VOLATILITY ESTIMATION FOR NON-LINEAR SPDE

MARTIN ANDERSSON, BENNY AVELIN, VALENTIN GARINO,
PAULIINA ILMONEN, AND LAURI VIITASAARI

ABSTRACT. BLAH

LATEST PLAN

- Lauri will continue cleaning up.
- Valentin asks B and M about simulations, and checks if Prop 3.6 goes with Fourier techniques.

Excluding Prop 3.6, everything works out at least if we assume for the measure γ that it satisfies

$$\widehat{\gamma}(\xi) \approx (1+|\xi|)^{\beta-d}.$$

Here $\approx$ means lower and upper bounded with a constant, and this would generalise of course the class of covariances. On top of that we would like to cover the case $\gamma(dx) = \gamma(x)dx$ where γ is a nice function, having integrability/regularity properties what ever we want. Point here to make though is the fact that if γ is very nice function, then the Fourier coefficient decay faster than any polynomial so assumption above is not okay.

LEFT TO DO:

(1) Add some context/references/explanations in the introduction. Valentin have some references, we need more.
(2) Selling point. Add bullet points to Intro, P will do magic.
(3) Do the simulations: Benny and Martin! multidim? Fractional in space?
(4) Give some examples other than the heat kernel. The Kolmogorov operator might work, but unfortunately we would need to modify the assumptions and the proof of the main theorem
(5) Verify the proofs and correct the typos. L cleans
(6) Do the proof of (**??**). This is fine, just explain clearly that it works.

1. INTRODUCTION

Some background consideration and litterature review.

Litterature review:

P.Bossert: *Parameter estimator for second-order SPDEs in multiple space dimensions*, https://arxiv.org/abs/2310.17828, 2023

- this paper uses a "quadratic variation" type estimator to estimate the (constant) volatility in a SPDE (with Dirichlet boundary conditions $u(0,t) = u(1,t) = 0$) of type

$$-Lu(t,x) + \sigma \dot{W}(t,x) = 0,$$

with $L = \frac{\partial}{\partial t} - \eta \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} + \sum_{i=1}^{d} \nu_i \frac{\partial}{\partial x_i}$ ($\eta > 0$ constant, ν_i constant)

They use the spectral approach, obtain both rates of convergence and a CLT for their estimator.

Y.Tonaki, Y.Kaino And M Uchida: *Parameter estimation for linear parabolic spdes in two sapce dimensions based on high frquency datas*, Scandinavian Journal of Statistics, 2023

- Does the same thing, but only in 2 dimensions (so less general)

M Bibinger and M Trabs: *Volatility estimation for stochastic PDEs using high frequency observations*, https://arxiv.org/abs/1710.03519, 2017

- Use quadratic variation aproach on a one-dimensional SPDE (with Dirichlet boundary conditions) driven by space time white noise only, of the form $-Lu(t,x) + \sigma_t \dot{W}(t,x)$, with $Lf = \frac{\partial f}{\partial t} - \nu_2 \frac{\partial^2 f}{\partial x^2} + \nu_1 \frac{\partial f}{\partial x} + \nu_0 f$ with $\nu_2 > 0$, ν_0, ν_1, ν_2 constant parameters. One interesting thing is that they have a variable (but deterministic) volatility σ_t, and they estimate $\int_0^T \sigma_t^2 dt$ (I guess that by shrinking the interval size they could also estimate σ_t^2 for all t). Under some regularity conditions on $t\sigma_t$, they also obtain a CLT for their estimator.

  Worth mentioning that this paper (and some others) also give a joint estimator for $(\sigma, \nu_1, \nu_2, \nu_0)$ although it is not possible to estimate all the parameters at the same time (only the ratios).

R. Almeyer and M. Reiss: *Non-parametric estimation for linear SPDEs from local measurements*, Annals of Applied Probability (2022)

- Pretty convoluted paper, I suggest you have a look because I'm not sure I understand everything. P. Bossert (see above) kind of miscaracterize this paper by saying that it only looks at 1-D SPDE but in reality it studies SPDEs in multiple space dimension driven by space time white noise in a distribution space (since real valued solutions do not exist in multiple space dimension)

  SPDEs they look at are of the form

$$-\frac{\partial X}{\partial t} + \sum_{i=1}^{d} \partial_{x_i}(\nu(x)\partial_{x_i} X) + a(x) \sum_{i=1}^{d} \partial_{x_i} X + b(x)X + \sigma(x)\dot{W}(t,x)$$

  with ν, a, b unknown functions, σ deterministic (they use the fact that the solution is a distributional Gaussian Field, so σ must be deterministic).

σ is not the target quantity, they actually use MLE estimators to estimate ν. Looks pretty impressive but also quite different than what we do.

R. Altmeyer, T. Bretschneider, J. Janak, M. Reiss: *Parameter Estimation in an SPDE Model for Cell Repolarisation.* SIAM Journal on Uncertainty Quantification, 2022

- Looks pretty similar to the previous paper (by partially the same authors), but in a more applied setting.

I Cialenco and Y. Huang: *A note on parameter estimation for discretely sampled SPDEs* Stochastics and Dynamics, 2020

- Looks pretty similar to the paper by Bibinger and Traps referenced above.

P. Bossert, M. Bibinger: *Efficient parameter estimation for parabolic spdes based on a log-linear model for realized volatilities.* Japanese Journal of Statistics and Data Science, 2023

- Another paper using the power variation approach for 1-D SPDE/Constant coefficients

I Cialenco: *Statistical Inference for SPDEs: an overview.* Statistical Inferences for Stochastic Processes, 2018.

- I think you know this paper already? It's a pretty general survey, but most references in it are either outdated or focus on drift estimation rather than volatility estimation. Probably still worth mentionning though

2. Non-parametric estimation of the diffusion function

Let us now introduce more precisely the setting of the present paper. Let us fix $T > 0$. We consider the following d-dimensional stochastic partial differential equation driven by a white-coloured noise W with spatial covariance Γ (see Section **??** for a precise definition) and with bounded regular initial condition.

$$\begin{cases} -Lu(t,x) + \sigma(u(t,x))\dot{W}(t,x) = 0 \\ u(\cdot,0) = u_0, \end{cases} \tag{1}$$

for $x \in \mathbb{R}^d$, $t \in [0,T]$. Here, u_0 belongs to the space of bounded Lipschitz functions, $\sigma : \mathbb{R} \to \mathbb{R}$ is an (unknown) measurable function which satisfies the following condition:

Assumption (A): σ is non negative and M-Lipschitz continuous for some $M > 0$.

Let us now introduce the Riesz potential. Let $0 < \beta < d$ and μ be a positive measure on $\mathbb{R}^d$. The Riesz potential $I_{d-\beta}$ is defined for $x \in \mathbb{R}^d$

$$\left(I_{d-\beta}\mu\right)(x) = \int_{\mathbb{R}^d} \|x-y\|^{-\beta} d\mu(y) = (K_{d-\beta} * \mu)(x), \tag{2}$$

with $K_{d-\beta}(y) = \|y\|^{-\beta}$, where $\|.\|$ denotes the L^2 norm.

Assumption (B) The spatial covariance Γ of the noise W can be written as

$$\Gamma = (I_{d-\beta}\delta), \tag{3}$$

with $\beta \leq d$ and $\beta < 2$.

On the other hand,we assume that the differential operator L satisfies the following assumptions:

Assumption (C)

$$L := \partial_t - \sum_{i,j=1}^{d} a_{i,j}(t,x)\partial^2_{x_i x_j} - \sum_{i=1}^{d} b_i(t,x)\partial_{x_i}, \tag{4}$$

where a, b are Lipschitz, with the matrix (a_{ij}) being uniformly elliptic. double check that this is enough to get the estimates we need instead of extra assumption, i.e. add more regularity to coefficient functions if needed

Remark 2.1. The non negativity assumption in **(A)** is necessary for identifiability purpose since we are actually going to estimate the quantity σ^2.

Remark 2.2. The assumption **(C1)** ensure that zero-order derivative $G(t, \cdot)$ satisfies the growth condition (i) in Assumption **(C2)**, see Section **??**.

A noteworthy particular case of the system **??** is the *stochastic heat equation*, given by

$$\frac{\partial u(t,x)}{\partial t} = \frac{1}{2}\Delta u(t,x) + \sigma(u(t,x))\dot{W}(t,x), \tag{5}$$

where Δ is the Laplace operator. We will use the abreviation SHE in what follows.

The goal of this paper is to build a consistent non parametric estimator of the process $(\sigma(u(t,x)))_{(x,t)\in\mathbb{R}^d\times[0,T]}$ for the L^1 norm. Let us first give a heuristic description of the method we will adopt:

Here we can provide an intuitive explanation, maybe by looking at the simplified case where W is smooth.

Since the noise W is not a proper function and u is not differentiable, we will have to introduce a discretisation of the operator L and perform a regularisation procedure over a small ball in order to apply this method. Let $(x_0, t_0) \in \mathbb{R} \times [0, T]$ and consider two discretisation parameters $h, \epsilon > 0$. Let $\mathcal{B}_{x_0,\epsilon}$ be the euclidian ball in $\mathbb{R}^d$ centered in x_0 and with radius ϵ. Let $V(\epsilon, d)$ denotes its volume, that is

$$V(\epsilon, d) = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}\epsilon^d. \tag{6}$$

Consider the estimator

$$\Sigma_{\epsilon,h}(x_0,t_0) := \mathbb{E}_{u(t_0,x_0)}\left[\left(\frac{1}{m(\epsilon)}\int_{\mathcal{B}_{x_0,\epsilon}}\int_{t_0}^{t_0+\epsilon}\right.\right.$$
$$\left.\left.\times\left(L^h u(s,y) - \int_{\mathbb{R}^d} L^h G(s,y-z)u_0(z)dz\right)dy,ds\right)^2\right], \tag{7}$$

where $\mathbb{E}_{u(t_0,x_0)}$ is the conditional expectation given $u(x_0,t_0)$, G is the fundamental solution of the associated PDE, m is given by

$$m(\epsilon) = \sqrt{\epsilon\int_{\mathcal{B}_{0,\epsilon}}\int_{\mathcal{B}_{0,\epsilon}}\Gamma(x-y)dxdy} = \sqrt{\epsilon}m_1(\epsilon) \tag{8}$$

and L^h is the finite difference approximation of the operator L. More precisely, for every function $f:[0,T]\times\mathbb{R}^d\mapsto\mathbb{R}$,

$$\begin{aligned} L^h : f(t,x) &\rightarrow (T_{h^2}f)(t,x) - \sum_{i,j=1}^d a_{ij}(t,x)(\mathcal{D}_{ij}^{2,h}f)(t,x) \\ &\quad + \sum_{i=1}^d b_i(t,x)(\mathcal{D}_i^{1,h}f)(t,x), \\ &= (T_{h^2}f)(t,x) - (S^h f)(t,x) \end{aligned} \tag{9}$$

where

$$\begin{aligned} (T^{h^2}f)(t,x) &= \frac{f(t+h^2,x)-f(t,x)}{h^2} \\ (\mathcal{D}_{ij}^{2,h}f)(t,x) &= \frac{f(t,x+h(e_i+e_j))+f(x,t)-f(t,x+he_i)-f(t,x+he_j)}{h^2} \\ (\mathcal{D}_i^{1,h}f)(t,x) &= \frac{f(t,x+he_i)-f(t,x)}{h}. \end{aligned}$$

Notice that when $\gamma=\delta$, $m_1(\epsilon)=\sqrt{V(\epsilon,d)}$.

The main result of the paper is as follow.

Theorem 2.3. *Let $h>0$, $p\in\mathbb{N}^*$. Let $\rho\in(0,1)$ and let $\epsilon=h^\rho$. Then, assuming that **(A)-(B)-(C1)-(C2)** are verified, we have for all $(x_0,t_0)\in\mathbb{R}^d\times[0,T]$, for all $\gamma\in(\rho,1)$ and for all $(\nu,\kappa)\in\left(1-\frac{\beta}{2},\frac{1}{2}-\frac{\beta}{4}\right)$,*

$$\begin{aligned} &\left\|\Sigma_{\epsilon,h}(x_0,t_0)-\sigma^2(u(x_0,t_0))\right\|_{L^p(\Omega)} \\ \le\;& K(|x_0|^2+|t_0|^2)\left(M(h^{\rho\kappa}+h^{\rho\nu})\right. \\ &\left.+h^{(1-\rho)(d+1-\beta)}+h^{2-2\gamma}+h^{(2-\beta)(\gamma-\rho)}\left(1+log(h)\mathbb{I}_{\beta=1}\right)\right), \end{aligned} \tag{10}$$

(11)

where K is a constant which depends on $\gamma, M, p, \nu, \kappa, \beta, d, G$ (where G is the fundamental solution of the homogeneous PDE associated to (**??**)*)*

somewhere heuristic derivation and link/comparison to existing literature and SDE case. Also explain the bias/variance tradeoff as with many non-parametric estimators

Remark 2.4. It is actually possible to optimize the rate in the previous result assuming a specific value for β. Indeed we have to solve the optimization problem

$$\begin{cases} \textit{Maximize } \min\left(\rho\nu, \rho\kappa, 2-2\gamma, (1-\rho)(d+1-\beta), (2-\beta)(\gamma-\rho)\right) \\ \textit{Given } 1 > \gamma > \rho > 0 \end{cases}$$

The solution to this problem is obtained for

$$\begin{aligned} \rho_0 &= \frac{4-2\beta}{4-2\beta+min(\kappa,\nu)(4-\beta)} \\ \gamma_0 = &= \frac{4+(2-\beta)\rho_0}{4-\beta}, \end{aligned}$$

yielding an overall rate of convergence

$$h^{\frac{(-2+\beta)\min(\kappa,\nu)}{4-2\beta+\min(\kappa,\nu)(4-\beta)}}.$$

Taking κ arbitrarily close to $\frac{2-\beta}{4}$, we can then see that the rate can be arbitrarily close to

$$h^{\frac{(-2+\beta)}{12-\beta}}$$

(of course, this will yield an exploding constant K as $\kappa \to \frac{2-\beta}{4}$) .

Remark 2.5. Notice that when $M \neq 0$ (that is, σ is not constant) and when $\beta = d = 1$ and taking κ, ν arbitrary close to $\frac{1}{4}$ and $\frac{1}{2}$ respectively, we can see by the previous remark that the rate of convergence is arbitrarily close to $\frac{1}{11}$. The low Hölder regularity of the solution to the SPDE (**??**) is mainly to blame for this surprisingly slow rate. Indeed, if we consider the much simpler problem of estimating a constant volatility $\sigma^2 \in \mathbb{R}_+$, then since $M = 0$, we can choose $\rho = 0$ and the optimal rate in (**??**) is then obtained for γ_0 such that $2 - 2\gamma_0 = (2-\beta)\gamma_0$ (when $\beta = 1$ for example, we have the much faster optimal speed of convergence $log(h)h^{\frac{2}{3}}$, obtained for $\gamma = \frac{2}{3}$).

discussion about discretisation of integral, what error it produces, and continue with discussion about regression then how conditional expectation can be computed in practice. That is, for a given point $u(t_0, x_0)$ we observe u on a ball around it but not for all points, so discuss approximation better (by assuming dense observations). That is, u is always measured for t_k and x_k around (t_0, x_0). Make another theorem out of this with the idea that each measurement consist of measures of u on (t_k, x_j), and then one can make a regression problem out of it

3. Numerical experiments

add simulations

Martin Andersson, Department of Mathematics, Uppsala University, S-751 06 Uppsala, Sweden

Benny Avelin, Department of Mathematics, Uppsala University, S-751 06 Uppsala, Sweden

Valentin Garino, Department of Mathematics, Uppsala University, S-751 06 Uppsala, Sweden

Pauliina Ilmonen, Department of Mathematics and Systems Analysis, Aalto University School of Science, 00076 Aalto, Finland

Lauri Viitasaari, Department of Mathematics, Uppsala University, S-751 06 Uppsala, Sweden