Robust SDE-Based Variational Formulations for Solving Linear PDEs via Deep Learning

Lorenz Richter, Julius Berner

(equal contribution)

ICML

July, 2022

Scope: Linear Kolmogorov PDEs

- We want to solve PDEs (*partial differential equations*) of the form

$$\left(\partial_t + \frac{1}{2}(\sigma\sigma^\top) : \nabla^2 + b \cdot \nabla\right) V(x,t) = 0, \qquad V(x,T) = g(x), \qquad (x,t) \in \mathbb{R}^d \times [0,T].$$

Scope: Linear Kolmogorov PDEs

- We want to solve PDEs (*partial differential equations*) of the form

$$\left(\partial_t + \frac{1}{2}(\sigma\sigma^\top) : \nabla^2 + b \cdot \nabla\right) V(x,t) = 0, \qquad V(x,T) = g(x), \qquad (x,t) \in \mathbb{R}^d \times [0,T].$$

- **Applications:** modelling of diffusion processes in physics, pricing of financial derivatives, reinforcement learning, diffusion-based generative modeling, ...

Scope: Linear Kolmogorov PDEs

- We want to solve PDEs (*partial differential equations*) of the form

$$\left(\partial_t + \frac{1}{2}(\sigma\sigma^\top) : \nabla^2 + b \cdot \nabla\right) V(x,t) = 0, \qquad V(x,T) = g(x), \qquad (x,t) \in \mathbb{R}^d \times [0,T].$$

- **Applications:** modelling of diffusion processes in physics, pricing of financial derivatives, reinforcement learning, diffusion-based generative modeling, ...

- **Idea:** minimize *variational formulations* using neural networks $u_\theta \in \mathcal{U}$ with parameters θ, i.e. consider losses

$$\mathcal{L} : \mathcal{U} \to \mathbb{R}_{\geq 0},$$

which shall be minimal iff $u \in \mathcal{U}$ fulfills the PDE.

SDE-Based Variational Formulations

- **Question:** which loss do we choose in practice?

SDE-Based Variational Formulations

- **Question:** which loss do we choose in practice?
- **Stochastic representation:** *Itô calculus* (cf. Feynman-Kac formula) shows

$$\underbrace{g(X_T) - V(\xi, \tau)}_{:=\Delta_V} - \underbrace{\int_\tau^T \sigma(X_s)^\top \nabla V(X_s, s) \cdot \mathrm{d}W_s}_{:=S_V} = 0,$$

where X is the solution to the SDE (*stochastic differential equation*)

$$\mathrm{d}X_s = b(X_s)\,\mathrm{d}s + \sigma(X_s)\,\mathrm{d}W_s, \quad X_\tau = \xi.$$

SDE-Based Variational Formulations

- **Question:** which loss do we choose in practice?
- **Stochastic representation:** *Itô calculus* (cf. Feynman-Kac formula) shows

$$\underbrace{g(X_T) - V(\xi,\tau)}_{:=\Delta_V} - \underbrace{\int_\tau^T \sigma(X_s)^\top \nabla V(X_s,s) \cdot \mathrm{d}W_s}_{:=S_V} = 0,$$

  where X is the solution to the SDE (*stochastic differential equation*)

$$\mathrm{d}X_s = b(X_s)\,\mathrm{d}s + \sigma(X_s)\,\mathrm{d}W_s, \quad X_\tau = \xi.$$

- As the stochastic integral S_u has vanishing expectation, this motivates the two losses

$$\mathcal{L}_{\mathrm{FK}}(u) := \mathbb{E}\left[\Delta_u^2\right] \qquad \text{and} \qquad \mathcal{L}_{\mathrm{BSDE}}(u) := \mathbb{E}\left[(\Delta_u - S_u)^2\right],$$

  where $(\xi,\tau) \sim \mathrm{Unif}(\mathbb{R}^d \times [0,T])$.

Robustness of the Losses

Robustness of the Losses

- The stochastic integral S_u in $\mathcal{L}_{\mathrm{BSDE}}$ can be interpreted as a *control variate*.

Robustness of the Losses

- The stochastic integral S_u in $\mathcal{L}_{\mathrm{BSDE}}$ can be interpreted as a *control variate*.
- It guarantees statistical advantages for the estimator versions $\mathcal{L}^{(K)}$ (with K samples) at the optimum $u_\theta = V$:

Robustness of the Losses

- The stochastic integral S_u in $\mathcal{L}_{\mathrm{BSDE}}$ can be interpreted as a *control variate*.
- It guarantees statistical advantages for the estimator versions $\mathcal{L}^{(K)}$ (with K samples) at the optimum $u_\theta = V$:

Proposition (Variance of Losses)

$$\mathbb{V}\left[\mathcal{L}_{\mathrm{FK}}^{(K)}(u_\theta)\right] = \frac{1}{K}\,\mathbb{V}\left[S_V^2\right] \qquad \textit{and} \qquad \mathbb{V}\left[\mathcal{L}_{\mathrm{BSDE}}^{(K)}(u_\theta)\right] = 0.$$

Robustness of the Losses

- The stochastic integral S_u in $\mathcal{L}_{\mathrm{BSDE}}$ can be interpreted as a *control variate*.
- It guarantees statistical advantages for the estimator versions $\mathcal{L}^{(K)}$ (with K samples) at the optimum $u_\theta = V$:

Proposition (Variance of Losses)

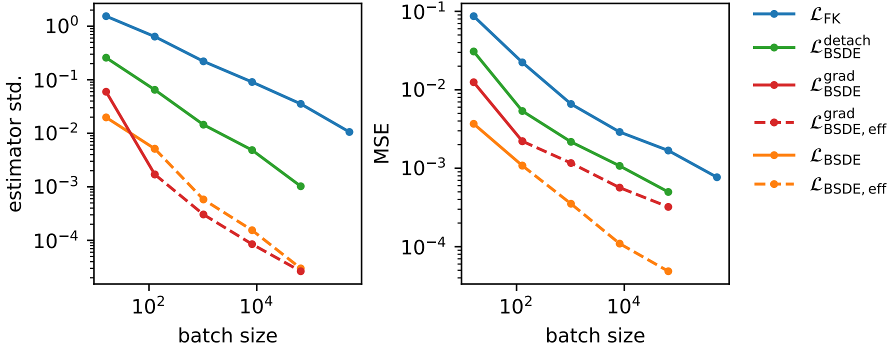
$$\mathbb{V}\left[\mathcal{L}_{\mathrm{FK}}^{(K)}(u_\theta)\right] = \frac{1}{K}\mathbb{V}\left[S_V^2\right] \qquad \textit{and} \qquad \mathbb{V}\left[\mathcal{L}_{\mathrm{BSDE}}^{(K)}(u_\theta)\right] = 0.$$

Proposition (Variance of Gradients)

$$\mathbb{V}\left[\nabla_\theta \mathcal{L}_{\mathrm{FK}}^{(K)}(u_\theta)\right] = \frac{4}{K}\mathbb{V}\left[S_V \nabla_\theta u_\theta(\xi,\tau)\right] \qquad \textit{and} \qquad \mathbb{V}\left[\nabla_\theta \mathcal{L}_{\mathrm{BSDE}}^{(K)}(u_\theta)\right] = 0.$$

- We propose various versions to include the control variate: $\mathcal{L}_{\mathrm{BSDE}}^{\mathrm{grad}}, \mathcal{L}_{\mathrm{BSDE}}^{\mathrm{detach}}$ and $\mathcal{L}_{\mathrm{BSDE,\,eff}}$.

- We propose various versions to include the control variate: $\mathcal{L}^{\mathrm{grad}}_{\mathrm{BSDE}}, \mathcal{L}^{\mathrm{detach}}_{\mathrm{BSDE}}$ and $\mathcal{L}_{\mathrm{BSDE,\,eff}}$.
- We improve state-of-the-art performance and analyze trade-offs between accuracy and complexity.



Thank you for your attention!

lorenz.richter@fu-berlin.de, julius.berner@univie.ac.at

Source code: `https://github.com/juliusberner/robust_kolmogorov`