STATIONARY TANGENT: THE DISCRETE AND NON-SMOOTH CASE

U. KEICH

Abstract. In [5] we define a stationary tangent process, or a locally optimal stationary approximation, to a real non-stationary smooth Gaussian process. This paper extends the idea by constructing a discrete tangent – a "locally" optimal stationary approximation – for a discrete time, real Gaussian process. Analogously to the smooth case, our construction relies on a generalization of the recursion formula for the orthogonal polynomials of the spectral distribution function. More precisely, we use a generalization of the Schur parameters to identify the stationary tangent. By way of discretizing, we later demonstrate how this tangent can be used to obtain "good" local stationary approximations to non-smooth continuous time, real Gaussian processes. Further, we demonstrate how, analogously to the curvatures in the smooth case, the Schur parameters can be used to determine the order of stationarity of a non-smooth process.

1. Introduction

The immediate context of this paper is its predecessor [5] in which we construct a locally optimal stationary approximation to a real non-stationary smooth Gaussian process. In the case where this construction is unique we name it the stationary tangent process at $t = t_0$. This paper offers an extension of this construction to discrete and non-smooth normal processes. The wider context of both papers is the area of so-called "locally stationary processes". Intuitively, these are *non*-stationary processes that on a sufficiently small time scale do not deviate considerably from stationarity. To mention just a couple of approaches to the subject: Priestley studied this problem through what he defines as the *evolutionary spectrum* [9, p. 148] and later he [10] – as well as Mallat, Papanicolaou & Zhang [7] – looked at the problem from wavelets point of view. These locally stationary processes, or the closely related processes with evolutionary spectral representation,

Key words and phrases. stationary Gaussian processes, Schur parameters, stationary tangent, locally stationary processes, trigonometric moment problem, curvature functions.

naturally show up in the study of statistical inference, e.g. [2] and [8]. In this paper we suggest a possible alternative path for defining locally stationary processes. Moreover, one may be able to utilize curvature functions and their discrete analogues, generalized Schur parameters, in the context of statistical inference.

Let $\mathbb{X}$ be a discrete time process (all processes in this paper are assumed to be 0-mean real Gaussian process). We say that the stationary process $\hat{\mathbb{X}}$ is a *locally optimal stationary approximation* to $\mathbb{X}$, at $t = 0$, if it is jointly Gaussian with $\mathbb{X}$[1] and it sequentially minimizes $\mathrm{E}\left|\mathbb{X}_k - \hat{\mathbb{X}}_k\right|^2$ for $k = 0, 1, \ldots$. More precisely, consider the following decreasing sets of processes:

$$A_0 \stackrel{d}{=} \left\{\hat{\mathbb{X}} : \hat{\mathbb{X}} \text{ is stationary and } \mathrm{E}\left|\mathbb{X}_0 - \hat{\mathbb{X}}_0\right|^2 = 0\right\}$$

$$A_k \stackrel{d}{=} \left\{\hat{\mathbb{X}} \in A_{k-1} : \mathrm{E}\left|\mathbb{X}_k - \hat{\mathbb{X}}_k\right|^2 = \min_{\mathbb{Y} \in A_{k-1}} \mathrm{E}\left|\mathbb{X}_k - \mathbb{Y}_k\right|^2\right\}.$$

Then, $\hat{\mathbb{X}}$ is a locally optimal stationary approximation if $\hat{\mathbb{X}} \in \cap_n A_n$. Our main result (Theorem 1) is that the stationary tangent we define is the unique locally optimal stationary approximation. Note that as the optimal approximation is defined sequentially, it suffices to consider a finite time process $\mathbb{X}_0, \mathbb{X}_1, \ldots, \mathbb{X}_n$, which is the main object of interest in this paper.

In [5] we show that the tangent, at $t = 0$, to a smooth process $\mathbb{X}$ sequentially minimizes $\mathrm{E}\left|\mathbb{X}_0^{(k)} - \hat{X}_0^{(k)}\right|^2$ for $k = 0, 1, \ldots$. In this paper we prove that the same result holds for the discrete tangent process provided we replace differentiation with asymmetric differences (Cor. 2).

In the smooth case our construction of the tangent relied on curvature functions. In the stationary case these curvatures are exactly the coefficients in the 3-term recursion formula for the orthogonal polynomials of the associated spectral distribution function (defined on $\mathbb{R}$). We define the discrete stationary tangent in an analogous way. Namely, we define it using a generalization of the Schur parameters to non-stationary (or, non-Toeplitz) correlation. In the stationary case, these Schur parameters can be described as the coefficients in the recursion formula for the orthogonal polynomials of any associated spectral distribution function (now defined on the circle $\mathbb{T}$).

In section 4 we provide an example where by way of discretizing we obtain a "tangent" stationary process to the Brownian motion at $t = 1$.

[1] This somewhat technical condition facilitates the analysis that follows.

This process turns out to be an Ornstein-Uhlenbeck process (Claim 4.1). As we show, this Ornstein-Uhlenbeck process is an optimal stationary approximation to first order order, that is, $\mathrm{E}\left|b_{t+1} - \tilde{\mathbb{X}}_t\right|^2 = t^2/4 + \mathrm{O}(t^3)$ and this is the best that can be done. We compare it with a couple of other stationary approximations and show that it is not the optimal one in this local mean square sense. It is likely that there is no optimal approximation in this sense. This leads us to a more general open question: for which processes does this discretizing procedure produce a stationary approximation, how good is this approximation, and does there exist a, possibly other, optimal stationary approximation in the local mean square sense?

Section 6 shows the relation between the curvatures and the Schur parameters runs deeper than previously mentioned. Indeed, the curvatures of a smooth correlation R can be obtained as a simple limit of the Schur parameters of discrete samplings of R. This was established in [5] for δ-curvatures which are a discrete variant of the curvatures based on a symmetric rather than the assyemtric scheme used here to define the generalized Schur parameters.

We conclude with an example which demonstrates how the Schur parameters can be used to determine the order of stationarity d of a process $\mathbb{X}$. The latter is essentially an optimal bound on the order of any stationary approximation to $\mathbb{X}$ at $t = t_0$. In the smooth case, d is determined by, roughly, the rate of change of the curvatures at t_0. Analogously, we show how in our example, the rate of change in the relevant Schur parameters determines d. Once again, the more general question of the order of stationarity of non-smooth processes and its connection with the Schur parameters is still open.

2. The Schur parameters and their generalization

In this section we provide a brief and somewhat restricted introduction to the subject of Schur parameters, more details can be found, for example, in [6], [3] and [1]. Let $\mathbb{T}$ be the unit circle in the complex plane and let σ be a positive finite measure on $\mathbb{T}$. Since we are interested in real processes we assume that σ is symmetric, i.e., we assume that the positive-definite sequence $\hat{\sigma}(k) \stackrel{d}{=} \int_{\mathbb{T}} z^k \, d\sigma = \int_0^{2\pi} e^{ik\omega} \, d\sigma(\omega)$ is symmetric, that is, $\hat{\sigma}(k) = \hat{\sigma}(-k)$. One can apply the Gram-Schmidt procedure to the powers $\{1, z, z^2, \ldots, z^n\}$ to obtain the orthonormal (trigonometric) polynomials $Q_0(z), Q_1(z), \ldots, Q_n(z)$, where Q_k is of degree k and $\langle Q_k, z^k \rangle_{L^2(\sigma)} > 0$. Since $\hat{\sigma}(k) = \langle z^{k+j}, z^j \rangle_{L^2(\sigma)}$, this procedure is completely determined by $\hat{\sigma}(k)$ for $k = 0, \ldots, n$. Thus, by identifying $R_{ij} = R(i-j) = \hat{\sigma}(i-j)$ we arrive at a perfect equivalence between stationary, or Toeplitz covariance matrices of order $(n+1)\times(n+1)$ and orthonormal polynomials of degree $\le n$. The latter in turn, can be completely described in terms of the following recursion formula they satisfy:

$$\begin{aligned}\sqrt{1-s_{k+1}^2}Q_{k+1}(z) &= zQ_k(z) - s_{k+1}Q_k^*(z)\\ \sqrt{1-s_{k+1}^2}Q_{k+1}^*(z) &= Q_k^*(z) - s_{k+1}zQ_k(z),\end{aligned} \tag{1}$$

where for a polynomial P of degree n, $P^*(z) \overset{d}{=} z^n P(1/z)$, and $Q_0 \equiv Q_0^* \equiv 1/\sqrt{R(0)}$. Clearly, the recursion formula is completely determined in terms of the coefficients s_k which are called the Schur parameters of the positive definite sequence R. One can show that for $k \ge 1$, $|s_k| \le 1$ and $|s_k| = 1$ if and only if the rank of the associated Toeplitz matrix is k. Thus, (1) establishes a perfect equivalence between positive definite, $n\times n$, Toeplitz matrices of rank n, and their n Schur parameters ($s_0 \overset{d}{=} \sqrt{R(0)}$).

Remark 2.1. It can be shown that $zQ_k(z)$ is the normalized coprojection of z^{k+1} on $W_k \overset{d}{=} \text{Span}\{z, z^2, \dots, z^k\}$, and that Q_{k+1}^* is the normalized coprojection of 1 on W_k. In particular, it follows from (1) that s_k is the correlation between the forward and backward prediction errors of length k [6]. Geometrically, the last statement amounts to the fact that s_k is the cosine of the angle between the coprojections of z^{k+1} and 1 on W_k. This is the key to generalizing the Schur parameters.

In order to generalize the notion of Schur parameters to the non-stationary case we need to recast (1) in a more general setting. Let R be an $n\times n$ covariance matrix and let A be any root of R, i.e., $R = AA^*$. Let $\boldsymbol{x}_k \in \mathbb{R}^n$ be the kth row of A. In analogy with the smooth case [5, Sec. 2], we say that $\boldsymbol{x}$ is a (discrete) *curve* that is associated with R. Clearly, $R(i,j) = \langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle$ where $\langle,\rangle$ is the standard inner product in $\mathbb{R}^n$. Note that in fact we associate with R an equivalence class of curves, where each class is determined up to an orthogonal transformation of $\mathbb{R}^n$. We say that the curve $\boldsymbol{x}$ is stationary if $\langle \boldsymbol{x}_j, \boldsymbol{x}_k\rangle$ is a function of $j-k$, or equivalently, if it is associated with a stationary correlation.

Let R be an $(n+1)\times(n+1)$ stationary covariance matrix and let $\boldsymbol{x}$ be an associated curve. Let $H_n \overset{d}{=} \text{Span}\{1, z, z^2, \dots, z^n\}$ be the space of trigonometric polynomials of degree $\le n$ equipped, as before, with the inner product $\langle z^i, z^j\rangle \overset{d}{=} R(i-j)$. The map defined by $z^k \mapsto \boldsymbol{x}_k$ is clearly an isometry between H_n and $\mathbb{R}^{n+1}$. Let $\boldsymbol{u}_0, \boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ be the result of the Gram-Schmidt procedure applied to $\boldsymbol{x}_0, \dots, \boldsymbol{x}_n$ (subject to $\langle \boldsymbol{x}_k, \boldsymbol{u}_k\rangle > 0$). Let $\boldsymbol{u}_k^*$ be the normalized coprojection of $\boldsymbol{x}_0$ on

$W_k \stackrel{d}{=} \text{Span}\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$ and let $\boldsymbol{u}_k(1)$ be the normalized coprojection of $\boldsymbol{x}_{k+1}$ on W_k. According to remark 2.1 our isometry translates (1) into,

$$\begin{aligned} \sqrt{1-s_{k+1}^2}\,\boldsymbol{u}_{k+1} &= \boldsymbol{u}_k(1) - s_{k+1}\boldsymbol{u}_k^* \\ \sqrt{1-s_{k+1}^2}\,\boldsymbol{u}_{k+1}^* &= \boldsymbol{u}_k^* - s_{k+1}\boldsymbol{u}_k(1), \end{aligned} \tag{2}$$

and $\boldsymbol{u}_0 = \boldsymbol{u}_0^* = \boldsymbol{x}_0/s_0$.

The latter recursion can be generalized to the non-stationary case: given an $(n+1)\times(n+1)$ covariance matrix R, let $\boldsymbol{x}$, $\boldsymbol{u}$, $\boldsymbol{u}^*$ and $\boldsymbol{u}(1)$ be as above.

Definition 2.2. The (generalized) Schur parameters of R are $s_0 \stackrel{d}{=} \sqrt{R(0)}$, and for $k \geq 1$, $s_k \stackrel{d}{=} \langle \boldsymbol{u}_{k-1}(1), \boldsymbol{u}_{k-1}^* \rangle$.

Remark.
- Note that here we define these generalized parameters at $t = 0$. Clearly, they will vary with t for a non-stationary process.
- Recall that if $\boldsymbol{y}$ is another curve with covariance R, then there exists an orthogonal map of $\mathbb{R}^{n+1}$, U such that, $\boldsymbol{y} = U\boldsymbol{x}$. It follows that the s_k are well defined. In section 5 we provide an expression for s_k directly in terms of R.
- It is not hard to see that s_k is the correlation between $\mathbb{X}_0$ and $\mathbb{X}_k$ given $\mathbb{X}_1, \dots, \mathbb{X}_{k-1}$.

Claim 2.3. The recursion described in (2) holds for non-stationary curves as well.

Proof. Let $H_k \stackrel{d}{=} \text{Span}\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$. Clearly, there exist $\alpha_k \in \mathbb{R}$ such that $P_{H_k}(\boldsymbol{x}_{k+1}) = P_{W_k}(\boldsymbol{x}_{k+1}) + \alpha_k \boldsymbol{u}_k^*$, where P_H stands for the orthogonal projection on H. Therefore,

$$\boldsymbol{x}_{k+1} - P_{H_k}(\boldsymbol{x}_{k+1}) = \boldsymbol{x}_{k+1} - P_{W_k}(\boldsymbol{x}_{k+1}) - \alpha_k \boldsymbol{u}_k^*,$$

which implies that for some $\beta_k, \gamma_k \in \mathbb{R}$, $\beta_k \boldsymbol{u}_{k+1} = \boldsymbol{u}_k(1) - \gamma_k \boldsymbol{u}_k^*$. Observing that $\|\boldsymbol{u}_k\| = 1$, that $\langle \boldsymbol{u}_k^*, \boldsymbol{u}_{k+1} \rangle = 0$, and that $s_{k+1} = \langle \boldsymbol{u}_k^*, \boldsymbol{u}_k(1) \rangle$ by definition, completes the proof of the first equality. The second one is proved analogously. ☐

Remark.
- We stress the fact that we consider $t = 0$. For a non-stationary processs and $t > 0$, (2) will still hold but s_k will vary.
- As in the stationary case, (2) provides a recipe for computing the Schur parameters.
- For a somewhat different and more detailed approach to generalized Schur parameters consult [4].

3. The stationary tangent

Suppose that $\mathbb{X}_k$ and $\hat{\mathbb{X}}_k$, defined for $k = 0 \dots n$, are jointly Gaussian. Then, there exist curves $\boldsymbol{x}, \hat{\boldsymbol{x}} \in \mathbb{R}^{n+1}$ and a vector $\boldsymbol{\xi}$ of $n+1$ independent standard normal random variables such that, $\mathbb{X}_k = \langle \boldsymbol{\xi}, \boldsymbol{x}_k \rangle$, and $\hat{\mathbb{X}}_k = \langle \boldsymbol{\xi}, \hat{\boldsymbol{x}}_k \rangle$. Since $\mathrm{E}\left|\mathbb{X}_k - \hat{\mathbb{X}}_k\right|^2 = \|\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k\|^2$, we can look for our locally optimal stationary process approximation by analyzing the analogue question for curves. Thus, we first define the stationary tangent curve. In order to do that we need the following claim.

In what follows $\boldsymbol{x}$ stands for a curve $\boldsymbol{x}_k \in \mathbb{R}^{n+1}$, where $k = 0, 1, \dots, n$. As before, $\boldsymbol{u}_0, \dots, \boldsymbol{u}_n$ is the result of Gram-Schmidt applied to $\boldsymbol{x}$, and we refer to $\boldsymbol{u}$ as the "frame" of $\boldsymbol{x}$. Finally, by the Schur parameters of $\boldsymbol{x}$ we mean those of R, where R is the covariance matrix associated with $\boldsymbol{x}$ (i.e., $R_{i,j} = \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$).

Claim 3.1. A *stationary* curve $\boldsymbol{x}$ is completely characterized by its frame and its $n+1$ Schur parameters $s_0, \dots, s_n$.

Proof. As noted earlier, the Schur parameters uniquely identify the stationary covariance matrix R that is associated with $\boldsymbol{x}$. If $\boldsymbol{y}$ is another curve associated with R, then $\boldsymbol{y} = U\boldsymbol{x}$ where U is an orthogonal map of $\mathbb{R}^{n+1}$. Let $\boldsymbol{v}_0, \dots, \boldsymbol{v}_n$ be the frame of $\boldsymbol{y}$, then $\boldsymbol{v}_i = U\boldsymbol{u}_i$ for $i = 0, \dots, n$. Therefore, if the frames of $\boldsymbol{x}$ and $\boldsymbol{y}$ are identical, necessarily $U = I$ and $\boldsymbol{y} = \boldsymbol{x}$. □

Definition. The stationary tangent curve (at $k = 0$), to the curve $\boldsymbol{x}$ is the unique stationary curve $\tilde{\boldsymbol{x}}$ whose frame and set of Schur parameters are identical to those of $\boldsymbol{x}$ (at $k = 0$).

We can now define the stationary tangent process. Let $\mathbb{X}$ be a finite process, in other words, only the finite strecth $\mathbb{X}_0, \dots, \mathbb{X}_n$ is considered. Then $\mathbb{X}$ can be represented as $\mathbb{X}_k = \langle \boldsymbol{\xi}, \boldsymbol{x}_k \rangle$ where $\boldsymbol{x}$ is a curve in $\mathbb{R}^{n+1}$ and $\boldsymbol{\xi}$ is a vector of $n+1$ independent standard normal random variables.

Definition. The stationary tangent process (at $k = 0$) to the process $\mathbb{X}$ is given by $\tilde{\mathbb{X}}_k \stackrel{d}{=} \langle \boldsymbol{\xi}, \tilde{\boldsymbol{x}}_k \rangle$ for $k = 0, \dots, n$, where $\tilde{\boldsymbol{x}}$ is the stationary tangent to the curve $\boldsymbol{x}$.

The following theorem is our main result.

Theorem 1. *The stationary tangent, $\tilde{\mathbb{X}}$, is a locally optimal stationary approximation. More explicitly, $\mathbb{X}_0 = \tilde{\mathbb{X}}_0$ and if $\hat{\mathbb{X}}$ is a stationary process with $\hat{\mathbb{X}}_i = \tilde{\mathbb{X}}_i$ for $i = 0, \dots, k$, then*

$$\mathrm{E}\left|\mathbb{X}_{k+1} - \tilde{\mathbb{X}}_{k+1}\right|^2 \le \mathrm{E}\left|\mathbb{X}_{k+1} - \hat{\mathbb{X}}_{k+1}\right|^2,$$

with equality if and only if $\hat{\mathbb{X}}_{k+1} = \tilde{\mathbb{X}}_{k+1}$.

Proof. As noted in the opening remarks to this section, it suffices to prove the analogue statement for curves, i.e., we need to show that $\|\boldsymbol{x}_{k+1} - \tilde{\boldsymbol{x}}_{k+1}\| \leq \|\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1}\|$. For that we need the following lemma.

Lemma 3.2. *For* $1 \leq k \leq n$, $\mathrm{Span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\} = \mathrm{Span}\{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_k\}$.

Proof. Once again, let $\boldsymbol{u}_k^*$ be the normalized coprojection of $\boldsymbol{x}_0$ on $W_k \stackrel{d}{=} \mathrm{Span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ and let $\boldsymbol{u}_k(1)$ be the normalized coprojection of $\boldsymbol{x}_{k+1}$ on W_k. It suffices to show that

$$\tilde{\boldsymbol{u}}_k(1) = \boldsymbol{u}_k(1) \qquad \text{for } k = 0, \ldots, n-1. \tag{3}$$

The case $k = 0$ follows immediately from (2):

$$\boldsymbol{u}_0(1) = s_1 \boldsymbol{u}_0^* + \sqrt{1 - s_1^2}\boldsymbol{u}_1 = s_1 \tilde{\boldsymbol{u}}_0^* + \sqrt{1 - s_1^2}\tilde{\boldsymbol{u}}_1 = \tilde{\boldsymbol{u}}_0(1).$$

Assuming (3) holds for $0 \leq k < n-1$, it follows that $W_{k+1} = \tilde{W}_{k+1}$, and in particular, $\boldsymbol{u}_{k+1}^* = \tilde{\boldsymbol{u}}_{k+1}^*$. Utilizing (2) we find that

$$\boldsymbol{u}_{k+1}(1) = s_{k+2}\boldsymbol{u}_{k+1}^* + \sqrt{1 - s_{k+2}^2}\boldsymbol{u}_{k+2} = s_{k+2}\tilde{\boldsymbol{u}}_{k+1}^* + \sqrt{1 - s_{k+2}^2}\tilde{\boldsymbol{u}}_{k+2} = \tilde{\boldsymbol{u}}_{k+1}(1),$$

which completes the proof of the lemma. □

Let $\hat{\boldsymbol{x}}$ be a stationary curve with $\hat{\boldsymbol{x}}_i = \tilde{\boldsymbol{x}}_i$ for $i = 0, \ldots, k$. Then, $\tilde{W}_k = \hat{W}_k$ and by the lemma $\tilde{W}_k = W_k$. Let P_W denote the orthogonal projection on W_k, and let $\hat{\boldsymbol{z}} = \hat{\boldsymbol{x}}_{k+1} - P_W(\hat{\boldsymbol{x}}_{k+1})$ and $\tilde{\boldsymbol{z}} = \tilde{\boldsymbol{x}}_{k+1} - P_W(\tilde{\boldsymbol{x}}_{k+1})$.

Since $\hat{\boldsymbol{x}}$ and $\boldsymbol{x}$ are stationary curves and since $\hat{\boldsymbol{x}}_i = \tilde{\boldsymbol{x}}_i$ for $i = 0, \ldots, k$,

$$P_W(\hat{\boldsymbol{x}}_{k+1}) = P_W(\tilde{\boldsymbol{x}}_{k+1}). \tag{4}$$

In particular, $P_W(\hat{\boldsymbol{x}}_{k+1} - \boldsymbol{x}_{k+1}) = P_W(\tilde{\boldsymbol{x}}_{k+1} - \boldsymbol{x}_{k+1})$, and with

$$W^\perp \stackrel{d}{=} \mathrm{Span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{k+1}\} \ominus W_k = \mathrm{Span}\{\boldsymbol{u}_k(1)\},$$

it suffices to show that

$$\|P_{W^\perp}(\boldsymbol{x}_{k+1}) - \hat{\boldsymbol{z}}\| \geq |P_{W^\perp}(\boldsymbol{x}_{k+1}) - \tilde{\boldsymbol{z}}\|\,. \tag{5}$$

Note that it follows from (4) and from the stationarity of $\tilde{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}$ that $\|\hat{\boldsymbol{z}}\| = \|\tilde{\boldsymbol{z}}\|$. By definition $P_{W^\perp}(\boldsymbol{x}_{k+1}) = \alpha \boldsymbol{u}_k(1)$, where $\alpha > 0$. Similarly, by the lemma and by (3), $\tilde{\boldsymbol{z}} = P_{W^\perp}(\tilde{\boldsymbol{x}}_{k+1}) = \beta \boldsymbol{u}_k(1)$ where $\beta > 0$. It follows that (5) holds with a strict inequality unless $\tilde{\boldsymbol{z}} = \hat{\boldsymbol{z}}$ which implies that $\tilde{\boldsymbol{x}}_{k+1} = \hat{\boldsymbol{x}}_{k+1}$. □

In [5] we show that the stationary tangent to a smooth (continuous time) process can be characterized as the stationary process which sequentially minimizes $\mathrm{E}\left|\mathbb{X}_0^{(k)} - \hat{\mathbb{X}}^{(k)}\right|^2$ for $k = 0, 1, \dots$. The next theorem shows that this characterization holds for our discrete tangent provided you interpret derivatives as asymmetric differences.

Let f be a discrete process or curve. We define the asymmetric difference of order k (at i) as:

$$(\Delta^0 f)(i) = f_i$$
$$(\Delta^{k+1} f)(i) = (\Delta^k f)(i+1) - (\Delta^k f)(i)$$

Our differences are all evaluated at $i = 0$ so we omit the i:

$$\Delta^k f \overset{d}{=} (\Delta^k f)(0) = \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} f_j.$$

Corollary 2. *The discrete stationary tangent is the unique sequential minimizer of* $\mathrm{E}\left|\Delta^k \mathbb{X} - \Delta^k \hat{\mathbb{X}}\right|^2$, *for* $k = 0, 1, 2, \dots$.

Proof. As in the previous proof, it suffices to establish the analogous statement for curves. Namely, suppose $\hat{\boldsymbol{x}}$ is a stationary curve with $\Delta^i(\boldsymbol{x} - \hat{\boldsymbol{x}}) = \Delta^i(\boldsymbol{x} - \tilde{\boldsymbol{x}})$ for $i = 0, \dots, k$. Then, we need to show that $\left\|\Delta^{k+1}(\boldsymbol{x} - \hat{\boldsymbol{x}})\right\| \geq \left\|\Delta^{k+1}(\boldsymbol{x} - \tilde{\boldsymbol{x}})\right\|$ with equality if and only if $\hat{\boldsymbol{x}}_{k+1} = \tilde{\boldsymbol{x}}_{k+1}$. The case $k = -1$ is trivial, so we can assume $k \geq 0$. Then, $\hat{\boldsymbol{x}}_0 = \boldsymbol{x}_0$ and therefore

$$\Delta^{k+1}(\boldsymbol{x} - \hat{\boldsymbol{x}}) = \sum_{j=1}^{k} n_j(\boldsymbol{x}_j - \hat{\boldsymbol{x}}_j) + (\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1}).$$

It follows from our assumption that $\hat{\boldsymbol{x}}_i = \tilde{\boldsymbol{x}}_i$ for $i = 0, \dots, k$, and therefore with the same notations as in the proof of the previous theorem,

$$\Delta^{k+1}(\boldsymbol{x} - \hat{\boldsymbol{x}}) = P_W\left(\Delta^{k+1}(\boldsymbol{x} - \hat{\boldsymbol{x}})\right) + P_{W^\perp}(\boldsymbol{x}_{k+1}) - \hat{\boldsymbol{z}}.$$

The same equations hold for $\tilde{\boldsymbol{x}}$ with the obvious modifications. Since (4) holds here as well, we see that

$$P_W\left(\Delta^{k+1}(\boldsymbol{x} - \hat{\boldsymbol{x}})\right) = P_W\left(\Delta^{k+1}(\boldsymbol{x} - \tilde{\boldsymbol{x}})\right),$$

and therefore the proof is completed by (5). ∎

4. The stationary tangent to Brownian motion

In this section we show how one can use discrete stationary tangents to obtain a "good" stationary approximations of the Brownian motion. As we mentioned in the introduction, it is unclear to us what is the extent to which this idea can be generalized to other non-smooth (continuous time) processes. Here we consider b_t which is the standard Brownian motion where $t \in [1, 1+T]$.

Fix $n \in \mathbb{N}$ and let $\delta = \delta_n = T/n$. Let $\mathbb{X}^n$ be a discrete sample of b: $\mathbb{X}^n_k \stackrel{d}{=} b_{1+k\delta}$ where $k = 0, \dots, n$. The Schur parameters of $\mathbb{X}^n$ can be found readily: $s^n_0 = 1$, $s^n_1 = (1+\delta)^{-1/2}$ and for $k \geq 2$, $s^n_k = 0$ since $\mathbb{X}^n$ is a Markov process and s^n_k is the correlation between $\mathbb{X}^n_k$ and $\mathbb{X}^n_0$ given $\mathbb{X}^n_1, \dots, \mathbb{X}^n_{k-1}$. It is convenient to introduce $\alpha = \alpha_n = [\log(1+\delta)]/2\delta$, so that $s^n_1 = e^{-\alpha\delta}$. Since $s^n_k = 0$ for $k \geq 2$, the stationary tangent, $\tilde{\mathbb{X}}^n$, is also a Markov process and it easy to see that it satisfies $\tilde{\mathbb{X}}^n_k = s^n_1(\tilde{\mathbb{X}}^n_{k-1} + \xi^n_k)$, where $\xi^n_k = \mathbb{X}^n_k - \mathbb{X}^n_{k-1} = b_{1+k\delta} - b_{1+(k-1)\delta}$. Thus,

$$\tilde{\mathbb{X}}^n_k = e^{-\alpha k\delta} \cdot b_1 + \sum_{i=0}^{k-1} e^{-\alpha(k-i)\delta} \cdot (b_{1+i\delta} - b_{1+(i-1)\delta}).$$

In what follows it is convenient to identify $\tilde{\mathbb{X}}^n$ with the continuous time process which linearly interpolates the discrete process (with time scaled in the obvious manner). These continuous time processes are not stationary but their limit is, as the next claim shows (note that $\alpha \to 1/2$ as $n \to \infty$).

Claim 4.1. $\tilde{\mathbb{X}}^n$ converges a.s. uniformly on $[0, T]$ to the Ornstein-Uhlenbeck process

$$\tilde{\mathbb{X}}_t \stackrel{d}{=} e^{-t/2} b_1 + \int_0^t e^{-(t-s)/2}\, db_{1+s}. \tag{6}$$

Remark. This result is consistent with the one obtained by considering *symmetric* finite differences as discussed in [5, Sec. 4.2] (here we use asymmetric differences).

Proof. Fix $t \in [0, T]$, let $k = k_n \stackrel{d}{=} [t/\delta]$ and let $Y_n \stackrel{d}{=} \tilde{\mathbb{X}}_t - \tilde{\mathbb{X}}^n_{k\delta}$. Then, Y_n is a normal random variable with mean 0 and

$$|Y_n| \leq \left|e^{-t/2} - e^{-\alpha k\delta}\right| |b_1| + \sum_{i=0}^{k-1} \left| \int_{i\delta}^{(i+1)\delta} \left[e^{-(t-s)/2} - e^{-\alpha(k-i)\delta}\right] db_{1+s} \right| + \left| \int_{k\delta}^{t} e^{-(t-s)/2}\, db_{1+s} \right|.$$

Thus, the variance of Y_n is bounded by

$$\left|e^{-t/2} - e^{-\alpha k\delta}\right|^2 + \sum_{i=0}^{k-1} \int_{i\delta}^{(i+1)\delta} \left[e^{-(t-s)/2} - e^{-\alpha(k-i)\delta}\right]^2 ds + \delta,$$

which is bounded by $c\delta \sim c/n$.

An application of Borel-Cantelli coupled with a simple estimate of the tail of the standard normal distribution will show that $Y_n \to 0$ a.s. In other words, $\tilde{\mathbb{X}}^n_{k\delta} \to \tilde{\mathbb{X}}_t$ a.s. for any fixed t. If we can show that $\tilde{\mathbb{X}}^n$ are a.s. a family of equicontinuous and uniformly bounded functions, then an application of Arzella-Ascoli will complete the proof. Clearly, it suffices to show that such bounds exist for the discrete processes $\tilde{\mathbb{X}}^n_i$.

Since the following result is independent n we suppress the n in $\mathbb{X}^n$. For $k = 0, \dots, n$, let $M_k \stackrel{d}{=} \max\{|\mathbb{X}_0|, \dots, |\mathbb{X}_k|\}$, and analogously for $\tilde{M}$. Then,

$$-M_k + \mathbb{X}_k \leq \tilde{\mathbb{X}}_k \leq M_k + \mathbb{X}_k. \tag{7}$$

Before we prove (7) by induction, we note that it implies the uniform bound:

$$\tilde{M}_k \leq 2M_k \leq 2 \max_{t\in[1,1+T]} |b_t|. \tag{8}$$

For $k = 0$, (7) holds trivially. Assume it holds for $k-1 \geq 0$. If $\tilde{\mathbb{X}}_k \geq 0$, then

$$\tilde{\mathbb{X}}_k \leq \tilde{\mathbb{X}}_{k-1} + \xi_k \leq M_{k-1} + \mathbb{X}_{k-1} + \xi_k = M_{k-1} + \mathbb{X}_k \leq M_k + \mathbb{X}_k,$$

while if $\tilde{\mathbb{X}}_k < 0$, then

$$\tilde{\mathbb{X}}_k < 0 \leq M_k + \mathbb{X}_k.$$

This establishes the second inequality in (7). The first one is obtained by replacing $\mathbb{X}_k$ with $-\mathbb{X}_k$.

Recall that $\tilde{\mathbb{X}}_k = e^{-\alpha\delta}(\tilde{X}_{k-1} + \xi_k)$, therefore,

$$\tilde{X}_k - \tilde{X}_{k-1} = (e^{-\alpha\delta} - 1)\tilde{X}_{k-1} + e^{-\alpha\delta}(b_{1+k\delta} - b_{1+(k-1)\delta}),$$

and the a.s. equicontinuity now follows from (8) and the a.s. uniform continuity of the Brownian motion in $[1, 1+T]$. $\square$

How good is the stationary approximation obtained with this limiting procedure? Let $\tilde{\mathbb{X}}$ be as in (6), and let $\mathbb{X}_t \stackrel{d}{=} b_{1+t}$. Then,

$$\mathrm{E}\left(\mathbb{X}_t - \tilde{\mathbb{X}}_t\right)^2 = t + 2e^{-t/2} - 2 = \frac{1}{4}t^2 - \frac{1}{24}t^3 + \mathrm{O}(t^4).$$

Compare this with the trivial stationary approximation $\hat{\mathbb{X}}_t \equiv b_1$:

$$\mathrm{E}\left(\mathbb{X}_t - \hat{\mathbb{X}}_t\right)^2 = |t|.$$

Moreover, $\tilde{\mathbb{X}}$ is a local stationary approximation which is optimal to first order: let $\hat{\mathbb{X}}$ be any (variance 1) stationary process, then

$$\mathrm{E}\left(\mathbb{X}_t - \hat{\mathbb{X}}_t\right)^2 \geq \left(\sqrt{t} - 1\right)^2 = \frac{1}{4}t^2 - \frac{1}{8}t^3 + \mathrm{O}(t^4).$$

Unfortunately, as we show next, $\tilde{\mathbb{X}}$ is not optimal. Consider stationary processes of the form

$$\hat{\mathbb{X}}_t \overset{d}{=} S(t)b_1 + \int_0^t f(t,r)\, db_{1+r}. \tag{9}$$

Note that $\tilde{\mathbb{X}}$ is also of this form (6). Since $\hat{\mathbb{X}}$ is assumed to be stationary, $S(t-s)$ is its covariance function and

$$S(t-s) = S(t)S(s) + \int_0^{t\wedge s} f(t,r)f(s,r)\, dr. \tag{10}$$

Utilizing Maple and a few informed guesses we were able to find $f(t,r)$ in two cases. We then used those to compute the corresponding distances to $\mathbb{X}$:

$$\mathrm{E}\left(\mathbb{X}_t - \hat{\mathbb{X}}_t\right)^2 = (1-S(t))^2 + \int_0^t (1-f(t,r))^2\, dr.$$

Our first correlation is $S(r) = 1 - |r|/2$. Here

$$f(t,r) = \frac{1-t/4}{1-r/4},$$

and

$$\mathrm{E}\left(\mathbb{X}_t - \hat{\mathbb{X}}_t\right)^2 = 2t + (8-2t)\log(1-t/4) = \frac{1}{4}t^2 + \frac{1}{48}t^3 + \mathrm{O}(t^4).$$

Thus, the corresponding stationary process (9) is doing a slightly worse job than $\tilde{\mathbb{X}}$ in approximating $\mathbb{X}$ about $t=1$. Our second correlation is $S(r) = 1 - |r|/2 + r^2/4$. In this case,

$$f(t,r) = \frac{1 - 3t/4 + t^2/8 + tr/4 - r^2/8 - t^2r/8 + r^3/24 + t^2r^2/32 - tr^3/48}{1 - 3r/4 + r^2/4 - r^3/12 + r^4/96}.$$

Since

$$\mathrm{E}\left(\mathbb{X}_t - \hat{\mathbb{X}}_t\right)^2 = \frac{1}{4}t^2 - \frac{1}{16}t^3 + \mathrm{O}(t^4),$$

this process yields a better local stationary approximation than $\tilde{\mathbb{X}}$. It can be shown that no process of the form (9) can improve on the $-1/16t^3$ term. However, it can also be shown there is no optimal approximation of the form (9), and there might not be an optimal local stationary approximation at all.

It is not hard to see that all three aforementioned stationary approximations to $\mathbb{X}$ satisfy

$$\hat{\mathbb{X}}_t - \mathbb{X}_t = -\frac{1}{2}b_1 t + \mathrm{o}(t^{3/2-\varepsilon}) \qquad \text{a.s.}$$

This is optimal, for if for some stationary $\hat{\mathbb{X}}$, $\hat{\mathbb{X}}_t - \mathbb{X}_t = \xi t + o(t)$ a.s., then ξ is a 0-mean normal random variable, and it follows from $\mathrm{E}\left(\mathbb{X}_t - \hat{\mathbb{X}}_t\right)^2 \geq 1/4t^2 + o(t^2)$, that $\mathrm{E}\,\xi^2 \geq 1/4$.

5. Schur parameters and determinants

In this section we express the generalized Schur parameters in terms of determinants of submatrices of R. The stationary case is well known and can be found for example in [1]. We show how to extend these formulas to the non-stationary case.

Let $R_{[i..j][k..l]}$ denote the submatrix of R obtained by taking rows i through j, and columns k through l of R.

Claim 5.1. (i)

$$\kappa_n \stackrel{d}{=} \sqrt{1 - s_n^2} = \left[\frac{\det R_{[0..n][0..n]} \cdot \det R_{[1..n-1][1..n-1]}}{\det R_{[0..n-1][0..n-1]} \cdot \det R_{[1..n][1..n]}}\right]^{1/2}.$$

(ii)

$$s_n = (-1)^{n-1} \frac{\det R_{[0..n-1][1..n]}}{\left[\det R_{[0..n-1][0..n-1]} \cdot \det R_{[1..n][1..n]}\right]^{1/2}}.$$

Proof. Recall that s_n is the cosine of the angle between $\boldsymbol{u}_{n-1}(1)$ and $\boldsymbol{u}^*_{n-1}$ (Definition 2.2). Thus, $\kappa_n \stackrel{d}{=} \sqrt{1 - s_n^2}$, the sine of the angle between $\boldsymbol{u}_{n-1}(1)$ and $\boldsymbol{u}^*_{n-1}$ is equal to the area of the parallelogram defined by those two vectors (denoted by $A(\boldsymbol{u}_{n-1}(1), \boldsymbol{u}^*_{n-1})$). Let $W \stackrel{d}{=} \mathrm{Span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}\}$. Clearly,

$$A(\boldsymbol{u}_{n-1}(1), \boldsymbol{u}^*_{n-1}) = \frac{A(\boldsymbol{x}_n - P_W(\boldsymbol{x}_n), \boldsymbol{x}_0 - P_W(\boldsymbol{x}_0))}{\|\boldsymbol{x}_n - P_W(\boldsymbol{x}_n)\| \cdot \|\boldsymbol{x}_0 - P_W(\boldsymbol{x}_0)\|}. \tag{11}$$

Let $\mathrm{Vol}(\boldsymbol{x}_i, \boldsymbol{x}_{i+1}, \ldots, \boldsymbol{x}_j)$ denote the volume of parallelepiped generated by $\boldsymbol{x}_i, \ldots, \boldsymbol{x}_j$. Then,

$$\begin{aligned} \|\boldsymbol{x}_n - P_W(\boldsymbol{x}_n)\| &= \mathrm{Vol}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)/\mathrm{Vol}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}), \\ \|\boldsymbol{x}_0 - P_W(\boldsymbol{x}_0)\| &= \mathrm{Vol}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{n-1})/\mathrm{Vol}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}), \\ A(\boldsymbol{x}_n - P_W(\boldsymbol{x}_n), \boldsymbol{x}_0 - P_W(\boldsymbol{x}_0)) &= \mathrm{Vol}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_n)/\mathrm{Vol}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}). \end{aligned}$$

Our proof of (i) is completed by (11) coupled with the observation that $\det R_{[i..j][i..j]} = \left[\mathrm{Vol}(\boldsymbol{x}_i, \ldots, \boldsymbol{x}_j)\right]^2$.

As for (ii), note that both sides are invariant under an application of Gram-Schmidt to $\{\boldsymbol{x}_1, \ldots \boldsymbol{x}_{n-1}\}$. Thus, without loss of generality,

$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}$ are an orthonormal basis of W. It is now a matter of an elementary computation to verify that

$$\langle \boldsymbol{x}_n - P_W(\boldsymbol{x}_n), \boldsymbol{x}_0 - P_W(\boldsymbol{x}_0) \rangle = (-1)^{n-1} \det R_{[0..n-1][1..n]},$$

which completes the proof. □

6. The curvatures and the Schur parameters

So far we saw that the question of local stationary approximations is adequately described in terms of either curvatures, in the smooth case [5], or generalized Schur parameters in the discrete case. One common theme these two fundamental objects share is that both can be defined in terms of a generalization of the recursion formula for the associated orthogonal polynomials. We next show how the curvatures of a smooth correlation can be obtained as a limit of (a simple function of) the Schur parameters of the sampled correlation.

Let R be a smooth correlation and let $\boldsymbol{x}$ be an associated smooth curve in l^2. Then, the ith positive curvature function of R (or $\boldsymbol{x}$) can be defined as [5, (6)-(7)]:

$$\kappa_i = \frac{\sqrt{D_i D_{i-2}}}{D_{i-1}}, \tag{12}$$

where D_i is the square of the volume of the parallelepiped generated by $\{\boldsymbol{x}, \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(i)}\}$, i.e.,

$$D_i \stackrel{d}{=} \begin{vmatrix} \langle \boldsymbol{x}, \boldsymbol{x} \rangle & \langle \boldsymbol{x}, \boldsymbol{x}^{(1)} \rangle & \langle \boldsymbol{x}, \boldsymbol{x}^{(2)} \rangle & \ldots & \langle \boldsymbol{x}, \boldsymbol{x}^{(i)} \rangle \\ \langle \boldsymbol{x}^{(1)}, \boldsymbol{x} \rangle & \langle \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(1)} \rangle & \langle \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \rangle & \ldots & \langle \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(i)} \rangle \\ \langle \boldsymbol{x}^{(2)}, \boldsymbol{x} \rangle & \langle \boldsymbol{x}^{(2)}, \boldsymbol{x}^{(1)} \rangle & \langle \boldsymbol{x}^{(2)}, \boldsymbol{x}^{(2)} \rangle & \ldots & \langle \boldsymbol{x}^{(2)}, \boldsymbol{x}^{(i)} \rangle \\ \ldots & & & & \\ \langle \boldsymbol{x}^{(i)}, \boldsymbol{x} \rangle & \langle \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(1)} \rangle & \langle \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(2)} \rangle & \ldots & \langle \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(i)} \rangle \end{vmatrix}. \tag{13}$$

Implicit in the definitions above is the assumption $t = 0$. Let R^δ be the δ-sampled correlation, i.e., $R^\delta_{ij} \stackrel{d}{=} R(i\delta, j\delta)$. Let s^δ_i be the ith Schur parameter of R^δ, and let $\kappa^\delta_i \stackrel{d}{=} \sqrt{1 - s^\delta_i}$ be the sine of the angle between $\boldsymbol{u}_{n-1}(1)$ and $\boldsymbol{u}^*_{n-1}$.

Claim 6.1.

$$\kappa_i = \lim_{\delta \to 0} \kappa^\delta_i / \delta.$$

Proof. As before let $\Delta \boldsymbol{y}(t) \stackrel{d}{=} \boldsymbol{y}(t + \delta) - \boldsymbol{y}(t)$, and let $\mathrm{Vol}(\boldsymbol{y}_0, \ldots, \boldsymbol{y}_i)$ be the volume of the parallelepiped generated by $\boldsymbol{y}_0, \ldots, \boldsymbol{y}_i$. Since $R^\delta_{[0..i][0..i]}$ is the Grammian matrix of $\boldsymbol{x}_0, \boldsymbol{x}_\delta \ldots, \boldsymbol{x}_{i\delta}$,

$$\sqrt{\det R^\delta_{[0..i][0..i]}} = \mathrm{Vol}\left(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{i\delta}\right) = \mathrm{Vol}\left((\Delta^0 \boldsymbol{x})(0), (\Delta^1 \boldsymbol{x})(0), \ldots, (\Delta^i \boldsymbol{x})(0)\right).$$

Thus,

$$\lim_{\delta\to 0}\frac{\sqrt{\det R^{\delta}_{[0..i][0..i]}}}{\delta^{1+2+\cdots+i}} = \operatorname{Vol}\left(\boldsymbol{x}_0, \dot{\boldsymbol{x}}_0 \ldots, \boldsymbol{x}_0^{(i)}\right) = \sqrt{D_i}.$$

Similarly,

$$\sqrt{\det R^{\delta}_{[1..i][1..i]}} = \operatorname{Vol}\left((\Delta^0\boldsymbol{x})(\delta), \ldots, (\Delta^{i-1}\boldsymbol{x})(\delta)\right),$$

and therefore,

$$\lim_{\delta\to 0}\frac{\sqrt{\det R^{\delta}_{[1..i][1..i]}}}{\delta^{1+\cdots+(i-1)}} = \sqrt{D_{i-1}}.$$

It follows from claim 5.1 that

$$\lim_{\delta\to 0}\frac{\kappa_i^{\delta}}{\delta} = \frac{\sqrt{D_i}\sqrt{D_{i-2}}}{\sqrt{D_{i-1}}\sqrt{D_{i-1}}} = \kappa_i.$$

□

It is worth noting that for Brownian motion and the Ornstein-Uhlenbeck processes, the correct scaling of κ_1^{δ} is $\sqrt{\delta}$. For Brownian motion, $\lim_{\delta}\kappa_1^{\delta}(t)/\sqrt{\delta} = 1/\sqrt{t}$, while for the Ornstein-Uhlenbeck, $\lim_{\delta}\kappa_1^{\delta}(t)/\sqrt{\delta} \equiv 1$. Of course, for both, s_i^{δ} vanish for $i \geq 2$, and s_1^{δ} converges without any scaling.

7. The order of stationarity – an example

In [5, Def. 2.19] we define d, the order of stationarity at t_0, of a smooth process $\mathbb{X}$. It is essentially defined in terms of the number of vanishing derivatives of the curvature functions of $\mathbb{X}$ at t_0. The importance of d is that it yields an optimal upper bound on the order of any stationary approximation to $\mathbb{X}$ at t_0 [5, Theorem 2b]. We next provide an example that demonstrates that at least some of these results can be extended to the non-smooth case.

Let S be a stationary correlation and let φ be a smooth real function. For reasons that will become clearer later on, we assume that $\varphi(t) = t + at^k + \mathrm{o}(t^k)$, where $a > 0$ and $k \geq 2$. Define $R(t,s) \stackrel{d}{=} S\big[\varphi(t) - \varphi(s)\big]$. If S is smooth then the curvatures of R satisfy $\kappa_i^R(t) = \dot{\varphi}(t)\kappa_i^S$ [5, Example 2.23]. In particular, the order of stationarity, at $t = 0$, of the governed process $\mathbb{X}$ is $d = k - 1$. Therefore, the stationary tangent which is an optimal stationary approximation, yields

$$\|\mathbb{X} - \tilde{\mathbb{X}}\| \stackrel{d}{=} \left[\mathrm{E}\,|\mathbb{X} - \tilde{\mathbb{X}}|^2\right]^{1/2} = t^k + \mathrm{o}(t^k) \tag{14}$$

Consider now the non-smooth stationary Ornstein-Uhlenbeck correlation, $S = \exp(-|t-s|/2)$, and let R be as defined above. We shall find the tangent and the order of stationarity of the governed process $\mathbb{X}$. A realization of $\mathbb{X}$ is given by

$$\mathbb{X}_t = e^{-\varphi(t)/2} b_1 + \int_0^t e^{-(\varphi(t)-\varphi(s))/2} \sqrt{\dot{\varphi}(s)}\, db_{1+s},$$

where b_t is the standard Brownian motion and we consider $t \in [0,T]$, where we assume φ is increasing. Let $\delta = \delta_n \stackrel{d}{=} T/n$ and let $\tilde{\mathbb{X}}^n_k$, $k = 0,\ldots,n$ be the discrete tangent of the sampled process $\mathbb{X}^n_k \stackrel{d}{=} \mathbb{X}_{k\delta}$. Once again we abuse our notations by using $\tilde{\mathbb{X}}^n$ to also denote the piecewise linear process obtained by identifying $\tilde{\mathbb{X}}^n_{k\delta} = \tilde{\mathbb{X}}^n_k$ for $k = 0,\ldots,n$.

Claim 7.1. With probability one,

$$\tilde{\mathbb{X}}^n_t \longrightarrow \tilde{\mathbb{X}}_t \stackrel{d}{=} e^{-t/2} b_1 + \int_0^t e^{-(t-s)/2}\, db_{1+s},$$

uniformly for $t \in [0,T]$.

Remark. Thus, the Ornstein-Uhlenbeck process, $\tilde{\mathbb{X}}$, is the stationary tangent at $t = 0$ to $\mathbb{X}$. Note that since $\mathbb{X}$ is a Markov process, $s^\delta_i = 0$ for all $i \geq 2$, and we should therefore expect its stationary tangent to be an Ornstein-Uhlenbeck process. The coefficient in the exponent is determined by $\dot{\varphi}(0) = 1$.

Proof. Let $s^\delta_1 = e^{-\varphi(\delta)/2}$ be the first Schur parameter of the sampled process $\mathbb{X}^n$, and let

$$\eta^n_k \stackrel{d}{=} \left[\frac{1-e^{-\varphi(\delta)}}{e^{-\varphi((k-1)\delta)} - e^{-\varphi(k\delta)}}\right]^{1/2} \int_{(k-1)\delta}^{k\delta} e^{-\varphi(s)/2} \sqrt{\dot{\varphi}(s)}\, db_{1+s}.$$

It is not difficult to verify that $\tilde{\mathbb{X}}^n$ is given by

$$\begin{aligned}
\tilde{\mathbb{X}}^n_k &= \sum_{j=1}^k \left(s^\delta_1\right)^{k-j} \eta^n_j + \left(s^\delta_1\right)^k \mathbb{X}_0 \\
&= e^{-k\varphi(\delta)/2} \left\{ b_1 + \sum_{j=1}^k e^{j\varphi(\delta)/2} \left[\frac{1-e^{-\varphi(\delta)}}{e^{-\varphi((j-1)\delta)} - e^{-\varphi(j\delta)}}\right]^{1/2} \int_{(j-1)\delta}^{j\delta} e^{-\varphi(s)/2} \sqrt{\dot{\varphi}(s)}\, db_{1+s} \right\}.
\end{aligned}$$

Let

$$f_n(s) \stackrel{d}{=} \sum_{j=1}^n 1_{[(j-1)\delta, j\delta)}(s) e^{j\varphi(\delta)/2} \left[\frac{1-e^{-\varphi(\delta)}}{e^{-\varphi((j-1)\delta)} - e^{-\varphi(j\delta)}}\right]^{1/2} e^{-\varphi(s)/2} \sqrt{\dot{\varphi}(s)},$$

and let $\mathbb{Y}_t^n \overset{d}{=} b_1 + \int_0^t f_n(s)\, db_{1+s}$. Finally, let $\mathbb{Y}_t \overset{d}{=} e^{t/2}\tilde{\mathbb{X}}_t = b_1 + \int_0^t e^{s/2}\, db_{1+s}$. Then, $\mathbb{Y}_t^n \longrightarrow \mathbb{Y}_t$ a.s. uniformly for $t \in [0,T]$. Indeed, let $\sigma_n = \|f_n(s) - \exp(s/2)\|_{L^2(0,T)}$, then by the Doob's submartingale inequality,

$$\operatorname{Prob}\left(\max_{t\in[0,T]} \mathbb{Y}_t^n - \mathbb{Y}_t > \varepsilon_n\right) = \operatorname{Prob}\left(\max_{t\in[0,T]} e^{\theta_n(\mathbb{Y}_t^n - \mathbb{Y}_t)} > e^{\theta_n\varepsilon_n}\right) \le e^{\theta_n^2\sigma_n^2/2 - \theta_n\varepsilon_n} = e^{-\varepsilon_n^2/(2\sigma_n^2)},$$

where we chose $\theta_n = \varepsilon_n/\sigma_n^2$. It is not hard to see that $\sigma_n \le c\delta$ for some constant c, and therefore a simple application of Borel-Cantelli yields the desired convergence of $\mathbb{Y}^n$. Note that $\mathbb{Y}_{k\delta}^n = e^{k\varphi(\delta)/2}\tilde{\mathbb{X}}_{k\delta}^n$. Thus, with $k = k_\delta \overset{d}{=} [t/\delta]$,

$$\tilde{\mathbb{X}}_{k\delta}^n = e^{-(k\delta)(\varphi(\delta)/2\delta)}\mathbb{Y}_{k\delta}^n \xrightarrow[\delta\to 0]{} e^{-t/2}\mathbb{Y}_t = \tilde{\mathbb{X}}_t,$$

where this convergence is uniform for $t \in [0,T]$ a.s. Since $\mathbb{X}_t^n$ is piecewise linear, our proof is complete. □

How good is this tangent approximation? Recall that $\varphi(t) = t + at^k + \mathrm{o}(t^k)$. An elementary computation yields

$$\left\|\mathbb{X}_t - \tilde{\mathbb{X}}_t\right\|^2 = \left[e^{-t/2} - e^{-\varphi(t)/2}\right]^2 + \int_0^t \left[e^{-(t-s)/2} - e^{-(\varphi(t)-\varphi(s))/2}\sqrt{\dot{\varphi}(s)}\right]^2 ds$$
$$= \frac{k^2a^2}{4(2k-1)}t^{2k-1} + \mathrm{o}(t^{2k-1}),$$

or

$$\left\|\mathbb{X}_t - \tilde{\mathbb{X}}_t\right\| = \frac{ka}{2\sqrt{2k-1}}t^{k-1/2} + \mathrm{o}(t^{k-1/2})$$

This is the non-smooth analogue of (14). We still have to show that this is the optimal order.

For a smooth process $\mathbb{X}$, we know that the rate of change at $t = 0$ of the first curvature $\kappa_1(t)$ yields an upper bound on the order of any stationary approximation to $\mathbb{X}$ at $t = 0$. By proving an analogous result for a non-smooth process, we show that the aforementioned tangent is a stationary approximation of optimal order.

Let $s_i^\delta(t)$ be the ith Schur parameter of the continuous process $\mathbb{X}$ sampled at $\mathbb{X}_t, \mathbb{X}_{t+\delta}, \mathbb{X}_{t+2\delta}, \ldots$, and let $\kappa_i^\delta(t) = \sqrt{1 - \left[s_i^\delta(t)\right]^2}$.

Claim 7.2. If $\kappa_1^\delta(\delta) - \kappa_1^\delta(0) = C\delta^{k-1/2} + \mathrm{o}(\delta^{k-1/2})$, then for any stationary process $\hat{\mathbb{X}}$, $\left\|\mathbb{X}_t - \hat{\mathbb{X}}_t\right\| \ge \mathrm{O}(t^{k-1/2})$.

Remark. It follows that $\tilde{\mathbb{X}}$, the tangent process in our example, is of optimal order. Indeed, in this case, $\kappa_1^\delta(\delta) = \sqrt{1 - \exp(-|\varphi(t+\delta) - \varphi(\delta)|)}$. Therefore, $\kappa_1^\delta(\delta) - \kappa_1^\delta(0) = a(2^k - 1)\delta^{k-1/2} + \mathrm{o}(\delta^{k-1/2})$, which is of the same order as $\left\|\mathbb{X}_t - \tilde{\mathbb{X}}_t\right\|$.

Proof of the Claim. Let $\hat{\mathbb{X}}$ be a stationary process which is jointly normal with $\mathbb{X}$ and consider the curves $\boldsymbol{x}_{k\delta}$ and $\hat{\boldsymbol{x}}_{k\delta}$ corresponding to the sampled processes $\mathbb{X}_{k\delta}$ and $\hat{\mathbb{X}}_{k\delta}$. Without loss of generality we can assume $\|\mathbb{X}_0\| = \|\hat{\mathbb{X}}_0\| = 1$. Let $\theta(\boldsymbol{x}_\delta, \boldsymbol{x}_{2\delta})$ denote the angle between these vectors. Recall that $\kappa_1^\delta(0) = \sin(\theta(\boldsymbol{x}_0, \boldsymbol{x}_\delta))$, then

$$\gamma \stackrel{d}{=} |\kappa_1^\delta(\delta) - \kappa_1^\delta(0)| \sim |\theta(\boldsymbol{x}_\delta, \boldsymbol{x}_{2\delta}) - \theta(\boldsymbol{x}_0, \boldsymbol{x}_\delta)|.$$

Since $\hat{\boldsymbol{x}}$ is stationary, $|\theta(\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_\delta) - \theta(\hat{\boldsymbol{x}}_\delta, \hat{\boldsymbol{x}}_{2\delta})| = 0$ and therefore, for small δ, $\max\{|\theta(\boldsymbol{x}_0, \boldsymbol{x}_\delta) - \theta(\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_\delta)|, |\theta(\boldsymbol{x}_\delta, \boldsymbol{x}_{2\delta}) - \theta(\hat{\boldsymbol{x}}_\delta, \hat{\boldsymbol{x}}_{2\delta})|\} > \gamma/3$. Suppose that $|\theta(\boldsymbol{x}_\delta, \boldsymbol{x}_{2\delta}) - \theta(\hat{\boldsymbol{x}}_\delta, \hat{\boldsymbol{x}}_{2\delta})| > \gamma/3$ (the other inequality is dealt with in exactly the same way). Since we seek to minimize $\max\{\|\boldsymbol{x}_\delta - \hat{\boldsymbol{x}}_\delta\|, \|\boldsymbol{x}_{2\delta} - \hat{\boldsymbol{x}}_{2\delta}\|\}$, we can assume, without loss of generality, that $\hat{\boldsymbol{x}}_\delta$ and $\hat{\boldsymbol{x}}_{2\delta}$ are in the same plane defined by $\boldsymbol{x}_\delta$ and $\boldsymbol{x}_{2\delta}$. By our assumption, either $\theta(\boldsymbol{x}_\delta, \hat{\boldsymbol{x}}_\delta) > \gamma/6$ or $\theta(\boldsymbol{x}_{2\delta}, \hat{\boldsymbol{x}}_{2\delta}) > \gamma/6$. Since δ and therefore γ are small and since we assume that $\|\mathbb{X}_0\| = 1$, it follows that $\max\{\|\boldsymbol{x}_\delta - \hat{\boldsymbol{x}}_\delta\|, \|\boldsymbol{x}_{2\delta} - \hat{\boldsymbol{x}}_{2\delta}\|\} > \gamma/12$, which completes the proof. □

8. ACKNOWLEDGEMENTS

We would like to thank an anonymous referee for insightful comments.

References

[1] Akhiezer, N.I. The classical moment problem and some related questions in analysis. Translated by N. Kemmer. Hafner Publishing Co., New York 1965.

[2] Dahlhaus, R. Fitting time series models to non-stationary processes. Ann. Statist. 25 (1997), no. 1, 1–37.

[3] Kailath, T. A Theorem of I. Schur and Its Impact on Modern Signal Processing. I. Schur methods in operator theory and signal processing, edited by I. Gohberg. Basel ; Boston : Birkhuser Verlag, 1986.

[4] Kailath, T. Time-variant and time-invariant lattice filters for non-stationary processes. (English. French summary) Mathematical tools and models for control, systems analysis and signal processing, Vol. 2 (Grenoble, 1980/Aussois, 1981), 417–464, Travaux Rech. Coop. Programme 567, CNRS, Paris, 1982.

[5] Keich, U. A possible definition of a stationary tangent. Stochastic Process. Appl. 88 (2000), no. 1, 1–36.

[6] Landau, H. J. Maximum entropy and the moment problem. Bull. Amer. Math. Soc. (N.S.) 16 (1987), no. 1, 47–77.

[7] Mallat, S.; Papanicolaou, G.C.; Zhang, Z. *Adaptive covariance estimation of locally stationary processes.* Ann. Statist. 26 (1998), no. 1, 1-47.

[8] Neumann, Michael H.; von Sachs, Rainer Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. Ann. Statist. 25 (1997), no. 1, 38–76.

[9] M.B. Priestley . Non-linear and Non-stationary Time Series Analysis. Academic Press, 1988.

[10] M.B. Priestley . Wavelets and time-dependent spectral analysis. Journal of Time Series Analysis, Vol 17, No. 1, (1996).

Department of Computer Science & Engineering, University of California, San Diego, CA 92093-0114

E-mail address: `keich@cs.ucsd.edu`