infinity, giving us the expected rate function $l(T)$ (cf. Proposition 23). Thus it suffices to study only the convergence of finite dimensional distributions.*

The above results are obviously just illustrations how our general theorems can be applied. For example, it is straightforward to check the applicability of Theorem 8 in the multidimensional versions of *the fractional Ornstein-Uhlenbeck process of the second kind* or *the bifractional Ornstein-Uhlenbeck process of the second kind* (see Sottinen and Viitasaari (2018) and the references therein for definitions). Indeed, it can be shown that, as in the univariate case, covariances $\gamma_{ij}(t)$ decay exponentially. Similarly, in the case of multidimensional fractional Ornstein-Uhlenbeck process where some of the Hurst indices H_i satisfy $H_i \geq 3/4$, we can obtain a limiting object, but with different rate and possibly different limiting object X_s. For example, if $\max H_i = 3/4$, then the rate is $\sqrt{T}/\sqrt{\log T}$ instead of standard $\sqrt{T}$, while the limiting process X_s is still Gaussian. If $\max H_i > 3/4$, one expects to have Rosenblatt components in X. Indeed, this is a well-known fact in dimension one (see, e.g. Hu et al. (2019)), and the tightness holds on the full range $H_i \in (0,1)$ (see Remark 11).

references

Applebaum, D. (2015) Infinite dimensional Ornstein-Uhlenbeck processes driven by Lévy processes. *Probab. Surv.*, **12**, 33–54.

Azmoodeh, E. and Viitasaari, L. (2015) Parameter estimation based on discrete observations of fractional Ornstein-Uhlenbeck process of the second kind. *Stat. Inference Stoch. Process.*, **18**, 205–227.

Bajja, S., Es-Sebaiy, K. and Viitasaari, L. (2017) Least squares estimator of fractional Ornstein-Uhlenbeck processes with periodic mean. *J. Korean Math. Soc.*, **46**, 608–622.

Balde, M. F., Es-Sebaiy, K. and Tudor, C. A. (2018) Ergodicity and drift parameter estimation for infinite-dimensional fractional Ornstein-Uhlenbeck process of the second kind. *Appl. Math. Optim.*

Behme, A. and Lindner, A. (2012) Multivariate generalized Ornstein–Uhlenbeck processes. *Stoch Process Their Appl.*, **122**, 1487–1518.

Bini, D. A., Iannazzo, B. and Meini, B. (2011) *Numerical solution of algebraic Riccati equations*, vol. 9. Siam.

Brouste, A. and Iacus, S. M. (2013) Parameter estimation for the discretely observed fractional Ornstein-Uhlenbeck process and the Yuima R package. *Comput. Statist.*, **28**, 1529–1547.

Byers, R. (1987) Solving the algebraic Riccati equation with the matrix sign function. *Linear Algebra Appl.*, **85**, 267–279.

Campese, S., Nourdin, I. and Nualart, D. (2020) Continuous Breuer-Major theorem: tightness and non-stationarity. *Ann. Probab.*, **48**, 147–177.

Cheridito, P., Kawaguchi, H. and Maejima, M. (2003) Fractional Ornstein-Uhlenbeck processes. *Electron. J. Probab.*, **8**, no. 3, 14 pp. (electronic).

Dehling, H., Franke, B. and Woerner, J. H. (2017) Estimating drift parameters in a fractional Ornstein-Uhlenbeck process with periodic mean. *Stat. Inference Stoch. Process.*, **20**, 1–14.

Douissi, S., Es-Sebaiy, K. and Tudor, C. A. (2020) Hermite Ornstein-Uhlenbeck processes mixed with a Gamma distribution. *Publ. Math. Debrecen*, **96**, 1–22.

Es-Sebaiy, K. and Ndiaye, D. (2014) On drift estimation for non-ergodic fractional Ornstein-Uhlenbeck process with discrete observations. *Afr. Stat.*, **9**, 615–625.

Es-Sebaiy, K. and Tudor, C. A. (2015) Fractional Ornstein-Uhlenbeck processes mixed with a gamma distribution. *Fractals*, **23**.

Horn, R. A. and Johnson, C. R. (1991) *Topics in matrix analysis*. Cambridge University Press.

Hu, Y. and Nualart, D. (2010) Parameter estimation for fractional Ornstein-Uhlenbeck processes. *Statist. Probab. Lett.*, **80**, 1030–1038.

Hu, Y., Nualart, D. and Zhou, H. (2019) Parameter estimation for fractional Ornstein-Uhlenbeck processes of general Hurst parameter. *Stat. Inference Stoch. Process.*, **22**, 111–142.

Janson, S. (1997) *Gaussian Hilbert spaces*, vol. 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge.

Kaarakka, T. and Salminen, P. (2011) On fractional Ornstein-Uhlenbeck processes. *Communications on Stochastic Analysis*, **5**, 121–133.

Kalman, R. (1960) Contributions to the theory of optimal control. *Bol. Soc. Mat. Mex.*, **5**, 102–119.

Kleptsyna, M. L. and Breton, A. L. (2002) Statistical analysis of the fractional Ornstein-Uhlenbeck type process. *Stat. Inference Stoch. Process.*, **5**, 229–248.

Kozachenko, Y., Melnikov, A. and Mishura, Y. (2015) On drift parameter estimation in models with fractional Brownian motion. *Statistics*, **49**, 35–62.

Kubilius, K., Mishura, Y., Ralchenko, K. and Seleznjev, O. (2015) Consistency of the drift parameter estimator for the discretized fractional Ornstein-Uhlenbeck process with Hurst index $H \in (0, \frac{1}{2})$. *Electron. J. Stat.*, **9**, 1799–1825.

Kucera, V. (1972) A contribution to matrix quadratic equations. *IEEE Trans. Automat. Control*, **17**, 344–347.

Lancaster, P. and Rodman, L. (1995) *Algebraic Riccati equations*. Clarendon Press.

Laub, A. (1979) A Schur method for solving algebraic Riccati equations. *IEEE Trans. Automat. Control*, **24**, 913–921.

Magdziarz, M. (2008) Fractional Ornstein-Uhlenbeck processes. Joseph effect in models with infinite variance. *Phys. A*, **387**, 123–133.

Maslowski, B. and Pospíšil, J. (2008) Ergodicity and drift parameter estimation for infinite-dimensional fractional Ornstein-Uhlenbeck process. *Appl. Math. Optim.*, **57**, 401–429.

Mishura, Y. S. (2008) *Stochastic calculus for fractional Brownian motion and related processes*, vol. 1929 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin.

Nourdin, I. and Tran, D. (2019) Statistical inference for Vasicek-type model driven by Hermite processes. *Stochastic Process. Appl.*, **129**, 3774–3791.

Shen, G., Yin, X. and Yan, L. (2016) Least squares estimation for Ornstein–Uhlenbeck processes driven by the weighted fractional Brownian motion. *Acta Math. Sci. Ser. B*, **36**, 394–408.

Shen, L. and Xu, Q. (2014) Asymptotic law of limit distribution for fractional Ornstein-Uhlenbeck process. *Adv. Difference Equ.*, 75.

Sottinen, T. and Viitasaari, L. (2018) Parameter estimation for the Langevin equation with stationary-increment Gaussian noise. *Stat. Inference Stoch. Process.*, **21**, 569–601.

Stewart, G. W. and Sun, J.-g. (1990) *Matrix perturbation theory*. Academic Press, Boston.

Sun, J.-g. (1998) Perturbation theory for algebraic Riccati equations. *SIAM J. Matrix Anal. Appl*, **19**, 39–65.

Sun, X. and Guo, F. (2015) On integration by parts formula and characterization of fractional Ornstein-Uhlenbeck process. *Statist. Probab. Lett.*, **107**, 170–177.

Tanaka, K. (2015) Maximum likelihood estimation for the non-ergodic fractional Ornstein-Uhlenbeck process. *Stat. Inference Stoch. Process.*, **18**, 315–332.

Viitasaari, L. (2016) Representation of stationary and stationary increment processes via Langevin equation and self-similar processes. *Statist. Probab. Lett.*, **115**, 45–53.

Voutilainen, M., Viitasaari, L. and Ilmonen, P. (2017) On model fitting and estimation of strictly stationary processes. *Mod. Stoch. Theory Appl.*, **4**, 381–406.

— (2019) Note on AR(1)-characterisation of stationary processes and model fitting. *Mod. Stoch. Theory Appl.*, **6**, 195–207.

Wonham, W. M. (1968) On a matrix Riccati equation of stochastic control. *SIAM Journal on Control*, **6**, 681–697.

Corresponding author:
Marko Voutilainen
mtvout@utu.fi
Turku School of Economics
Department of Accounting and Finance
FI-20014 University of Turku
Turku, Finland

A | PROOFS

For the reader's convenience, we divide this section into six subsections. The first subsection, Subsection A.1, provides a proof of Theorem 1 motivating our model. Subsection A.2 contains a proof of Theorem 2 that leads to the definition of our estimator $\hat{\Theta}_T$. In Subsection A.3 we prove our results, Theorem 3 and Theorem 4, concerning the asymptotic behaviour of $\hat{\Theta}_T$. In Subsection A.4, we provide representations for the cross-covariance $\gamma(t)$. These representations will then be applied in Subsection A.5, where we prove results related to our Gaussian example.

A.1 | Proof of Theorem 1

The proof of Theorem 1 follows the strategy of Viitasaari (2016). However, in multidimensional setting one has to be carefully, e.g. whether matrices commute or not. In addition, we need to extend concepts such as self-similarity to the matrix-valued case. For this reason, we do not omit the proof even though it is partly very similar to the univariate case.

We begin with the following definition of Θ-self-similar processes, where Θ is a matrix.

Definition 6 *Let $\Theta > 0$. An n-dimensional stochastic process $X = (X_t)_{t\geq 0}$ with $X_0 = 0$ is Θ-self-similar if*

$$(X_{at})_{t\geq 0} \overset{law}{=} a^{\Theta}(X_t)_{t\geq 0}$$

for every $a > 0$ in the sense of finite dimensional distributions. Here the matrix exponent is defined through the matrix exponential function $a^{\Theta} = e^{\Theta \log a}$.

The following remark illustrates the necessity of positive definiteness of Θ.

Remark 12 *The assumption $\Theta > 0$ is natural, as otherwise we may reduce the number of dimensions. Indeed, if $\Theta \geq 0$ with one eigenvalue $\lambda_1 = 0$, then the eigendecomposition $\Theta = Q\Lambda Q^{\top}$ gives*

$$X_a \overset{law}{=} Qe^{\Lambda \log a}Q^{\top}X_1 = Q\mathrm{diag}(e^{\lambda_i \log a})Z,$$

with $Z = Q^\top X_1$. Since Q is orthogonal, it follows that

$$\|X_a\| \overset{law}{=} \|Q\mathrm{diag}(e^{\lambda_i \log a})Z\| \geq |Z^{(1)}| = |(Q^\top X_1)^{(1)}|.$$

In particular, using continuity of X and letting $a \to 0$ yields $(Q^\top X_1)^{(1)} = 0$. This means that X is an $(n-1)$-dimensional process. Similarly, if $G \in \mathcal{G}_\Theta$ with Θ having zero as an eigenvalue with algebraic multiplicity equal k, then G degenerates to an $(n-k)$-dimensional process.

Definition 7 *Let $\Theta > 0$. In addition, let $U = (U_t)_{t\in\mathbb{R}}$ and $X = (X_t)_{t\geq 0}$ be n-dimensional stochastic processes. We define*

$$(\mathcal{L}_\Theta U)_t = t^\Theta U_{\log t}, \quad \text{for } t > 0$$
$$(\mathcal{L}_\Theta^{-1} X)_t = e^{-\Theta t} X_{e^t}, \quad \text{for } t \in \mathbb{R}.$$

The following result extends the one-to-one correspondence between Θ-self-similar processes and stationary processes to the matrix-valued case. We use the name Lamperti transform for our matrix-valued version in honour to the original univariate result.

Theorem 10 (Lamperti) *Let $\Theta > 0$. Let $(U_t)_{t\in\mathbb{R}}$ be an n-dimensional stationary process. Then $(\mathcal{L}_\Theta U)_t$ is Θ-self-similar. Conversely, let $(X_t)_{t\geq 0}$ be an n-dimensional Θ-self-similar process. Then $(\mathcal{L}_\Theta^{-1} X)_t$ is stationary.*

Proof *Suppose first that $(U_t)_{t\in\mathbb{R}}$ is stationary. Define $Y_t = (\mathcal{L}_\Theta U)_t = t^\Theta U_{\log t}$. Let $a > 0$ and $[t_1, t_2, \cdots, t_m]^\top \in \mathbb{R}_+^m$. Then*

$$\begin{aligned}(Y_{at_1}, Y_{at_2}, \cdots, Y_{at_m}) &= (a^\Theta t_1^\Theta U_{\log at_1}, a^\Theta t_2^\Theta U_{\log at_2}, \cdots, a^\Theta t_m^\Theta U_{\log at_m})\\ &= (a^\Theta t_1^\Theta U_{\log a + \log t_1}, a^\Theta t_2^\Theta U_{\log a+\log t_2}, \cdots, a^\Theta t_m^\Theta U_{\log a+\log t_m}) \overset{law}{=} (a^\Theta Y_{t_1}, a^\Theta Y_{t_2}, \cdots, a^\Theta Y_{t_m}).\end{aligned}$$

Now let $\Theta = Q\Lambda Q^\top$ be an eigendecomposition of Θ. Then

$$t^\Theta = e^{\Theta \log t} = Q \sum_{k=0}^{\infty} \frac{\Lambda^k (\log t)^k}{k!} Q^\top = Q e^{\Lambda \log t} Q^\top, \tag{23}$$

where $e^{\Lambda \log t}$ is a diagonal matrix with diagonal elements of the form $e^{\lambda_i \log t}$. Since $\lambda_i > 0$ for every $i = 1, 2, \ldots, n$, we notice that $\lim_{t\to 0} Y_t = 0$ in probability, and hence, Y is Θ-self-similar.
Next, suppose that $(X_t)_{t\geq 0}$ is Θ-self-similar. Define $Y_t = (\mathcal{L}_\Theta^{-1} X)_t = e^{-\Theta t} X_{e^t}$. Let $s \in \mathbb{R}$ and $[t_1, t_2, \cdots, t_m]^\top \in \mathbb{R}^m$. Then

$$\begin{aligned}(Y_{t_1+s}, Y_{t_2+s}, \cdots, Y_{t_m+s}) &= (e^{-\Theta(t_1+s)} X_{e^{t_1+s}}, e^{-\Theta(t_2+s)} X_{e^{t_2+s}}, \cdots, e^{-\Theta(t_m+s)} X_{e^{t_m+s}})\\ &\overset{law}{=} (e^{-\Theta t_1} X_{e^{t_1}}, e^{-\Theta t_2} X_{e^{t_2}}, \cdots, e^{-\Theta t_m} X_{e^{t_m}}) = (Y_{t_1}, Y_{t_2}, \cdots, Y_{t_m})\end{aligned}$$

concluding the proof.

The following lemma is a straightforward extension of a similar univariate result of Viitasaari (2016). For the reader's convenience, we present the proof here.

Lemma 11 *Let $(X_t)_{t\geq 0}$ be an n-dimensional Θ-self-similar process. Define $Y = (Y_t)_{t\in\mathbb{R}}$ by*

$$Y_t = \int_0^t e^{-\Theta u} dX_{e^u}.$$

Then $Y \in \mathcal{G}_\Theta$.

Proof *Clearly $Y_0 = 0$. In addition*

$$\int_{-\infty}^0 e^{\Theta u} dY_u = \int_{-\infty}^0 dX_{e^u} = X_1 - \lim_{t\to-\infty} X_{e^t} \overset{\mathbb{P}}{=} X_1.$$

Now, let $t, s, h \in \mathbb{R}$. Then

$$Y_t - Y_s = \int_s^t e^{-\Theta u} dX_{e^u} = \int_{s+h}^{t+h} e^{-\Theta(v-h)} dX_{e^{v-h}} \overset{law}{=} \int_{s+h}^{t+h} e^{-\Theta v} dX_{e^v} = Y_{t+h} - Y_{s+h},$$

where we have used the change of variable $u = v - h$. The penultimate equation can be verified by approximating the integral with Riemann sums, using self-similarity, and passing to the limit. Similarly, for multidimensional distributions

$$\begin{bmatrix} Y_{t_1} - Y_{s_1} \\ Y_{t_2} - Y_{s_2} \\ \vdots \\ Y_{t_m} - Y_{s_m} \end{bmatrix} = \begin{bmatrix} \int_{s_1}^{t_1} e^{-\Theta u} dX_{e^u} \\ \int_{s_2}^{t_2} e^{-\Theta u} dX_{e^u} \\ \vdots \\ \int_{s_m}^{t_m} e^{-\Theta u} dX_{e^u} \end{bmatrix} = \begin{bmatrix} \int_{s_1+h}^{t_1+h} e^{-\Theta(v-h)} dX_{e^{v-h}} \\ \int_{s_2+h}^{t_2+h} e^{-\Theta(v-h)} dX_{e^{v-h}} \\ \vdots \\ \int_{s_m+h}^{t_m+h} e^{-\Theta(v-h)} dX_{e^{v-h}} \end{bmatrix} \overset{law}{=} \begin{bmatrix} \int_{s_1+h}^{t_1+h} e^{-\Theta v} dX_{e^v} \\ \int_{s_2+h}^{t_2+h} e^{-\Theta v} dX_{e^v} \\ \vdots \\ \int_{s_m+h}^{t_m+h} e^{-\Theta v} dX_{e^v} \end{bmatrix} = \begin{bmatrix} Y_{t_1+h} - Y_{s_1+h} \\ Y_{t_2+h} - Y_{s_2+h} \\ \vdots \\ Y_{t_m+h} - Y_{s_m+h} \end{bmatrix}.$$

We split the proof of Theorem 1 into three lemmas. The first one gives us the stationary solution to (3).

Lemma 12 *Let $\Theta > 0$ and $G \in \mathcal{G}_\Theta$. Then the unique solution to the Langevin equation* (3) *with the initial condition*

$$U_0 = \int_{-\infty}^0 e^{\Theta s} dG_s$$

is given by

$$U_t = e^{-\Theta t} \int_{-\infty}^t e^{\Theta s} dG_s.$$

The solution is stationary.

Proof *By integration by parts*

$$U_t = e^{-\Theta t} \int_{-\infty}^t e^{\Theta s} dG_s = G_t - e^{-\Theta t}\Theta \int_{-\infty}^t e^{\Theta s} G_s ds,$$

giving

$$dU_t = dG_t - d\left(e^{-\Theta t}\Theta\right) \int_{-\infty}^t e^{\Theta s} G_s ds - e^{-\Theta t}\Theta d\left(\int_{-\infty}^t e^{\Theta s} G_s ds\right) = dG_t - \Theta d\left(e^{-\Theta t}\right) \int_{-\infty}^t e^{\Theta s} G_s ds - \Theta e^{-\Theta t} e^{\Theta t} G_t dt.$$

Here we have used the fact that $e^{-\Theta t}$ and Θ commute. Now

$$\begin{aligned} dU_t &= dG_t - \Theta d\left(e^{-\Theta t}\right)\int_{-\infty}^t e^{\Theta s}G_s ds - \Theta G_t dt = dG_t + \left(\Theta^2 e^{-\Theta t}dt\right)\int_{-\infty}^t e^{\Theta s}G_s ds - \Theta G_t dt \\ &= dG_t - \Theta\left(G_t - \Theta e^{-\Theta t}\int_{-\infty}^t e^{\Theta s}G_s ds\right)dt = dG_t - \Theta U_t dt \end{aligned}$$

completing the proof of the first assertion. To show stationarity, change of variable $u = s - t$ gives us

$$U_t = e^{-\Theta t}\int_{-\infty}^t e^{\Theta s}dG_s = e^{-\Theta t}\int_{-\infty}^0 e^{\Theta(u+t)}dG_{u+t} = \int_{-\infty}^0 e^{\Theta u}dG_{u+t} \overset{law}{=} \int_{-\infty}^0 e^{\Theta u}dG_u = U_0,$$

where we have used that G has stationary increments. Again, the penultimate equation can be verified by approximating the integral with finite Riemann sums, using stationarity of increments, and passing to the limit. Treating multidimensional distributions similarly concludes the proof.

The next result gives us the other direction, i.e. it shows that stationary processes solve Langevin equation.

Lemma 13 *Let $\Theta > 0$ be fixed and let $U = (U_t)_{t\in\mathbb{R}}$ be stationary process with continuous paths. Then U is the unique solution to the Langevin equation* (3) *for some $G \in \mathcal{G}_H$ and the initial condition*

$$U_0 = \int_{-\infty}^0 e^{\Theta s}dG_s.$$

Proof *Assume that $(U_t)_{t\in\mathbb{R}}$ is stationary. Then by Theorem 10 there exists a Θ-self-similar $(X_t)_{t\geq 0}$ such that $U_t = (\mathcal{L}_\Theta^{-1}X)_t = e^{-\Theta t}X_{e^t}$. Consequently*

$$dU_t = d(e^{-\Theta t})X_{e^t} + e^{-\Theta t}dX_{e^t} = -\Theta e^{-\Theta t}X_{e^t}dt + e^{-\Theta t}dX_{e^t} = -\Theta U_t dt + e^{-\Theta t}dX_{e^t}.$$

Now, define $Y = (Y_t)_{t\in\mathbb{R}}$ as in Lemma 11. Then $Y \in \mathcal{G}_\Theta$ and $dY_t = e^{-\Theta t}dX_{e^t}$ concluding the proof.

Finally, the next lemma provides us with the uniqueness of the noise.

Lemma 14 *Let $\Theta > 0$ be fixed. Then a process $U = (U_t)_{t\in\mathbb{R}}$ satisfies the Langevin equation* (3) *with the initial*

$$U_0 = \int_{-\infty}^0 e^{\Theta s}dG_s \tag{24}$$

for one process $G \in \mathcal{G}_\Theta$ at the most.

Proof *Suppose that $G, \tilde{G} \in \mathcal{G}_\Theta$ yield the same solution U of the Langevin equation with the initial* (24)*. Then for every $t \in \mathbb{R}$*

$$e^{\Theta t}U_t = \int_{-\infty}^t e^{\Theta u}dG_u = \int_{-\infty}^t e^{\Theta u}d\tilde{G}_u.$$

Let $s < t$, then

$$\int_s^t e^{\Theta u}dG_u = \int_s^t e^{\Theta u}d\tilde{G}_u.$$

Integration by parts gives

$$\int_s^t e^{\Theta u} dG_u = e^{\Theta t} G_t - e^{\Theta s} G_s - \Theta \int_s^t e^{\Theta u} G_u du$$

yielding

$$e^{\Theta t}(G_t - \tilde{G}_t) - e^{\Theta s}(G_s - \tilde{G}_s) = \Theta \int_s^t e^{\Theta u}(G_u - \tilde{G}_u) du.$$

By denoting $h(t) = e^{\Theta t}(G_t - \tilde{G}_t)$, *we obtain*

$$h(t) - h(s) = \Theta \int_s^t h(u) du$$

or equivalently

$$dh(t) = \Theta h(t) dt. \tag{25}$$

The general solution to (25) *reads*

$$h(t) = Q e^{\Lambda t} C,$$

where the columns of Q *are equal to the eigenvectors of* Θ *and* Λ *is the corresponding eigenvalue diagonal matrix, and* C *is a constant vector. The initial* $G_0 = \tilde{G}_0$ *gives* $QC = 0$. *Since* Q *is invertible, we conclude that* $C = 0$.

The proof of Theorem 1 now follows directly.

Proof of Theorem 1 The existence of a unique stationary solution to the Langevin equation is in fact the statement of Lemma 12. Conversely, the fact that stationary processes solve Langevin equations is the statement of Lemma 13. Finally, Lemma 14 gives us the uniqueness of the noise.

A.2 | Proof of Theorem 2

In order to prove Theorem 2, we begin by showing that for $\Theta > 0$, the (7) implies $G \in \mathcal{G}_\Theta$, i.e. we show that for $\Theta > 0$,

$$\int_{-\infty}^0 e^{\Theta s} G_s ds$$

defines an almost surely finite random variable. For this we begin with the following very elementary lemma.

Lemma 15 *Let* $G = (G_s)_{s\in\mathbb{R}}$ *be a* 1*-dimensional centred process with stationary increments,* $G_0 = 0$, *and* $\sup_{s\in[0,1]} \mathbb{E}G_s^2 < \infty$. *Then*

$$\mathrm{var}(G_s) \le C(s+1)^2$$

for every $s \ge 0$, *where* C *is some positive constant depending only on the process* G.

Proof *Let $s \geq 0$. Writing $G_s = G_s - G_{\lfloor s \rfloor} + \sum_{k=1}^{\lfloor s \rfloor}(G_k - G_{k-1})$, where $\lfloor \cdot \rfloor$ is the standard floor-function, and using stationarity of the increments together with the Minkowski's inequality gives $\sqrt{\mathbb{E}G_s^2} \leq (s+1)\sup_{s\in[0,1]}\sqrt{\text{var}(G_s)}$.*

Lemma 16 *Let $\Theta > 0$ and let G satisfy (7). Then*

$$\lim_{u\to\infty}\int_{-u}^{0} e^{\Theta s} G_s ds \tag{26}$$

exists almost surely.

Proof *Let $\Theta = Q\Lambda Q^\top$ be an eigendecomposition of Θ. Then $e^{\Theta s} = Qe^{\Lambda s}Q^\top$, where $e^{\Lambda s}$ is a diagonal matrix with diagonal entries of the form $e^{\lambda_i s}$. Thus $\|e^{\Lambda s}\| = e^{\lambda_{min} s}$, where λ_{min} is the smallest eigenvalue of Θ. Moreover, by orthogonality of Q, we have $\|Q\|\|Q^\top\| = 1$. Thus*

$$\|e^{\Theta s}G_s\| \leq \|Q\|\|e^{\Lambda s}\|\|Q^\top\|\|G_s\| = e^{\lambda_{min} s}\|G_s\| \leq e^{\lambda_{min} s}\sqrt{n}\max_i |G_s^{(i)}|. \tag{27}$$

On the other hand, Lemma 15 gives

$$\mathbb{P}\left(\left|e^{\frac{\lambda_{min}}{2}s}G_s^{(i)}\right| > \epsilon\right) \leq \frac{\text{var}\left(e^{\frac{\lambda_{min}}{2}s}G_s^{(i)}\right)}{\epsilon^2} \leq \frac{C_i e^{\lambda_{min} s}(1+|s|)^2}{\epsilon^2}.$$

Thus Borel-Cantelli implies $\left|e^{\frac{\lambda_{min}}{2}s}G_s^{(i)}\right| \to 0$ almost surely as $s \to -\infty$, which further implies

$$\lim_{s\to-\infty}\max_i |e^{\frac{\lambda_{min}}{2}s}G_s^{(i)}| = 0 \tag{28}$$

almost surely. Hence we observe

$$\begin{aligned}\int_{-\infty}^{0}\|e^{\Theta s}G_s\|ds &\leq \int_{-\infty}^{0} C_n e^{\lambda_{min} s}\max_i |G_s^{(i)}|ds = C_n\int_{-\infty}^{0} e^{\frac{\lambda_{min}}{2}s}\max_i |e^{\frac{\lambda_{min}}{2}s}G_s^{(i)}|ds\\ &\leq C_n \sup_{s\in(-\infty,0]}\left\{\max_i |e^{\frac{\lambda_{min}}{2}s}G_s^{(i)}|\right\}\int_{-\infty}^{0} e^{\frac{\lambda_{min}}{2}s}ds,\end{aligned}$$

where the supremum term is finite almost surely by (28). This concludes the proof.

We are now ready to prove Theorem 2.

Proof of Theorem 2 Lemma 16 together with assumption (7) gives us $G \in \mathcal{G}_\Theta$, and by Theorem 1 the solution with initial (5) U is stationary. Now (3) and $G_0 = 0$ gives, for every $t \geq 0$, that

$$G_t - G_0 = G_t = U_t - U_0 + \Theta\int_0^t U_s ds. \tag{29}$$

In the sequel, we use the short notation $\Delta_t U = U_t - U_0$. Noticing that $\Theta^\top = \Theta$, we now get from (29) that

$$
\begin{aligned}
G_t G_t^\top &= \Delta_t U (\Delta_t U)^\top + \Delta_t U \left(\int_0^t U_s ds\right)^\top \Theta + \Theta \int_0^t U_s ds\ (\Delta_t U)^\top + \Theta \int_0^t U_s ds \left(\int_0^t U_s ds\right)^\top \Theta \\
&= \Delta_t U (\Delta_t U)^\top + \int_0^t \Delta_t U U_s^\top ds\ \Theta + \Theta \int_0^t U_s (\Delta_t U)^\top ds + \Theta \int_0^t \int_0^t U_s U_u^\top du ds \Theta.
\end{aligned}
$$

Taking expectation on both sides completes the proof. Indeed, the first order term with respect to Θ is

$$
\begin{aligned}
\int_0^t (\gamma(t-s) - \gamma(-s))\, ds\, \Theta + \Theta \int_0^t (\gamma(s-t) - \gamma(s))\, ds &= \int_0^t (\gamma(s) - \gamma(-s))\, ds\, \Theta + \Theta \int_0^t (\gamma(-s) - \gamma(s))\, ds \\
&= \int_0^t \left(\gamma(s) - \gamma(s)^\top\right) ds\, \Theta + \Theta \int_0^t \left(\gamma(s)^\top - \gamma(s)\right) ds \\
&= B_t \Theta + \Theta B_t^\top,
\end{aligned}
$$

where $\gamma(t)$ is the cross-covariance matrix of U given by (8). Computing other expectations similarly and rearranging terms gives us (12).

A.3 | Proofs of Theorem 3 and Theorem 4

We begin with some preliminary notation. Let B, C, and D denote the coefficient matrices of the original CARE (12), and let $\Phi = B - C\Theta$. We define a linear operator $L : S^n \to S^n$ by

$$
L(M) = \Phi^\top M + M\Phi.
$$

The operator L is bounded and invertible (see e.g. Stewart and Sun (1990) and Sun (1998)). In addition, we define linear operators $Q : S^n \to S^n$ and $P : \mathbb{R}^{n\times n} \to S^n$ by

$$
\begin{aligned}
Q(M) &= L^{-1}(\Theta M \Theta) \\
P(M) &= L^{-1}(\Theta M + M^\top \Theta).
\end{aligned}
$$

Since L is bounded, also Q and P are bounded operators. We set

$$
l = \frac{1}{\|L^{-1}\|}, \quad p = \|P\|, \quad q = \|Q\|
$$

and

$$
\begin{aligned}
\epsilon_T &= \frac{1}{l}\|\Delta_T D\| + p\|\Delta_T B\| + q\|\Delta_T C\| \\
\delta_T &= \|\Delta_T B\| + \|\Delta_T C\|\|\Theta\| \\
g_T &= \|C\| + \|\Delta_T C\| \\
\epsilon_T^* &= \frac{2l\epsilon_T}{l - 2\delta_T + \sqrt{(l-2\delta_T)^2 - 4lg_T\epsilon_T}}.
\end{aligned}
\tag{30}
$$

The following theorem is taken from Sun (1998), but stated using our notation.

Theorem 17 *Let Θ be the unique positive semidefinite solution to the CARE* (13)*. Then if the coefficient matrices $\hat{D}_T$ and $\hat{C}_T$ of the perturbed CARE* (14) *are positive semidefinite, and if*

$$\delta_T + \sqrt{l g_T \epsilon_T} < \frac{l}{2}, \tag{31}$$

then (14) *has a unique solution $\hat{\Theta}_T \geq 0$ satisfying*

$$\|\hat{\Theta}_T - \Theta\| \leq \epsilon_T^*.$$

Remark 13 *Theorem 17 holds for any unitarily invariant submultiplicative matrix norm $\|\cdot\|$, not just the spectral norm we are using. On the other hand, for our purposes the choice of the norm does not matter since in finite dimensions all norms are equivalent.*

Note that $\epsilon_T = O(\|\Delta_T D\| + \|\Delta_T B\| + \|\Delta_T C\|)$, where O denotes the usual Landau notation with meaning $X_T = O(Y_T)$ if $|X_T| \leq c|Y_T|$ for some constant c. We also note that Theorem 17 gives us the following first-order perturbation bound

$$\|\hat{\Theta}_T - \Theta\| \leq \epsilon_T + O(\|\Delta_T D\|^2 + \|\Delta_T B\|^2 + \|\Delta_T C\|^2). \tag{32}$$

Recall that $\hat{\gamma}_T(s)$ and $\hat{\gamma}_{T,i,j}(s)$ denote some fixed estimators of $\gamma(s)$ and $\gamma_{i,j}(s)$ respectively. Next result gives bound (32) in terms of cross-covariance estimators.

Lemma 18 *For ϵ_T given in* (30)*, we have*

$$\epsilon_T \leq (2p + 2qt) \int_0^t \|\hat{\gamma}_T(s) - \gamma(s)\| ds + \frac{2}{l} \left(\|\hat{\gamma}_T(0) - \gamma(0)\| + \|\hat{\gamma}_T(t) - \gamma(t)\| \right).$$

In particular,

$$\epsilon_T \leq \sup_{s \in [0,t]} \|\hat{\gamma}_T(s) - \gamma(s)\| \left(2pt + 2qt^2 + \frac{4}{l} \right).$$

Proof *We have*

$$\Delta_T B = \int_0^t \left(\hat{\gamma}_T(s) - \hat{\gamma}_T(s)^\top \right) ds - \int_0^t \left(\gamma(s) - \gamma(s)^\top \right) ds = \int_0^t \left(\hat{\gamma}_T(s) - \gamma(s) \right) ds + \int_0^t \left(\gamma(s)^\top - \hat{\gamma}_T(s)^\top \right) ds$$

implying

$$\|\Delta_T B\| \leq 2 \int_0^t \|\hat{\gamma}_T(s) - \gamma(s)\| ds. \tag{33}$$

Similarly, for

$$\Delta_T C = \int_0^t \int_0^t \hat{\gamma}_T(s-u) du ds - \int_0^t \int_0^t \gamma(s-u) du ds$$

we get

$$\|\Delta_T C\| \leq \int_0^t \int_0^t \|\hat{\gamma}_T(s-u) - \gamma(s-u)\| du ds \leq 2t \int_0^t \|\hat{\gamma}_T(s) - \gamma(s)\| ds. \tag{34}$$

Finally, for $\Delta_T D$ *we have*

$$\Delta_T D = \hat{\Sigma}_{G_t} - \hat{\Sigma}_{U_t - U_0} - \Sigma_{G_t} + \Sigma_{U_t - U_0} = \Sigma_{U_t - U_0} - \hat{\Sigma}_{U_t - U_0},$$

where Σ_X *denotes the covariance matrix of a vector* X*. Now*

$$\Sigma_{U_t - U_0} = 2\gamma(0) - \gamma(t) - \gamma(t)^\top$$

giving us

$$\|\Delta_T D\| \leq 2\|\hat{\gamma}_T(0) - \gamma(0)\| + 2\|\hat{\gamma}_T(t) - \gamma(t)\|. \tag{35}$$

Combining the estimates (33), (34) *and* (35) *concludes the proof.*

Corollary 19 *Let* Θ *be the unique positive definite solution to the CARE* (13)*. If the coefficient matrices* $\hat{D}_T$ *and* $\hat{C}_T$ *of the perturbed CARE* (14) *are positive semidefinite, and if*

$$\delta_T + \sqrt{l g_T \epsilon_T} < \frac{l}{2}, \tag{36}$$

then (14) *has a unique positive semidefinite solution* $\hat{\Theta}_T$*. Moreover,*

$$\|\hat{\Theta}_T - \Theta\| \leq c \sup_{s \in [0,t]} \|\hat{\gamma}_T(s) - \gamma(s)\| + O\left(\sup_{s \in [0,t]} \|\hat{\gamma}_T(s) - \gamma(s)\|^2\right),$$

where the constant c *depends on* n, Θ, t *and the cross-covariance of* G.

Proof *The claim follows directly from the bound* (32)*, Lemma 18 and the proof of Lemma 18.*

We are now in position to proof our consistency result, Theorem 3.

Proof of Theorem 3 We first pick $\rho > 0$ small enough such that if

$$\sup_{s \in [0,t]} \|\hat{\gamma}_T(s) - \gamma(s)\| \leq \rho, \tag{37}$$

then $\delta_T + \sqrt{l g_T \epsilon_T} < l/2$. Since $C, D > 0$, we have that $\hat{C}_T, \hat{D}_T$ are positive definite whenever ρ is chosen small enough. Moreover, in this case Corollary 19 gives

$$\|\hat{\Theta}_T - \Theta\| \leq \tilde{C} \sup_{s \in [0,t]} \|\hat{\gamma}_T(s) - \gamma(s)\|. \tag{38}$$

Next, let $\varepsilon > 0$ be arbitrary such that $\varepsilon/\tilde{C} \le \rho$ and set

$$A_{T,\varepsilon} := \left\{\omega : \sup_{s\in[0,t]} \|\hat{\gamma}_T(s) - \gamma(s)\| \le \frac{\varepsilon}{\tilde{C}}\right\}.$$

Now a unique positive semidefinite solution $\hat{\Theta}_T$ to (14) exists for $\omega \in A_{T,\varepsilon}$, and we have (38). Moreover, assumption $\sup_{s\in[0,t]} \|\hat{\gamma}_T(s) - \gamma(s)\| \xrightarrow{\mathbb{P}} 0$ implies that for any $\xi > 0$ there exists $T_{\varepsilon,\xi}$ such that for every $T \ge T_{\varepsilon,\xi}$ we have $\mathbb{P}(A_{T,\varepsilon}) \ge 1 - \xi/2$. Thus we can conclude

$$\begin{aligned}\mathbb{P}\left(\|\hat{\Theta}_T - \Theta\| > \varepsilon\right) &\le \mathbb{P}\left(1_{A_{T,\varepsilon}}\|\hat{\Theta}_T - \Theta\| > \varepsilon\right) + \mathbb{P}\left(1_{A^c_{T,\varepsilon}}\|\hat{\Theta}_T - \Theta\| > \varepsilon\right)\\ &\le \mathbb{P}\left(\sup_{s\in[0,t]} \|\hat{\gamma}_T(s) - \gamma(s)\| > \frac{\varepsilon}{\tilde{C}}\right) + \mathbb{P}(A^c_{T,\varepsilon}) \le \frac{\xi}{2} + \frac{\xi}{2} = \xi.\end{aligned}$$

This concludes the proof.

Before proving Theorem 4 we recall an auxiliary lemma, taken from Horn and Johnson (1991).

Lemma 20 *Let E and F be square matrices of sizes m and n, respectively. Then all the eigenvalues of the Kronecker sum*

$$E \oplus F = (I_n \otimes E) + (F \otimes I_m)$$

are of the form $\lambda_i + \lambda_j$, where λ_i is an eigenvalue of E and λ_j is an eigenvalue of F.

Proof of Theorem 4 **Item (1)**:
Recall that

$$\begin{aligned}\Delta_T D &= 2(\gamma(0) - \hat{\gamma}_T(0)) + \hat{\gamma}_T(t) - \gamma(t) + \hat{\gamma}_T(t)^\top - \gamma(t)^\top\\ \Delta_T B &= \int_0^t \hat{\gamma}_T(s) - \gamma(s)ds + \int_0^t \gamma(s)^\top - \hat{\gamma}_T(s)^\top ds.\end{aligned}$$

Similarly, using

$$\int_0^t\int_0^t \gamma(u-s)duds = \int_0^t\int_0^s \gamma(u-s)duds + \int_0^t\int_s^t \gamma(u-s)duds,$$

where

$$\begin{aligned}\int_0^t\int_0^s \gamma(u-s)duds &= \int_0^t\int_0^s \gamma(s-u)^\top duds = \int_0^t\int_0^s \gamma(z)^\top dzds\\ &= \int_0^t\int_z^t \gamma(z)^\top dsdz = \int_0^t (t-z)\gamma(z)^\top dz\end{aligned}$$

and

$$\int_0^t\int_s^t \gamma(u-s)duds = \int_0^t (t-z)\gamma(z)dz,$$

we get

$$\Delta_T C = \int_0^t \int_0^t \hat{\gamma}_T(s-u) - \gamma(s-u)\,du\,ds = \int_0^t (t-s)(\hat{\gamma}_T(s) - \gamma(s))ds + \int_0^t (t-s)(\hat{\gamma}_T(s)^\top - \gamma(s)^\top)ds.$$

Now by assumption, we have

$$l(T)\,\mathrm{vec}(\hat{\gamma}_T(s) - \gamma(s)) \xrightarrow{\text{law}} X_s,$$

where $X = (X_s)_{s\in[0,t]}$ is an n^2-dimensional stochastic process with continuous paths. Now it is clear that the linear mapping $L_1 : C[0,t]^{n^2} \to \mathbb{R}^{3n^2}$ defined by

$$L_1(X) = \begin{bmatrix} \int_0^t (t-s)(X_s + \tilde{X}_s)ds \\ \int_0^t X_s - \tilde{X}_s ds \\ -2X_0 + X_t + \tilde{X}_t, \end{bmatrix}$$

where $\tilde{X}_s$ denotes the permutation of the elements of X_s that corresponds to the order of the elements of $\mathrm{vec}(\gamma(s)^\top)$, is a continuous operator. Thus we may apply continuous mapping theorem to conclude that

$$L_1(l(T)\,\mathrm{vec}(\hat{\gamma}_T(s) - \gamma(s))) = l(T)\,\mathrm{vec}(\Delta_T C, \Delta_T B, \Delta_T D) \xrightarrow{\text{law}} L_1(X).$$

Item (2):
As in the proof of Theorem 3, set

$$A_T := \left\{\omega : \sup_{s\in[0,t]} \|\hat{\gamma}_T(s) - \gamma(s)\| \le \rho\right\},$$

where ρ is chosen as in the proof of Theorem 3. Then the unique (positive semidefinite) solution $\hat{\Theta}_T$ to the perturbed CARE (14) exists for all $\omega \in A_T$. Let $\Delta_T\Theta = \hat{\Theta}_T - \Theta$. We write

$$l(T)\,\mathrm{vec}(\Delta_T\Theta) = l(T)1_{A_T}\,\mathrm{vec}(\Delta_T\Theta) + l(T)1_{A_T^c}\,\mathrm{vec}(\Delta_T\Theta). \tag{39}$$

Since (17) implies (15), Theorem 3 implies $\mathrm{vec}(\Delta_T\Theta) \xrightarrow{\mathbb{P}} 0$. Moreover, we have

$$\mathbb{P}(l(T)1_{A_T^c} > \varepsilon) \le \mathbb{P}(A_T^c) \to 0$$

for every $\varepsilon > 0$. Thus the second term in (39) converges to zero in probability, and hence, by Slutsky's theorem, it suffices to consider the first term in (39). For this, we first observe that, by the proof of Theorem 17 in Sun (1998), we have

$$(B - C\Theta)^\top \Delta_T\Theta + \Delta_T\Theta(B - C\Theta) = -E + h_1(\Delta_T\Theta) + h_2(\Delta_T\Theta),$$

where

$$\begin{aligned} E_T &= \Delta_T D + \Delta_T B^\top \Theta + \Theta \Delta_T B - \Theta \Delta_T C \Theta \\ h_1(\Delta_T \Theta) &= -[(\Delta_T B - \Delta_T C\Theta)^\top \Delta_T \Theta + \Delta_T \Theta(\Delta_T B - \Delta_T C \Theta)] \\ h_2(\Delta_T\Theta) &= \Delta_T\Theta(C + \Delta_T C)\Delta_T\Theta. \end{aligned}$$

Recall the notation $\Phi = B - C\Theta$. Now, by compatibility of vectorization operator and Kronecker product we obtain

$$\begin{aligned} \operatorname{vec}\left((B - C\Theta)^\top \Delta_T\Theta + \Delta_T\Theta(B - C\Theta)\right) &= \operatorname{vec}(\Phi^\top \Delta_T\Theta + \Delta_T\Theta\Phi) = (I \otimes \Phi^\top)\operatorname{vec}(\Delta_T\Theta) + (\Phi^\top \otimes I)\operatorname{vec}(\Delta_T\Theta) \\ &= (I \otimes \Phi^\top + \Phi^\top \otimes I)\operatorname{vec}(\Delta_T\Theta) = (\Phi^\top \oplus \Phi^\top)\operatorname{vec}(\Delta_T\Theta), \end{aligned}$$

where $\oplus$ denotes the Kronecker sum. By Lemma 5 the matrix Φ is stable, i.e. all its eigenvalues are in the open left half-plane. Hence, by Lemma 20, $(\Phi^\top \oplus \Phi^\top)$ is invertible, and consequently

$$I(T)1_{A_T}\operatorname{vec}(\Delta_T\Theta) = I(T)1_{A_T}(\Phi^\top \oplus \Phi^\top)^{-1}\left(-\operatorname{vec}(E_T) + \operatorname{vec}(h_1(\Delta_T\Theta)) + \operatorname{vec}(h_2(\Delta_T\Theta))\right).$$

By compatibility of vectorization and Kronecker product, the terms of $I(T)1_{A_T}\operatorname{vec}(h_1(\Delta_T\Theta))$ are given by

$$\begin{aligned} I(T)1_{A_T}\operatorname{vec}(\Delta_T\Theta\Delta_T B) &= I(T)1_{A_T}(I \otimes \Delta_T\Theta)\operatorname{vec}(\Delta_T B) \\ I(T)1_{A_T}\operatorname{vec}(\Delta_T B^\top \Delta_T\Theta) &= I(T)1_{A_T}(\Delta_T\Theta \otimes I)\operatorname{vec}(\Delta_T B^\top) \\ I(T)1_{A_T}\operatorname{vec}(\Theta\Delta_T C\Delta_T\Theta) &= I(T)1_{A_T}(\Delta_T\Theta \otimes \Theta)\operatorname{vec}(\Delta_T C) \\ I(T)1_{A_T}\operatorname{vec}(\Delta_T\Theta\Delta_T C\Theta) &= I(T)1_{A_T}(\Theta \otimes \Delta_T\Theta)\operatorname{vec}(\Delta_T C). \end{aligned}$$

By item (1), Theorem 3, and Slutsky's theorem all these terms converge to zero in probability. Similarly, for the asymptotically dominant term of $I(T)1_{A_T}\operatorname{vec}(h_2(\Delta_T\Theta))$, we get

$$1_{A_T}\|\Delta_T\Theta C\Delta_T\Theta\| \leq 1_{A_T}\|\Delta_T\Theta\|^2\|C\|.$$

Now on the set A_T we have, by (32) and item (1), that

$$1_{A_T}\|\Delta_T\Theta\| = 1_{A_T}O(\|\Delta_T D\| + \|\Delta_T B\| + \|\Delta_T C\|).$$

Thus $I(T)1_{A_T}\operatorname{vec}(h_2(\Delta_T\Theta))$ converges to zero in probability as well, and it suffices to study asymptotics of

$$I(T)1_{A_T}(\Phi^\top \oplus \Phi^\top)^{-1}(-\operatorname{vec}(E_T)). \tag{40}$$

For this we write

$$\begin{aligned} &-I(T)1_{A_T}\operatorname{vec}(E_T) = I(T)1_{A_T}\operatorname{vec}\left(\Theta\Delta_T C\Theta - \Delta_T D - \Delta_T B^\top\Theta - \Theta\Delta_T B\right) \\ &= I(T)1_{A_T}\left((\Theta \otimes \Theta)\operatorname{vec}(\Delta_T C) - \operatorname{vec}(\Delta_T D) - (\Theta \otimes I)\operatorname{vec}(\Delta_T B^\top) - (I \otimes \Theta)\operatorname{vec}(\Delta_T B)\right). \end{aligned}$$

We define a linear function $f : \mathbb{R}^{3n^2} \to \mathbb{R}^{n^2}$ by

$$f(\mathrm{vec}(C,B,D)) = (\Theta \otimes \Theta)\,\mathrm{vec}(C) - \mathrm{vec}(D) - (\Theta \otimes I)\,\mathrm{vec}(B^\top) - (I \otimes \Theta)\,\mathrm{vec}(B). \tag{41}$$

Then the usual Delta method and Slutsky's theorem implies

$$-l(T)1_{A_T}\,\mathrm{vec}(E_T) = l(T)1_{A_T}\left(f(\mathrm{vec}(\hat{C}_T,\hat{B}_T,\hat{D}_T)) - f(\mathrm{vec}(C,B,D)\right) \xrightarrow{\text{law}} L_2^\star L_1(X),$$

where the linear mapping $L_2^\star$ is given by the Jacobian (i.e. matrix representation) of the function f defined in (41). It remains to apply continuous mapping theorem to (40) to conclude that

$$l(T)1_{A_T}\,\mathrm{vec}(\Delta_T\Theta) \xrightarrow{\text{law}} (\Phi^\top \oplus \Phi^\top)^{-1} L_2^\star L_1(X) =: L_2 L_1(X).$$

This completes the proof.

A.4 | Representation of cross-covariance matrix $\gamma(t)$

Our main results relies on the cross-covariance matrix $\gamma(s)$ of U. However, in many models one assumes that only the variance of the noise G is known. In this subsection we give several representations for $\gamma(s)$ in terms of the variance matrix $v(t) = \mathbb{E}(G_t G_t^\top)$, provided that G has independent components. In particular, then $\mathbb{E}(G_t G_s^\top)$ is a diagonal matrix for all t, s, and satisfies

$$\mathbb{E}(G_t G_s^\top) = \frac{1}{2}(v(t) + v(s) - v(t-s)).$$

Our first representation is the following.

Lemma 21 *Let G have independent components. Then*

$$\begin{aligned}\gamma(r) = \frac{e^{-\Theta r}\Theta}{2}&\left(\int_{-\infty}^{r} e^{\Theta x} v(x)dx - \int_{-\infty}^{r}\int_{-\infty}^{0} e^{\Theta x}e^{\Theta s}v(x)e^{\Theta s}\Theta ds dx - \int_{r}^{\infty}\int_{-\infty}^{r-x} e^{\Theta x}e^{\Theta s}v(x)e^{\Theta s}\Theta ds dx\right)\\ &+\frac{1}{2}\int_{r}^{\infty} v(x)e^{\Theta(r-x)}\Theta dx - \frac{1}{2}v(r).\end{aligned} \tag{42}$$

Proof *Using representation $U_t = G_t - e^{-\Theta t}\Theta\int_{-\infty}^{t} e^{\Theta s}G_s ds$ gives us*

$$U_r U_0^\top = -\int_{-\infty}^{0} G_r G_s^\top e^{\Theta s}\Theta ds + e^{-\Theta r}\Theta\int_{-\infty}^{0}\int_{-\infty}^{r} e^{\Theta u}G_u G_s^\top e^{\Theta s}\Theta du ds.$$

Taking expectation and using Fubini's theorem thus yields

$$\gamma(r) = -\frac{1}{2}\int_{-\infty}^{0}(v(r)+v(s)-v(r-s))e^{\Theta s}\Theta ds + e^{-\Theta r}\frac{\Theta}{2}\int_{-\infty}^{0}\int_{-\infty}^{r} e^{\Theta u}(v(u)+v(s)-v(u-s))e^{\Theta s}\Theta du ds.$$

Here

$$e^{-\Theta r}\frac{\Theta}{2}\int_{-\infty}^{0}\int_{-\infty}^{r}e^{\Theta u}v(s)e^{\Theta s}\Theta du ds = e^{-\Theta r}\frac{1}{2}\int_{-\infty}^{0}e^{\Theta r}v(s)e^{\Theta s}\Theta ds = \frac{1}{2}\int_{-\infty}^{0}v(s)e^{\Theta s}\Theta ds$$

leading to

$$\gamma(r) = -\frac{1}{2}\int_{-\infty}^{0}(v(r)-v(r-s))e^{\Theta s}\Theta ds + e^{-\Theta r}\frac{\Theta}{2}\int_{-\infty}^{0}\int_{-\infty}^{r}e^{\Theta u}(v(u)-v(u-s))e^{\Theta s}\Theta du ds.$$

By the change of variable $x = u - s$, we get

$$\begin{aligned}\int_{-\infty}^{0}\int_{-\infty}^{r}e^{\Theta u}v(u-s)e^{\Theta s}\Theta du ds &= \int_{-\infty}^{0}\int_{-\infty}^{r-s}e^{\Theta(x+s)}v(x)e^{\Theta s}\Theta dx ds = \int_{-\infty}^{\infty}\int_{-\infty}^{\min\{0,r-x\}}e^{\Theta x}e^{\Theta s}v(x)e^{\Theta s}\Theta ds dx\\ &= \int_{-\infty}^{r}\int_{-\infty}^{0}e^{\Theta x}e^{\Theta s}v(x)e^{\Theta s}\Theta ds dx + \int_{r}^{\infty}\int_{-\infty}^{r-x}e^{\Theta x}e^{\Theta s}v(x)e^{\Theta s}\Theta ds dx.\end{aligned}$$

Finally, we have the identities

$$\int_{-\infty}^{0}v(r-s)e^{\Theta s}\Theta ds = \int_{r}^{\infty}v(x)e^{\Theta(r-x)}\Theta dx$$
$$\int_{-\infty}^{0}\int_{-\infty}^{r}e^{\Theta u}v(u)e^{\Theta s}\Theta du ds = \int_{-\infty}^{r}e^{\Theta x}v(x)dx$$
$$\int_{-\infty}^{0}v(r)e^{\Theta s}\Theta ds = v(r).$$

Combining all the results above gives us (42).

Lemma 22 *Let G have independent components. Then*

$$\gamma(r) = \frac{\Theta}{2}\int_{-\infty}^{0}\int_{-\infty}^{0}e^{\Theta x}\left(v(x+r)-v(r)+v(r-s)-v(x+r-s)\right)e^{\Theta s}\Theta dx ds. \tag{43}$$

Proof *The expression*

$$\gamma(r) = -\frac{1}{2}\int_{-\infty}^{0}(v(r)-v(r-s))e^{\Theta s}\Theta ds + e^{-\Theta r}\frac{\Theta}{2}\int_{-\infty}^{0}\int_{-\infty}^{r}e^{\Theta u}(v(u)-v(u-s))e^{\Theta s}\Theta du ds$$

can be written as

$$\gamma(r) = -\frac{1}{2}\int_{-\infty}^{0}e^{-\Theta r}\Theta\int_{-\infty}^{r}e^{\Theta u}(v(r)-v(u)-v(r-s)+v(u-s))e^{\Theta s}\Theta du ds,$$

from which the claim follows by the change of variable $u - r = x$.

Remark 14 *If the matrices $v(t)$ and Θ commute for every $t \in \mathbb{R}$, we obtain even simpler expression*

$$\gamma(r) = \frac{\Theta}{4}\int_{-\infty}^{0}e^{\Theta x}\left(v(x+r)+v(r-x)-2v(r)\right)dx.$$

In particular, this is the case if G consists of independent processes with equal variances.

A.5 | Proofs related to Gaussian examples

This section is devoted to the proofs of Proposition 7, Theorem 8, and Corollary 9. We begin with the proof of Proposition 7.

Proof of Proposition 7 We have

$$\hat{\gamma}_{T,i,j}(s) - \gamma_{i,j}(s) = \frac{1}{T}\int_0^T \left(U_{r+s}^{(i)}U_r^{(j)} - \mathbb{E}\left(U_s^{(i)}U_0^{(j)}\right)\right) dr$$

implying, with straightforward computations, that

$$\mathbb{E}\left[\hat{\gamma}_{T,i,j}(s) - \gamma_{i,j}(s)\right]^2 \leq \frac{2}{T}\int_0^T |\gamma_{i,j}(r+s)|^2 dr.$$

Now $\|\gamma(r)\| \to 0$ implies $|\gamma_{i,j}(r)| \to 0$ as well, and thus, for each fixed s,

$$|\hat{\gamma}_{T,i,j}(s) - \gamma_{i,j}(s)| \to 0$$

in L^2. These further implies

$$\sup_{s\in[0,t]} \mathbb{E}\|\hat{\gamma}_T(s) - \gamma(s)\|^2 \to 0,$$

from which we conclude that conditions of Remark 5 are satisfied. The claim then follows.

In order to prove Theorem 8 one needs to prove the convergence of finite dimensional distributions and tightness. For the latter we present the following result that might be interesting on its own. In the sequel, we use the short notation

$$F_T(\tau) = l(T)\,\mathrm{vec}(\hat{\gamma}_T(\tau) - \gamma(\tau)).$$

Proposition 23 *Suppose that $\gamma(r)$ is locally absolutely continuous and*

$$\max\left(\|\gamma'(r)\|, \|\gamma(r)\|\right) \leq h(r)$$

for some non-increasing function $h(r)$ such that, for some $K > 0$, we have $h(r) \in L^1([0,K])$ and

$$\int_K^T h(r)^2 dr = O\left(\frac{T}{l(T)^2}\right), \quad T > K.$$

Then there exists T_0 such that for all $\tau, s \in [0,t]$, all $a \in \mathbb{R}^{n^2}$ and all $p \geq 2$, we have

$$\mathbb{E}\left|a^\top(F_T(\tau) - F_T(s))\right|^p \leq c|\tau - s|^{\frac{p}{2}}, \quad T \geq T_0,$$

where c depends only on p, t, γ, and $a^\top$.

Proof *We have*

$$F_T(\tau) - F_T(s) = \frac{l(T)}{T} \int_0^T \mathrm{vec}(U_{u+\tau}U_u^\top - \gamma(\tau) - U_{u+s}U_u^\top + \gamma(s))du.$$

First we note that it suffices to prove the claim only for $p = 2$. Indeed, since U is Gaussian, the expression $a^\top(F_T(\tau) - F_T(s))$ belongs to the so-called second Wiener chaos (for details, see e.g. Janson (1997)), implying the hypercontractivity property

$$\mathbb{E}|a^\top(F_T(\tau) - F_T(s))|^p \le c_p \left[\mathbb{E}|a^\top(F_T(\tau) - F_T(s))|^2\right]^{\frac{p}{2}}.$$

Thus, let $p = 2$. We have

$$|a^\top(F_T(\tau) - F_T(s))|^2 \le c_a \|F_T(\tau) - F_T(s)\|^2,$$

where

$$\|F_T(\tau) - F_T(s)\|^2 \le \frac{l(T)^2}{T^2} \sum_{i,j=1}^{n} \left(\int_0^T \left(U_{u+\tau}U_u^\top - \gamma(\tau) - U_{u+s}U_u^\top + \gamma(s)\right)_{i,j} du\right)^2.$$

Here

$$\begin{aligned}\left(\int_0^T (U_{u+\tau}U_u^\top - \gamma(\tau) - U_{u+s}U_u^\top + \gamma(s))_{i,j} du\right)^2 &= \int_0^T U_{u+\tau}^{(i)}U_u^{(j)} - \gamma_{i,j}(\tau) - U_{u+s}^{(i)}U_u^{(j)} + \gamma_{i,j}(s)du \\ &\times \int_0^T U_{v+\tau}^{(i)}U_v^{(j)} - \gamma_{i,j}(\tau) - U_{v+s}^{(i)}U_v^{(j)} + \gamma_{i,j}(s)dv.\end{aligned}$$

Taking expectation and with some straightforward computations, we get

$$\begin{aligned}&\mathbb{E}\left(\int_0^T (U_{u+\tau}U_u^\top - \gamma(\tau) - U_{u+s}U_u^\top + \gamma(s))_{i,j} du\right)^2 = \int_0^T (T-x)\gamma_{j,j,}(x)(\gamma_{i,i}(x) - \gamma_{i,i}(x+\tau-s))dx \\ &+ \int_0^T (T-x)\gamma_{j,j}(x)(\gamma_{i,i}(x) - \gamma_{i,i}(-x+\tau-s))dx + \int_0^T (T-x)\gamma_{j,j,}(x)(\gamma_{i,i}(x) - \gamma_{i,i}(x+s-\tau))dx \\ &+ \int_0^T (T-x)\gamma_{j,j}(x)(\gamma_{i,i}(x) - \gamma_{i,i}(-x+s-\tau))dx + \int_0^T (T-x)\gamma_{i,j}(x+\tau)(\gamma_{i,j}(-x+\tau) - \gamma_{i,j}(-x+s))dx \\ &+ \int_0^T (T-x)\gamma_{i,j}(-x+\tau)(\gamma_{i,j}(x+\tau) - \gamma_{i,j}(x+s))dx + \int_0^T (T-x)\gamma_{i,j}(x+s)(\gamma_{i,j}(-x+s) - \gamma_{i,j}(-x+\tau))dx \\ &+ \int_0^T (T-x)\gamma_{i,j}(-x+s)(\gamma_{i,j}(x+s) - \gamma_{i,j}(x+\tau))dx.\end{aligned}$$

Thus it suffices to show that all eight terms, when multiplied with $l(T)^2/T^2$, admit a bound of the form $C|\tau - s|$. We show how the first term can be treated, while the rest can be shown with similar arguments. Without loss of generality, let $s < \tau$ and denote $u = \tau - s$. For the first term above, we apply the mean value theorem to obtain

$$\left|\int_0^T (T-x)\gamma_{j,j,}(x)(\gamma_{i,i}(x) - \gamma_{i,i}(x+u))dx\right| \le Tu\left(\int_0^K |\gamma_{j,j}(x)| \operatorname*{ess\,sup}_{x<\xi<x+u} |\gamma_{i,i}'(\xi)|dx + \int_K^T |\gamma_{j,j}(x)| \operatorname*{ess\,sup}_{x<\xi<x+u} |\gamma_{i,i}'(\xi)|dx\right).$$

By assumption,

$$\operatorname*{ess\,sup}_{x<\xi<x+u} |\gamma'_{i,i}(\xi)| \leq \operatorname*{ess\,sup}_{x<\xi<x+u} \|\gamma'(\xi)\| \leq h(x).$$

Since $|\gamma_{j,j}(x)|$ is bounded and $h(x) \in L^1([0,K])$, we get

$$\frac{l(T)^2}{T^2} \cdot Tu \int_0^K |\gamma_{j,j}(x)| \operatorname*{ess\,sup}_{x<\xi<x+u} |\gamma'_{i,i}(\xi)| dx \leq cu \frac{l(T)^2}{T}.$$

Now, the best possible rate $l(T)$ that one can have is $\sqrt{T}$, giving $\sup_{T\geq T_0} \frac{l(T)^2}{T} < \infty$. Similarly, we have

$$\int_K^T |\gamma_{j,j}(x)| \operatorname*{ess\,sup}_{x<\xi<x+u} |\gamma'_{i,i}(\xi)| dx \leq c \int_K^T h(x)^2 dx \leq c \frac{T}{l(T)^2}$$

by assumption. Treating the rest of the terms similarly concludes the proof.

The proof of Theorem 8 is now rather straightforward.

Proof of Theorem 8 By Cramer-Wold device it suffices to prove the convergence of linear combinations, and then the tightness follows from Proposition 23. In order to obtain convergence of multidimensional distributions, we have to prove that all the combinations of the form

$$\sqrt{T} \sum_{k=1}^{d} a_k^\top \operatorname{vec}(\hat{\gamma}_T(\tau_k) - \gamma(\tau_k))$$

converges towards $\sum_{k=1}^d a_k^\top X_{\tau_k}$, where a_k are some n^2-dimensional vectors. Now the above expression can be written as

$$\frac{1}{\sqrt{T}} \int_0^T \sum_{k=1}^{d} \sum_{i,j=1}^{n} a_k^{(i,j)} \left[U_{r+\tau_k}^{(i)} U_r^{(j)} - \gamma_{i,j}(\tau_k) \right] dr,$$

and thus the convergence towards a Gaussian limit follows from the continuous time Breuer-Major Theorem (see, e.g. Campese et al. (2020) and references therein) together with the fact that now $\int_0^\infty \|\gamma(s)\|^2 ds < \infty$. This completes the proof.

Finally, we verify the result for multidimensional fractional Ornstein-Uhlenbeck process.

Proof of Corollary 9 Let $H_{max} = \max_{1\leq i\leq n} H_i$ and $H_{min} = \min_{1\leq i\leq n} H_i$. We prove that asymptotically, as $t \to \infty$, we have

$$\|\gamma(t)\| = O\left(t^{2H_{max}-2}\right), \tag{44}$$

$$\|\gamma'(t)\| = O\left(t^{2H_{max}-2}\right), \tag{45}$$

and that $\|\gamma'(t)\| = O\left(\max\left(t^{2H_{min}-1}, 1\right)\right)$ for $t \leq K$ with K fixed. Since $H_{max} < 3/4$ and $H_{min} > 0$, the statement then follows from Theorem 8. For this, let g be an auxiliary function such that $g(r)/r \to 0$ and $\log r/g(r) \to 0$. Then it follows that $r^{4-2H_{max}}e^{-cg(r)} \to 0$ for all $c > 0$. We begin by showing (44). We divide the expression (43) for $\gamma(r)$ into

$$\frac{\Theta}{2}\int_{-\infty}^{-g(r)}\int_{-\infty}^{-g(r)} e^{\Theta x}\left(v(x+r)-v(r)+v(r-s)-v(x+r-s)\right)e^{\Theta s}\Theta dxds \tag{46}$$

$$\frac{\Theta}{2}\int_{-g(r)}^{0}\int_{-g(r)}^{0} e^{\Theta x}\left(v(x+r)-v(r)+v(r-s)-v(x+r-s)\right)e^{\Theta s}\Theta dxds. \tag{47}$$

$$\frac{\Theta}{2}\int_{-g(r)}^{0}\int_{-\infty}^{-g(r)} e^{\Theta x}\left(v(x+r)-v(r)+v(r-s)-v(x+r-s)\right)e^{\Theta s}\Theta dxds. \tag{48}$$

$$\frac{\Theta}{2}\int_{-\infty}^{-g(r)}\int_{-g(r)}^{0} e^{\Theta x}\left(v(x+r)-v(r)+v(r-s)-v(x+r-s)\right)e^{\Theta s}\Theta dxds. \tag{49}$$

Since r is large, we obtain for (47) that

$$\begin{aligned}
v_{i,i}(x+r)-v_{i,i}(r)+v_{i,i}(r-s)-v_{i,i}(x+r-s) &= |x+r|^{2H_i}-r^{2H_i}+|r-s|^{2H_i}-|x+r-s|^{2H_i}\\
&= r^{2H_i}((1+\frac{x}{r})^{2H_i}+(1-\frac{s}{r})^{2H_i}-1-(1+\frac{x-s}{r})^{2H_i})\\
&= r^{2H_i}\left(O\left(\left(\frac{x}{r}\right)^2\right)+O\left(\left(\frac{s}{r}\right)^2\right)\right), \quad \text{when } H_i \neq \frac{1}{2}
\end{aligned}$$

and

$$v_{i,i}(x+r)-v_{i,i}(r)+v_{i,i}(r-s)-v_{i,i}(x+r-s)=0, \quad \text{when } H_i = \frac{1}{2}.$$

Hence

$$\begin{aligned}
&\frac{\Theta}{2}\int_{-g(r)}^{0}\int_{-g(r)}^{0} e^{\Theta x}\left(v(x+r)-v(r)+v(r-s)-v(x+r-s)\right)e^{\Theta s}\Theta dxds\\
=&\frac{\Theta}{2}r^{2H_{max}}\int_{-g(r)}^{0}\int_{-g(r)}^{0} e^{\Theta x}\left(O\left(\left(\frac{x}{r}\right)^2\right)+O\left(\left(\frac{s}{r}\right)^2\right)\right)\operatorname{diag}\left(1_{H_i\neq\frac{1}{2}}r^{2(H_i-H_{max})}\right)e^{\Theta s}\Theta dxds,
\end{aligned}$$

and taking the norm gives

$$\begin{aligned}
&\left\|\frac{\Theta}{2}\int_{-g(r)}^{0}\int_{-g(r)}^{0} e^{\Theta x}\left(v(x+r)-v(r)+v(r-s)-v(x+r-s)\right)e^{\Theta s}\Theta dxds\right\|\\
&\leq Cr^{2H_{max}-2}\int_{-g(r)}^{0}\int_{-g(r)}^{0} e^{\lambda_{min}(x+s)}(x^2+s^2)dxds = O(r^{2H_{max}-2}).
\end{aligned}$$

Next, we study (46). We have four terms from which one satisfies

$$\begin{aligned}
\left\|\int_{-\infty}^{-g(r)}\int_{-\infty}^{-g(r)} e^{\Theta x}v(r-s)e^{\Theta s}\Theta dxds\right\| &\leq Ce^{-\lambda_{min}g(r)}\int_{-\infty}^{-g(r)} e^{\lambda_{min}s}(r-s+1)^2ds\\
&= O\left(e^{-\lambda_{min}g(r)}r^2\right)=O\left(r^{2H_{max}-2}\right)
\end{aligned} \tag{50}$$

by our choice of $g(r)$. By utilizing the exponential decay similarly with the remaining three terms, and with (48) and (49), proves (44). Let us next consider the derivative $\gamma'(r)$. We first observe that, by representation (42), the matrix

$\gamma(r)$ is continuously differentiable except at the origin, and we have

$$\|\gamma'(r)\| = O\left(\max\left(r^{2H_{min}-1}, 1\right)\right)$$

as $r \to 0$. To prove (45), by a standard application of dominated convergence theorem, we express $\gamma'(r)$ as the sum of

$$\frac{\Theta}{2}\int_{-\infty}^{-g(r)}\int_{-\infty}^{-g(r)} e^{\Theta x}\left(v'(x+r) - v'(r) + v'(r-s) - v'(x+r-s)\right)e^{\Theta s}\Theta dx ds$$
$$\frac{\Theta}{2}\int_{-g(r)}^{0}\int_{-g(r)}^{0} e^{\Theta x}\left(v'(x+r) - v'(r) + v'(r-s) - v'(x+r-s)\right)e^{\Theta s}\Theta dx ds.$$
$$\frac{\Theta}{2}\int_{-g(r)}^{0}\int_{-\infty}^{-g(r)} e^{\Theta x}\left(v'(x+r) - v'(r) + v'(r-s) - v'(x+r-s)\right)e^{\Theta s}\Theta dx ds$$
$$\frac{\Theta}{2}\int_{-\infty}^{-g(r)}\int_{-g(r)}^{0} e^{\Theta x}\left(v'(x+r) - v'(r) + v'(r-s) - v'(x+r-s)\right)e^{\Theta s}\Theta dx ds$$

Now all terms can be treated similarly as (46), (47), (48) and (49), which concludes the proof.