# Robust SDE-based variational formulations for solving linear PDEs via deep learning

Lorenz Richter[⋆,1,2,3], Julius Berner[⋆,4]

[⋆] Equal contribution, [1] dida Datenschmiede GmbH, [2] Zuse Institute Berlin, [3] Freie Universität Berlin, [4] University of Vienna

## Solving Kolmogorov PDEs

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

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

- **Applications:** modelling of diffusion processes in physics, pricing of financial derivatives, diffusion-based generative modeling, reinforcement learning, ...
- **Idea:** minimize *variational formulations* using neural networks $u_\theta \in \mathcal{U}$ with parameters $\theta$, i.e., consider losses

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

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

## Stochastic Representations

- *Itô calculus* (cf. Feynman-Kac formula) shows that

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

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

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

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

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

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

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

## Robust 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 1** (Variance of Losses)**.**

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

$$\mathbb{V}\left[\mathcal{L}^{(K)}_{\mathrm{BSDE}}(u_\theta)\right] = 0.$$

**Proposition 2** (Variance of Gradients)**.**

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

$$\mathbb{V}\left[\nabla_\theta\mathcal{L}^{(K)}_{\mathrm{BSDE}}(u_\theta)\right] = 0.$$

- For $\mathcal{L}_{\mathrm{BSDE}}$ we can expect small variances also close to the solution $u_\theta \approx V$:

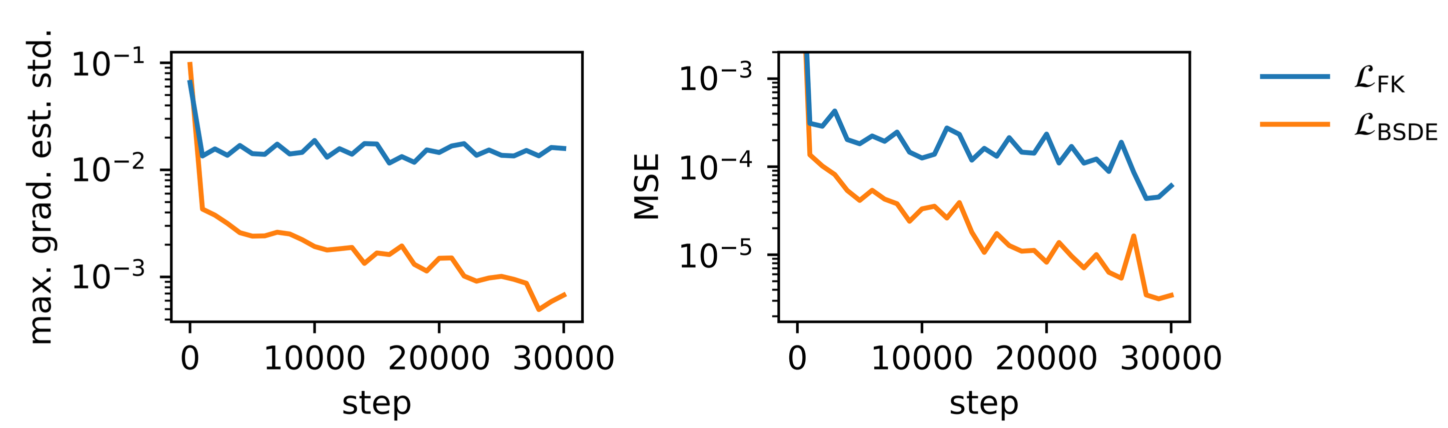
**Proposition 3** (Stability Close to Solution)**.** *Assume that*

$$|u_\theta(\xi,\tau) - V(\xi,\tau)| \leq \varepsilon, \quad \|\nabla_x(u_\theta - V)(x,t)\| \leq \varepsilon(1+\|x\|^\gamma),$$

*for some* $\gamma \in \mathbb{R}_{>0}$*. Then it holds that*

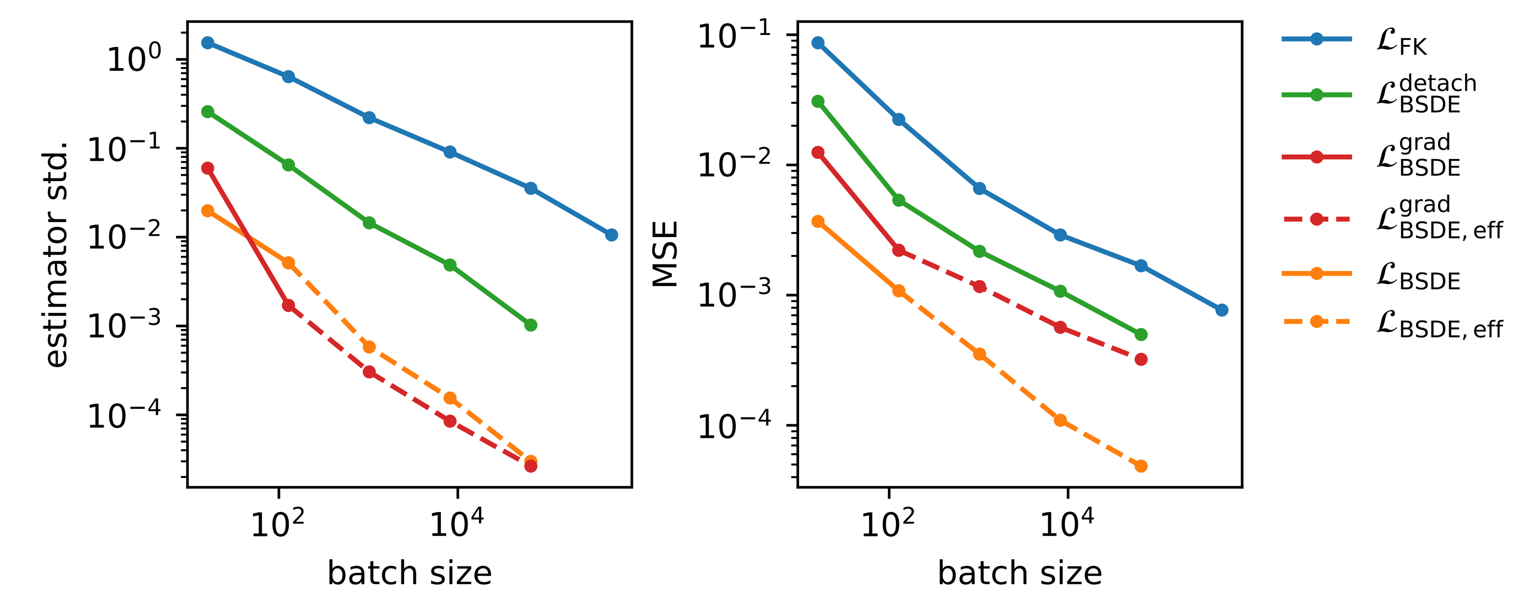
$$\mathbb{V}\left[\nabla_\theta\mathcal{L}^{(K)}_{\mathrm{BSDE}}(u_\theta)\right] \lesssim \frac{\varepsilon^2}{K}.$$

- This can also be observed empirically:

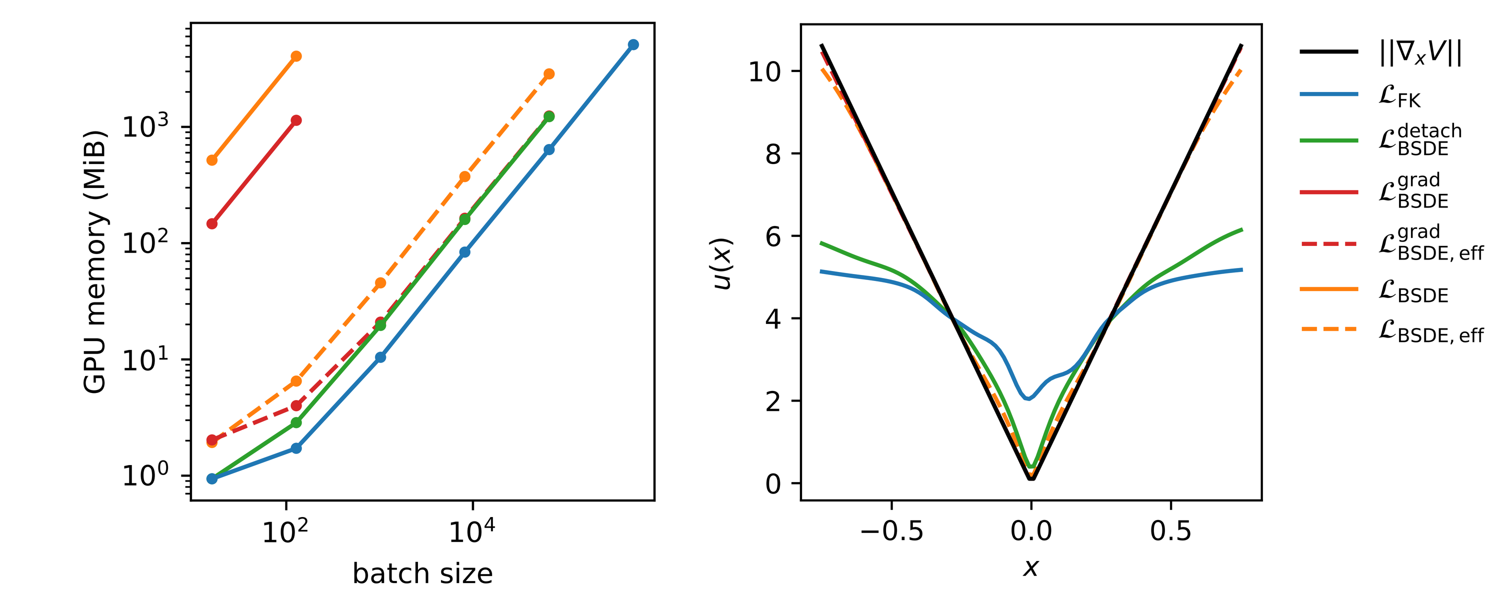



## Numerical Experiments

- We propose various (more efficient) 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:




- The efficient versions of the BSDE-based method combine accuracy and memory efficiency:




## Takeaways and Reference

Our paper provides:

- variational formulations for (linear) PDEs,
- techniques for analyzing the variance of (gradient) estimators,
- novel estimators with reduced variance,
- empirical studies of complexity vs. performance.

arxiv.org/abs/2206.10588
github.com/juliusberner/robust_kolmogorov