Deep PPDEs for rough local stochastic volatility

Antoine Jacquier
(Imperial College London and Alan Turing Institute)

Fourth Conference on Ambit Fields and Related Topics
Sandbjerg Estate, Sønderborg, Denmark

Joint work with Mugad Oumgari (Lloyds Bank)

August 7, 2019

(Instantaneous) Volatility is rough!

Consider a one-dimensional asset price process

$$\frac{\mathrm{d}S_t}{S_t} = \sigma_t \, \mathrm{d}W_t,$$

for some Brownian motion W.

- Classical approach (stochastic volatility models: Heston, SABR, Stein-Stein,...):

$$\mathrm{d}\sigma_t = b(t, \sigma_t) \, \mathrm{d}t + \xi(t, \sigma_t) \, \mathrm{d}B_t,$$

  with $\mathrm{d}\langle W, B\rangle_t = \rho \, \mathrm{d}t$, for $\rho \in [-1, 1]$.

(Instantaneous) Volatility is rough!

Consider a one-dimensional asset price process

$$\frac{\mathrm{d}S_t}{S_t} = \sigma_t \, \mathrm{d}W_t,$$

for some Brownian motion W.

- Classical approach (stochastic volatility models: Heston, SABR, Stein-Stein,...):

  $$\mathrm{d}\sigma_t = b(t, \sigma_t) \, \mathrm{d}t + \xi(t, \sigma_t) \, \mathrm{d}B_t,$$

  with $\mathrm{d}\langle W, B\rangle_t = \rho \, \mathrm{d}t$, for $\rho \in [-1, 1]$.
- Gatheral-Jaisson-Rosenbaum (2014), Bennedsen-Lunde-Pakkanen (2016): Across large financial data sets

  $$\mathbb{E}\left[|\log(\sigma_{t+\Delta}) - \log(\sigma_t)|^q\right] \approx C\Delta^{Hq}, \quad \text{for any } q > 0, \qquad \text{with } H \approx 0.1;$$

  $\longrightarrow$ consistent with $\sigma_t = \exp(B_t^H)$, where B^H is a fractional Brownian motion;
- **Rough volatility models**: fBm with $H < \frac{1}{2}$ is less 'regular' than standard BM.

(Instantaneous) Volatility is rough!

Consider a one-dimensional asset price process

$$\frac{\mathrm{d}S_t}{S_t} = \sigma_t \,\mathrm{d}W_t,$$

for some Brownian motion W.

- Classical approach (stochastic volatility models: Heston, SABR, Stein-Stein,...):

$$\mathrm{d}\sigma_t = b(t, \sigma_t)\,\mathrm{d}t + \xi(t, \sigma_t)\,\mathrm{d}B_t,$$

  with $\mathrm{d}\langle W, B\rangle_t = \rho\,\mathrm{d}t$, for $\rho \in [-1, 1]$.

- Gatheral-Jaisson-Rosenbaum (2014), Bennedsen-Lunde-Pakkanen (2016): Across large financial data sets

$$\mathbb{E}\left[|\log(\sigma_{t+\Delta}) - \log(\sigma_t)|^q\right] \approx C\Delta^{Hq}, \quad \text{for any } q > 0, \qquad \text{with } H \approx 0.1;$$

  $\longrightarrow$ consistent with $\sigma_t = \exp(B_t^H)$, where B^H is a fractional Brownian motion;

- **Rough volatility models**: fBm with $H < \frac{1}{2}$ is less 'regular' than standard BM.

$\longrightarrow$ **Rough Volatility website**

What kind of volatility models? $\frac{\mathrm{d}S_t}{S_t} = \sigma_t \, \mathrm{d}W_t$

- Simplest rough model: $\sigma_t = \exp(B_t^H)$;
- Rough SDE: $\mathrm{d}\sigma_t = (\cdots)\,\mathrm{d}t + (\cdots)\,\mathrm{d}B_t^H$ (Comte-Renault (1998), El Euch-Rosenbaum (2019), Horvath-Jacquier-Lacombe (2019))
- More natural: $\sigma_t = \exp(X_t)$, for some Gaussian Volterra process

$$X_t = \int_0^t \mathrm{K}(t-s)^\top \cdot \mathrm{d}B_s,$$

  with B is a (possibly multidimensional) Brownian motion and K a (singular) kernel such that $\int_0^t \|\mathrm{K}(t-s)\|^2 ds < \infty$ (Bennedsen-Lunde-Pakkanen (2017));
  - Power-law kernel;
  - Gamma kernel;
  - fBm kernel;
- BSS process: $\sigma_t = \exp(X_t)$, with $X_t = \int_0^t \sqrt{\Gamma_s}\mathrm{K}(t-s)^\top \mathrm{d}B_s$, where Γ is a strictly positive affine process independent of B (Horvath-Jacquier-Tankov (2019));

What kind of volatility models? $\frac{\mathrm{d}S_t}{S_t} = \sigma_t \, \mathrm{d}W_t$

- Simplest rough model: $\sigma_t = \exp(B_t^H)$;
- Rough SDE: $\mathrm{d}\sigma_t = (\cdots)\,\mathrm{d}t + (\cdots)\,\mathrm{d}B_t^H$ (Comte-Renault (1998), El Euch-Rosenbaum (2019), Horvath-Jacquier-Lacombe (2019))
- More natural: $\sigma_t = \exp(X_t)$, for some Gaussian Volterra process

$$X_t = \int_0^t \mathrm{K}(t-s)^\top \cdot \mathrm{d}B_s,$$

  with B is a (possibly multidimensional) Brownian motion and K a (singular) kernel such that $\int_0^t \|\mathrm{K}(t-s)\|^2 ds < \infty$ (Bennedsen-Lunde-Pakkanen (2017));
  - Power-law kernel;
  - Gamma kernel;
  - fBm kernel;
- BSS process: $\sigma_t = \exp(X_t)$, with $X_t = \int_0^t \sqrt{\Gamma_s}\mathrm{K}(t-s)^\top \mathrm{d}B_s$, where Γ is a strictly positive affine process independent of B (Horvath-Jacquier-Tankov (2019));

That said, practitioners' authoritative choice: local stochastic volatility models!!

Rough local stochastic volatility models (I)

$$
\begin{aligned}
S_t &= S_0 + \int_0^t l(r, S_r, V_r) S_r \mathrm{d}W_r, \\
V_t &= V_0 + \int_0^t \mathrm{K}(t-r)\Big(b(V_r)\mathrm{d}r + \xi(V_r)\mathrm{d}B_r\Big), \\
\mathrm{d}\left\langle W, B \right\rangle_t &= \rho \, \mathrm{d}t,
\end{aligned}
$$

where $V = \sigma^2$. Setting $\mathbf{X} = (S, V)$, we can rewrite the system as

$$
\mathbf{X}_t = \mathbf{X}_0 + \int_0^t \mathbf{b}(t, r, \mathbf{X}_r)\mathrm{d}r + \int_0^t \boldsymbol{\sigma}(t, r, \mathbf{X}_r) \cdot \mathrm{d}\mathbf{B}_r,
$$

where

$$
\mathbf{b}(t, r, \mathbf{X}_r) = \begin{pmatrix} 0 \\ \mathrm{K}(t-r)b(V_r) \end{pmatrix} \quad \text{and} \quad \boldsymbol{\sigma}(t, r, \mathbf{X}_r) = \begin{pmatrix} \overline{\rho}\, l(t, S_r, V_r)S_r & \rho\, l(t, S_r, V_r)S_r \\ 0 & \mathrm{K}(t-r)\xi(V_r) \end{pmatrix},
$$

with $\overline{\rho} := \sqrt{1-\rho^2}$, $\mathbf{B} = (B^\perp, B)$, and $W := \rho B + \overline{\rho} B^\perp$.

Rough local stochastic volatility models (II)

- The function $l(\cdot)$ is of the form $l(t, S, v) = \mathfrak{L}(t, S)\varsigma(v)$, where $\mathfrak{L}(\cdot, \cdot)$ is the *leverage function*. From Dupire and Gyöngy, to ensure that this model calibrates exactly to European option prices, we require

$$\sigma_{\mathrm{loc}}^2(t, s) = \mathbb{E}\left[l(t, S_t, V_t)^2 | S_t = s\right] = \mathfrak{L}(t, s)^2 \mathbb{E}\left[\varsigma(V_t)^2 | S_t = s\right]$$

for every $t, s \geq 0$, where σ_{loc} is the so-called *local volatility*, inferred directly from European option prices; inverting this yields

$$\mathfrak{L}(t, s) = \frac{\sigma_{\mathrm{loc}}(t, s)}{\sqrt{\mathbb{E}\left[\varsigma(V_t)^2 | S_t = s\right]}}.$$

Rough local stochastic volatility models (II)

- The function $l(\cdot)$ is of the form $l(t,S,v) = \mathfrak{L}(t,S)\varsigma(v)$, where $\mathfrak{L}(\cdot,\cdot)$ is the *leverage function*. From Dupire and Gyöngy, to ensure that this model calibrates exactly to European option prices, we require

$$\sigma^2_{\mathrm{loc}}(t,s) = \mathbb{E}\left[l(t,S_t,V_t)^2|S_t = s\right] = \mathfrak{L}(t,s)^2\mathbb{E}\left[\varsigma(V_t)^2|S_t = s\right]$$

  for every $t, s \geq 0$, where σ_{loc} is the so-called *local volatility*, inferred directly from European option prices; inverting this yields

$$\mathfrak{L}(t,s) = \frac{\sigma_{\mathrm{loc}}(t,s)}{\sqrt{\mathbb{E}\left[\varsigma(V_t)^2|S_t = s\right]}}.$$

Important Questions:

- Existence of the denominator?
- Is the stock price a martingale?

Rough local stochastic volatility models (II)

- The function $l(\cdot)$ is of the form $l(t,S,v) = \mathfrak{L}(t,S)\varsigma(v)$, where $\mathfrak{L}(\cdot,\cdot)$ is the *leverage function*. From Dupire and Gyöngy, to ensure that this model calibrates exactly to European option prices, we require

$$\sigma_{\mathrm{loc}}^2(t,s) = \mathbb{E}\left[l(t,S_t,V_t)^2 | S_t = s\right] = \mathfrak{L}(t,s)^2\mathbb{E}\left[\varsigma(V_t)^2 | S_t = s\right]$$

for every $t, s \geq 0$, where σ_{loc} is the so-called *local volatility*, inferred directly from European option prices; inverting this yields

$$\mathfrak{L}(t,s) = \frac{\sigma_{\mathrm{loc}}(t,s)}{\sqrt{\mathbb{E}\left[\varsigma(V_t)^2 | S_t = s\right]}}.$$

Important Questions:

- Existence of the denominator?
- Is the stock price a martingale?

Answer: Yes, under some assumptions (boundedness of the coefficients,...).

Evaluating options on $(S_t)_{t \geq 0}$?

- **Goal**: $P\left(t, S_t, (\Theta_u^t)_{t \leq u \leq T}\right)$.
- PDE / Finite differences: ruled out by absence of Markovianity.
- Fourier transform: usually not available.
- Monte Carlo: yes but slow (recent improvements, though).

Evaluating options on $(S_t)_{t\geq 0}$?

- **Goal**: $P\left(t, S_t, (\Theta^t_u)_{t\leq u\leq T}\right)$.
- PDE / Finite differences: ruled out by absence of Markovianity.
- Fourier transform: usually not available.
- Monte Carlo: yes but slow (recent improvements, though).

Goal: solve via path-dependent PDEs and propose a numerical scheme.

Pricing formalism

Lemma. $P_t = \mathbb{E}[g(S_T)|\mathcal{F}_t]$

There exists a function $P : [0, T] \times [0, \infty) \times \mathcal{C}([0, T] \to \mathbb{R})$, such that, for any $t \in [0, T]$,

$$P_t = P\left(t, S_t, (\Theta_u^t)_{t \le u \le T}\right),$$

where for any $u \in [t, T]$,

$$\Theta_u^t := \mathbb{E}\left[V_u - \int_t^u \mathrm{K}(u - r) b(r, V_r) \mathrm{d}r \middle| \mathcal{F}_t\right] \in \mathcal{F}_t.$$

Note in particular that $\Theta_t^t = V_t$.

Notation:

$$\mathrm{K}^t := \mathrm{K}(\cdot - t), \qquad \text{for any } t \ge 0,$$

is the curve K^t seen at time t.

Pricing PPDE

Theorem: Pricing PPDE

The option price is the unique solution to the linear path-dependent PDE

$$\Big(\partial_t + \mathcal{L}_{xx} + \mathcal{L}_{x\omega} + \mathcal{L}_{\omega} + \mathcal{L}_{\omega\omega}\Big) P(t, S_t, \Theta^t) = 0,$$

for $t \in [0, T)$, with boundary condition $P(T, S_T, \Theta^T) = g(S_T)$, where

$$\begin{aligned} \mathcal{L}_{x\omega} &:= \rho l(t,x,v)\xi(\Theta_t^t)x\langle \partial_{x,\omega}, \mathrm{K}^t\rangle, & \mathcal{L}_{xx} &:= \frac{1}{2} l(t,x,\Theta_t^t)^2 x^2 \partial_x^2, \\ \mathcal{L}_{\omega\omega} &:= \frac{1}{2}\xi(\Theta_t^t)^2 \langle \partial_\omega^2, (\mathrm{K}^t, \mathrm{K}^t)\rangle, & \mathcal{L}_{\omega} &:= b(\Theta_t^t)\langle \partial_\omega, \mathrm{K}^t\rangle. \end{aligned}$$

- Nice theoretically, but very few algorithms to solve it (Zhang-Zuo (2014), Ren-Tan (2017)).
- Existence and uniqueness guaranteed (Ekren-Touzi-Zhang (2016)).
- Meaning of the derivatives?

A short detour through functional Itô

$$\mathbf{X}_t = \mathbf{X}_0 + \int_0^t \mathbf{b}(t, r, \mathbf{X}_\cdot)\mathrm{d}r + \int_0^t \boldsymbol{\sigma}(t, r, \mathbf{X}_\cdot) \cdot \mathrm{d}\mathbf{W}_r, \tag{1}$$

where $\mathbf{X}_0 \in \mathbb{R}^d$, $\mathbf{W}$ is a Brownian motion on $\mathbb{R}^n$, and $\mathbf{b} : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ and $\boldsymbol{\sigma} : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ satisfy

Assumptions:

- $\mathbf{b}$ and $\boldsymbol{\sigma}$ are adapted, and $\partial_t \mathbf{b}$ and $\partial_t \boldsymbol{\sigma}$ exist. For $\boldsymbol{\varphi} \in \{\mathbf{b}, \boldsymbol{\sigma}, \partial_t \mathbf{b}, \partial_t \boldsymbol{\sigma}\}$, $|\boldsymbol{\varphi}(t, s, \omega)| \leq C\left(1 + \|\omega\|_T^a\right)$ for some $C, a > 0$.
- (1) admits a weak solution and $\mathbb{E}[\sup_{t\in[0,T]} |\mathbf{X}_t|^p] < \infty$ for all $p \geq 1$.
- The following hold:
  - (i) (Regular case) For any $s \in [0, T]$, $\partial_t \mathbf{b}(t, s, \cdot)$ and $\partial_t \boldsymbol{\sigma}(t, s, \cdot)$ exist on $[s, T]$, and for $\boldsymbol{\varphi} \in \{\mathbf{b}, \boldsymbol{\sigma}, \partial_t \mathbf{b}, \partial_t \boldsymbol{\sigma}\}$,
  $$|\boldsymbol{\varphi}(t, s, \omega)| \leq C\left(1 + \|\omega\|_T^a\right), \qquad \text{for some } a, C > 0;$$
  - (ii) (Singular case) Let $\boldsymbol{\varphi} \in \{\mathbf{b}, \boldsymbol{\sigma}\}$. For any $s \in [0, T]$, $\partial_t \boldsymbol{\varphi}(t, s, \cdot)$ exists on $(s, T]$, and there exists $h \in \left(0, \frac{1}{2}\right)$ such that, for some $a, C > 0$,
  $$|\boldsymbol{\varphi}(t, s, \omega)| \leq C\left(1 + \|\omega\|_T^a\right)(t-s)^{h-\frac{1}{2}} \quad \text{and} \quad |\partial_t \boldsymbol{\varphi}(t, s, \omega)| \leq C\left(1 + \|\omega\|_T^a\right)(t-s)^{h-\frac{3}{2}}.$$

For any $0 \leq t \leq u$, decompose $\mathbf{X}$ as

$$\mathbf{X}_u = \underbrace{\mathbf{X}_0 + \int_0^t (\cdots)\mathrm{d}r + \int_0^t (\cdots)\mathrm{d}\mathbf{W}_r}_{\Theta_u^t \in \mathcal{F}_t} + \underbrace{\int_t^u (\cdots)\mathrm{d}r + \int_t^u (\cdots)\mathrm{d}\mathbf{W}_r}_{I_u^t \notin \mathcal{F}_t}.$$

Denote the concatenation of the paths $\mathbf{X}$ and Θ^t before and after time t,

$$\left(\mathbf{X} \otimes^t \Theta^t\right)_u := \mathbf{X}_u \, \mathbb{1}_{\{0<u<t\}} + \Theta_u^t \, \mathbb{1}_{\{t<u<T\}}, \qquad \text{for any } u, t \in [0, T].$$

Now, the space derivatives are defined as Gâteaux derivatives:

$$\langle \partial_\omega u(t,\omega), \eta \rangle := \lim_{\varepsilon \downarrow 0} \frac{u(t, \omega + \varepsilon \eta_{[t,T]}) - u(t,\omega)}{\varepsilon}, \qquad \text{for any } \eta \in \mathcal{C}_t,$$

$$\left\langle \partial_\omega^2 u(t,\omega), (\eta,\zeta) \right\rangle := \lim_{\varepsilon \downarrow 0} \frac{\left\langle \partial_\omega u(t, \omega + \varepsilon \eta_{[t,T]}), \zeta \right\rangle - \langle \partial_\omega u(t,\omega), \zeta \rangle}{\varepsilon}, \qquad \text{for any } \eta, \zeta \in \mathcal{C}_t.$$

Functional Itô Theorem (Viens-Zhang)

For $t \in [0, T]$, let $\mathbf{Z}^t := \mathbf{X} \otimes^t \Theta^t$, and $\boldsymbol{\varphi}^{t,\omega} := \boldsymbol{\varphi}(\cdot, t, \omega)$ for $\boldsymbol{\varphi} \in \{\mathbf{b}, \boldsymbol{\sigma}\}$. Then

$$\mathrm{d}u\left(t, \mathbf{Z}^t\right) = \left[\partial_t u\left(t, \mathbf{Z}^t\right) + \left\langle \partial_\omega u\left(t, \mathbf{Z}^t\right), \mathbf{b}^{t,\mathbf{X}} \right\rangle + \frac{1}{2}\left\langle \partial_\omega^2 u\left(t, \mathbf{Z}^t\right), \left(\boldsymbol{\sigma}^{t,\mathbf{X}}, \boldsymbol{\sigma}^{t,\mathbf{X}}\right) \right\rangle\right] \mathrm{d}t$$
$$+ \left\langle \partial_\omega u\left(t, \mathbf{Z}^t\right), \boldsymbol{\sigma}^{t,\mathbf{X}} \right\rangle \mathrm{d}\mathbf{W}_t,$$

Discretisation of $(\partial_t + \mathcal{L}_{xx} + \mathcal{L}_{x\omega} + \mathcal{L}_{\omega} + \mathcal{L}_{\omega\omega})P = 0$

Discretisation of $(\partial_t + \mathcal{L}_{xx} + \mathcal{L}_{x\omega} + \mathcal{L}_{\omega} + \mathcal{L}_{\omega\omega})P = 0$

For each $t \in [0, T]$, consider a basis of càdlàg functions $(\psi_i^t)_{i=1,\ldots,n}$, and approximate

$$\widehat{\Theta}^t := \sum_{i=1}^n \theta_i^t \psi_n^t \qquad \text{and} \qquad \widehat{\mathrm{K}}^t := \sum_{i=1}^n \kappa_i^t \psi_n^t.$$

Since

$$\left\langle \partial_\omega P\left(t, x, \Theta^t\right), \mathrm{K}^t \right\rangle := \partial_\varepsilon P\left(t, x, \Theta^t + \varepsilon \mathrm{K}^t\, \mathbb{1}_{[t,T]}\right)\Big|_{\varepsilon=0} = \partial_\varepsilon P\left(t, x, \Theta^t + \varepsilon \mathrm{K}^t\right)\Big|_{\varepsilon=0},$$

then introduce the approximation

$$\begin{aligned}\left\langle \partial_\omega P(t, x, \widehat{\Theta}^t), \widehat{\mathrm{K}}^t \right\rangle &:= \partial_\varepsilon P(t, x, \widehat{\Theta}^t + \varepsilon \widehat{\mathrm{K}}^t)|_{\varepsilon=0} = \partial_\varepsilon P\left(t, x, \sum_{i=1}^n \left(\theta_n^t + \varepsilon \kappa_n^t\right)\psi_n\right)\Bigg|_{\varepsilon=0} \\ &=: \partial_\varepsilon \widehat{P}\left(t, x, \left(\theta_n^t + \varepsilon\kappa_n^t\right)_{i=1}^n\right)\Big|_{\varepsilon=0} = \nabla_{\boldsymbol{\theta}^t} \widehat{P}\left(t, x, \boldsymbol{\theta}^t\right) \cdot \boldsymbol{\kappa}^t, \\ \left\langle \partial_\omega^2 \left(t, x, \widehat{\Theta}^t\right), \left(\widehat{\mathrm{K}}^t, \widehat{\mathrm{K}}^t\right)\right\rangle &:= \left(\boldsymbol{\kappa}^t\right)^\top \cdot \Delta_{\boldsymbol{\theta}^t} \widehat{P}\left(t, x, \boldsymbol{\theta}^t\right) \cdot \boldsymbol{\kappa}, \\ \left\langle \partial_{x,\omega}\left(t, x, \widehat{\Theta}^t\right), \widehat{\mathrm{K}}^t\right\rangle &:= \partial_x \nabla_{\boldsymbol{\theta}^t} \widehat{P}\left(t, x, \boldsymbol{\theta}^t\right) \cdot \boldsymbol{\kappa}^t.\end{aligned}$$

The PPDE becomes a (very) high-dimensional PDE:

$$\left(\partial_t + \mathcal{L}_{xx} + \sum_{i=1}^n \mathcal{L}_{x\theta_i^t} + \sum_{i=1}^n \mathcal{L}_{\theta_i^t} + \sum_{i,j=1}^n \mathcal{L}_{\theta_i^t \theta_j^t}\right)\widehat{P} = 0.$$

Discretisation of the PPDE

$$\left(\partial_t + \mathcal{L}_{xx} + \sum_{i=1}^{n} \mathcal{L}_{x\theta_i^t} + \sum_{i=1}^{n} \mathcal{L}_{\theta_i^t} + \sum_{i,j=1}^{n} \mathcal{L}_{\theta_i^t\theta_j^t}\right) \widehat{P} = 0,$$

where the differential operators are defined, for each $i, j = 1, \dots, n$, as

$$\begin{aligned} \mathcal{L}_{x\theta_i} &:= \rho\, l(t,x,\boldsymbol{\theta}^t)\xi(\boldsymbol{\theta}^t)\kappa_i^t \partial_{x\theta_i^t}, & \mathcal{L}_{xx} &:= \frac{l(t,x,\boldsymbol{\theta}^t)^2}{2}\partial_x^2, \\ \mathcal{L}_{\theta_i^t\theta_j^t} &:= \frac{\xi(\boldsymbol{\theta}^t)^2}{2}\kappa_i^t\kappa_j^t \partial_{\theta_i^t\theta_j^t}, & \mathcal{L}_{\theta_i} &:= b(\boldsymbol{\theta}^t)\kappa_i^t \partial_{\theta_i^t}. \end{aligned}$$

We can rewrite this system in a more concise way as

$$\partial_t \widehat{P} + \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\Sigma}\cdot\boldsymbol{\Sigma}^\top \cdot \Delta\widehat{P}\right) + \boldsymbol{\mu}\cdot\nabla\widehat{P} = 0,$$

where $\boldsymbol{\mu}(t,x,\boldsymbol{\theta}^t) := \left(0, b(\boldsymbol{\theta}^t)\kappa_1^t, \dots, b(\boldsymbol{\theta}^t)\kappa_n^t\right)^\top$ and

$$\left(\boldsymbol{\Sigma}\cdot\boldsymbol{\Sigma}^\top\right)(t,x,\boldsymbol{\theta}^t) := \begin{pmatrix} \frac{1}{2}l(t,x,\boldsymbol{\theta}^t)^2 & \rho l(t,x,\boldsymbol{\theta}^t)\xi(\boldsymbol{\theta}^t)\kappa_1^t & \cdots & \rho l(t,x,\boldsymbol{\theta}^t)\xi(\boldsymbol{\theta}^t)\kappa_n^t \\ \rho\, l(t,x,\boldsymbol{\theta}^t)\xi(\boldsymbol{\theta}^t)\kappa_1^t & \xi(\boldsymbol{\theta}^t)^2(\kappa_1^t)^2 & \cdots & \xi(\boldsymbol{\theta}^t)^2\kappa_1^t\kappa_n^t \\ \vdots & \vdots & \ddots & \vdots \\ \rho\, l(t,x,\boldsymbol{\theta}^t)\xi(\boldsymbol{\theta}^t)\kappa_n^t & \xi(\boldsymbol{\theta}^t)^2\kappa_1^t\kappa_n^t & \cdots & \xi(\boldsymbol{\theta}^t)^2\left(\kappa_n^t\right)^2 \end{pmatrix}$$

Easy Peasy?

$$\partial_t \widehat{P} + \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\Sigma} \cdot \boldsymbol{\Sigma}^\top \cdot \Delta \widehat{P}\right) + \boldsymbol{\mu} \cdot \nabla \widehat{P} = 0,$$

Easy Peasy?

$$\partial_t \widehat{P} + \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\Sigma} \cdot \boldsymbol{\Sigma}^\top \cdot \Delta \widehat{P}\right) + \boldsymbol{\mu} \cdot \nabla \widehat{P} = 0,$$

- Very high-dimensional PDE;
- Curse of dimensionality;
- Neural network is an old idea, but recent numerics show surprisingly good results;
- Mathematically, universal approximation theorems show that a wide class of functions on compact subsets can be approximated by single hidden layer neural network;
- More work needed for multi-layer NN.

The BSDE formulation and simulation

$$\partial_t \widehat{P} + \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\Sigma}\cdot\boldsymbol{\Sigma}^\top\cdot\Delta\widehat{P}\right) + \boldsymbol{\mu}\cdot\nabla\widehat{P} = 0,$$

In the classical BSDE literature, one can associate a BSDE solution (X, Y, Z) to a given PDE. Here, one can show that $Y = \widehat{P}$ is an element of the BSDE; written in backward form, for any $t \in [0, T]$:

$$\widehat{P}\left(t, S_t, \boldsymbol{\theta}^t\right) = \widehat{P}\left(T, S_T, \boldsymbol{\theta}^T\right) - \int_t^T \boldsymbol{\Sigma}\left(u, S_u, \boldsymbol{\theta}^u\right)^\top \cdot \nabla\widehat{P}\left(u, S_u, \boldsymbol{\theta}^u\right)\mathrm{d}W_u,$$

Discretising (Euler) the backward SDE gives

$$\begin{cases} \widehat{P}\left(t_n, S_{t_n}, \boldsymbol{\theta}^{t_n}\right) & = & g(S_T), \\ \widehat{P}\left(t_i, S_{t_i}, \boldsymbol{\theta}^{t_i}\right) & = & \widehat{P}\left(t_{i+1}, S_{t_{1+1}}, \boldsymbol{\theta}^{t_{i+1}}\right) - \boldsymbol{\Sigma}\left(t_i, S_{t_i}, \boldsymbol{\theta}^{t_i}\right)^\top \nabla\widehat{P}\left(t_i, S_{t_i}, \boldsymbol{\theta}^{t_i}\right)\Delta W_{t_i}. \end{cases}$$

Simulation of the Volterra process is performed using the hybrid scheme (Bennedsen-Pakkanen-Lunde (2017)).

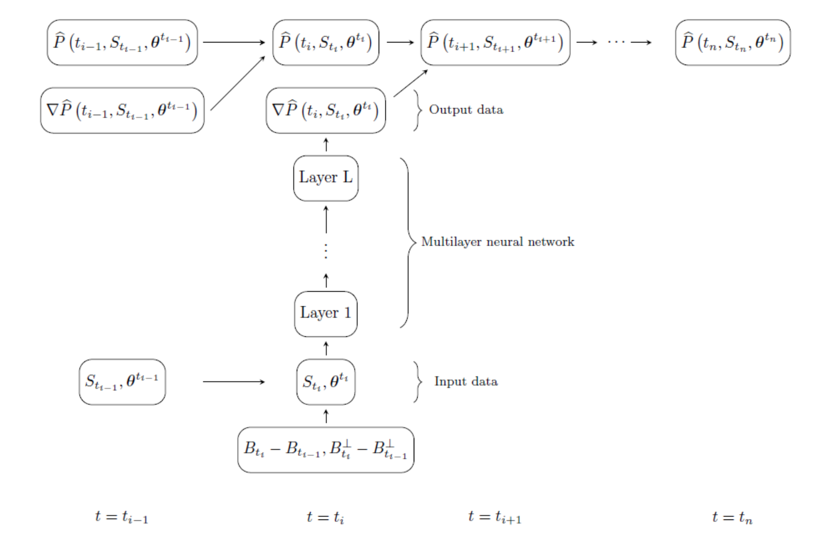
Computing via neural network (Han-Jentzen)

- At each time step, price and model parameters are known.
- Unknown: $\mathbf{\Sigma}\left(t_i, S_{t_i}, \boldsymbol{\theta}^{t_i}\right)^\top \nabla\widehat{P}\left(t_i, S_{t_i}, \boldsymbol{\theta}^{t_i}\right)$. We use a NN to infer its value.
- n_l: number of layers per sub-network;
- n_N: number of neurons per layer;

Loss function to minimise:

$$\widehat{\mathrm{L}}\left(\mathbf{w}, \boldsymbol{\delta}, \beta, \gamma, \widehat{P}_0, (\nabla\widehat{P})_0\right) := \frac{1}{N}\sum_{k=1}^{N}\left|g\left(S_{t_n}^k\right) - \widehat{P}\left((t_i)_{1\leq i\leq n}, \left(S_{t_i}^k\right)_{1\leq i\leq n}, (\boldsymbol{\theta}^{t_i})_{1\leq i\leq n}^k\right)\right|^2.$$

Network structure



Example

We consider

$$
\begin{aligned}
S_t &= S_0 + \int_0^t \mathfrak{L}(r, S_r)\varsigma(V_r), S_r \mathrm{d}W_r,\\
V_t &= V_0 + \int_0^t \mathrm{K}(t-r)\Big(b(V_r)\mathrm{d}r + \xi(V_r)\mathrm{d}B_r\Big),
\end{aligned}
$$

with $b(V) = \kappa(V_\infty - V), \qquad \xi(V) = \nu V, \qquad \varsigma(V) = \mathrm{e}^V, \qquad \mathrm{K}(t) = \dfrac{t^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)}.$

Regarding the neural network, we use the default configuration

- Two hidden layers;
- 5 neurons per hidden layer;
- Between 100 and $2,600$ time steps per year depending on the maturity;
- Between 20 and 520 BSDE time steps per year depending on the maturity.

Parameter values:

$$
S_0 = 100, \quad T = 0.1, \quad V_0 = \log(0.15), \quad \nu = 0.2, \quad V_\infty = 0.15, \quad \kappa = 0.5, \quad \rho = -0.4.
$$

Maturity: 1 week				
Strike	0.95	0.98	1.02	1.05
MC Price	15.09	15.02	14.96	14.94
DL Price	15.12	15.06	14.99	14.91

Maturity: 1 month						
Strike	0.89	0.93	0.98	1.02	1.07	1.11
MC Price	15.18	15.11	15.05	14.99	14.92	14.85
DL Price	15.29	15.	15.04	14.98	14.90	14.84

Maturity: 6 months								
Strike	0.77	0.83	0.9	0.97	1.00	1.1	1.17	1.23
MC Price	15.37	15.24	15.16	15.03	14.93	14.85	14.76	14.72
DL Price	15.43	15.28	15.16	15.06	14.95	14.89	14.82	14.74

Maturity: 1 year								
Strike	0.72	0.83	0.94	1.06	1.17	1.28	1.39	1.5
MC Price	15.49	15.29	15.11	14.96	14.81	14.67	14.55	14.44
DL Price	15.50	15.31	15.12	14.89	14.83	14.71	14.61	14.53

Maturity: 2 years								
Strike	0.72	0.83	0.94	1.06	1.17	1.28	1.39	1.5
MC Price	15.49	15.26	15.10	14.98	14.84	14.72	14.60	14.51
DL Price	15.51	15.34	15.17	15.02	14.89	14.78	14.72	14.60

References

- M. Bennedsen, A. Lunde, M.S. Pakkanen. Hybrid scheme for Brownian semistationary processes. Finance and Stochastics, 2017.
- B. Dupire. Functional Itô Calculus. 2009.
- I. Ekren, N. Touzi, J. Zhang. Viscosity solutions of fully nonlinear parabolic path-dependent PDEs: Part I. Annals of Probability, 2016.
- I. Ekren, N. Touzi, J. Zhang. Viscosity solutions of fully nonlinear parabolic path-dependent PDEs: Part II. Annals of Probability, 2016.
- J. Gatheral, T. Jaisson, M. Rosenbaum. Volatility is rough. Quant. Fin., 2018.
- J. Han, A. Jentzen. Solving high-dimensional partial differential equations using deep learning. Proceedings of the Natural Academy of Sciences, 2018.
- A. Jacquier, M. Oumgari. Deep PPDEs for rough local stochastic volatility, 2019.
- Z. Ren, X. Tan. On the convergence of monotone schemes for path-dependent PDEs. Stochastic Processes and their Applications, 2017.
- F. Viens, J. Zhang. A martingale approach for fractional Brownian motions and related path-dependent PDEs, Annals of Applied Probability, 2019.
- J. Zhang, J. Zhuo. Monotone schemes for fully nonlinear parabolic path dependent PDEs. Journal of Financial Engineering, 2014.