Error Analysis of Deep PDE Solvers for Option Pricing

12th General AMaMeF Conference

Jasper Rou

June 23, 2025

Support for this presentation was provided by G-Research.

Option Pricing

$$\frac{\partial u}{\partial t} - \nabla \cdot (A \nabla u) + \mathbf{b} \cdot \nabla u + ru = 0$$
$$u(0, \mathbf{x}) = \Psi(\mathbf{x})$$

Option Pricing

$$\frac{\partial u}{\partial t} - \nabla \cdot (A \nabla u) + \mathbf{b} \cdot \nabla u + ru = 0$$

$$u(0, \mathbf{x}) = \Psi(\mathbf{x})$$

$$f(t, \mathbf{x}; \theta) \approx u(t, \mathbf{x})$$

Black–Scholes

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sigma S_t\mathrm{d}W_t$$

Black–Scholes

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sigma S_t \mathrm{d}W_t$$

$$0 = \frac{\partial u}{\partial t} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 u}{\partial S^2} - rS\frac{\partial u}{\partial S} + ru$$

Black–Scholes

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sigma S_t\mathrm{d}W_t$$

$$\begin{aligned} 0 &= \frac{\partial u}{\partial t} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 u}{\partial S^2} - rS\frac{\partial u}{\partial S} + ru \\ &= \frac{\partial u}{\partial t} - \frac{\partial}{\partial S}\left(\frac{1}{2}\sigma^2 S^2 \frac{\partial u}{\partial S}\right) + \sigma^2 S \frac{\partial u}{\partial S} - rS\frac{\partial u}{\partial S} + ru \end{aligned}$$

Black–Scholes

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sigma S_t \mathrm{d}W_t$$

$$\begin{aligned}
0 &= \frac{\partial u}{\partial t} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 u}{\partial S^2} - rS\frac{\partial u}{\partial S} + ru \\
&= \frac{\partial u}{\partial t} - \frac{\partial}{\partial S}\left(\frac{1}{2}\sigma^2 S^2 \frac{\partial u}{\partial S}\right) + \sigma^2 S \frac{\partial u}{\partial S} - rS\frac{\partial u}{\partial S} + ru \\
&= \frac{\partial u}{\partial t} - \frac{\partial u}{\partial S}\left(\frac{1}{2}\sigma^2 S^2 \frac{\partial u}{\partial S}\right) + (\sigma^2 - r)S\frac{\partial u}{\partial S} + ru
\end{aligned}$$

Heston model

$$dS_t = rS_t dt + \sqrt{V_t} S_t dW_t \qquad S_0 > 0$$
$$dV_t = \kappa(\theta - V_t)dt + \eta\sqrt{V_t} dB_t \qquad V_0 > 0$$

Heston model

$$dS_t = rS_t dt + \sqrt{V_t} S_t dW_t \qquad S_0 > 0$$

$$dV_t = \kappa(\theta - V_t)dt + \eta\sqrt{V_t} dB_t \qquad V_0 > 0$$

$$0 = \frac{\partial u}{\partial t} - rS\frac{\partial u}{\partial S} - \kappa(\theta - V)\frac{\partial u}{\partial V} - \frac{1}{2}S^2 V\frac{\partial^2 u}{\partial S^2} - \frac{1}{2}\eta^2 V\frac{\partial^2 u}{\partial V^2} - \rho\eta SV\frac{\partial^2 u}{\partial S \partial V} + ru$$

Heston model

$$dS_t = rS_t dt + \sqrt{V_t} S_t dW_t \qquad S_0 > 0$$
$$dV_t = \kappa(\theta - V_t)dt + \eta\sqrt{V_t} dB_t \qquad V_0 > 0$$

$$\begin{aligned}
0 =& \frac{\partial u}{\partial t} - rS\frac{\partial u}{\partial S} - \kappa(\theta - V)\frac{\partial u}{\partial V} - \frac{1}{2}S^2 V\frac{\partial^2 u}{\partial S^2} - \frac{1}{2}\eta^2 V\frac{\partial^2 u}{\partial V^2} - \rho\eta SV\frac{\partial^2 u}{\partial S\partial V} + ru \\
=& \frac{\partial u}{\partial t} - rS\frac{\partial u}{\partial S} - \kappa(\theta - V)\frac{\partial u}{\partial V} - \frac{\partial}{\partial S}\left(\frac{1}{2}S^2 V\frac{\partial u}{\partial S}\right) + SV\frac{\partial u}{\partial S} \\
& - \frac{\partial}{\partial V}\left(\frac{1}{2}\eta^2 V\frac{\partial u}{\partial V}\right) + \frac{1}{2}\eta^2\frac{\partial u}{\partial V} - \frac{\partial}{\partial S}\left(\frac{1}{2}\rho\eta SV\frac{\partial u}{\partial V}\right) + \frac{1}{2}\rho\eta V\frac{\partial u}{\partial V} \\
& - \frac{\partial}{\partial V}\left(\frac{1}{2}\rho\eta SV\frac{\partial u}{\partial S}\right) + \frac{1}{2}\rho\eta S\frac{\partial u}{\partial S} + ru
\end{aligned}$$

Heston model

$$dS_t = rS_t dt + \sqrt{V_t} S_t dW_t \qquad S_0 > 0$$

$$dV_t = \kappa(\theta - V_t)dt + \eta\sqrt{V_t} dB_t \qquad V_0 > 0$$

$$\begin{aligned}
0 =& \frac{\partial u}{\partial t} - rS\frac{\partial u}{\partial S} - \kappa(\theta - V)\frac{\partial u}{\partial V} - \frac{1}{2}S^2 V\frac{\partial^2 u}{\partial S^2} - \frac{1}{2}\eta^2 V\frac{\partial^2 u}{\partial V^2} - \rho\eta SV\frac{\partial^2 u}{\partial S\partial V} + ru \\
=& \frac{\partial u}{\partial t} - rS\frac{\partial u}{\partial S} - \kappa(\theta - V)\frac{\partial u}{\partial V} - \frac{\partial}{\partial S}\left(\frac{1}{2}S^2 V\frac{\partial u}{\partial S}\right) + SV\frac{\partial u}{\partial S} \\
& - \frac{\partial}{\partial V}\left(\frac{1}{2}\eta^2 V\frac{\partial u}{\partial V}\right) + \frac{1}{2}\eta^2\frac{\partial u}{\partial V} - \frac{\partial}{\partial S}\left(\frac{1}{2}\rho\eta SV\frac{\partial u}{\partial V}\right) + \frac{1}{2}\rho\eta V\frac{\partial u}{\partial V} \\
& - \frac{\partial}{\partial V}\left(\frac{1}{2}\rho\eta SV\frac{\partial u}{\partial S}\right) + \frac{1}{2}\rho\eta S\frac{\partial u}{\partial S} + ru \\
=& \frac{\partial u}{\partial t} - \nabla\cdot\left(\frac{1}{2}\begin{bmatrix} S^2 V & \rho\eta SV \\ \rho\eta SV & \eta^2 V\end{bmatrix}\nabla u\right) + \begin{bmatrix} SV - rS + \frac{1}{2}\rho\eta S \\ \kappa(V-\theta) + \frac{1}{2}\rho\eta V + \frac{1}{2}\eta^2\end{bmatrix}\cdot\nabla u + ru
\end{aligned}$$

Deep **Galerkin** Method

$$\frac{\partial u}{\partial t} - \nabla \cdot (A \nabla u) + \mathbf{b} \cdot \nabla u + ru = 0$$

$$u(0, \mathbf{x}) = \Psi(\mathbf{x})$$

Deep **Galerkin** Method

$$\frac{\partial u}{\partial t} - \nabla \cdot (A \nabla u) + \mathbf{b} \cdot \nabla u + ru = 0$$
$$u(0, \mathbf{x}) = \Psi(\mathbf{x})$$

Minimize

$$\left\| \frac{\partial u}{\partial t} - \nabla \cdot (A \nabla u) + \mathbf{b} \cdot \nabla u + ru \right\|^2 + \| u(0, \mathbf{x}) - \Psi(\mathbf{x}) \|^2 .$$

Deep Galerkin Method

$$\left\| \frac{\partial u}{\partial t} - \nabla \cdot (A \nabla u) + \mathbf{b} \cdot \nabla u + ru \right\|^2 + \| u(0, \mathbf{x}) - \Psi(\mathbf{x}) \|^2 .$$

Deep Galerkin Method

$$\left\| \frac{\partial u}{\partial t} - \nabla \cdot (A \nabla u) + \mathbf{b} \cdot \nabla u + ru \right\|^2 + \| u(0, \mathbf{x}) - \Psi(\mathbf{x}) \|^2 .$$

$$\begin{aligned} L(\theta; t, \mathbf{x}) =& \frac{T}{M_1} \sum_{m=1}^{M_1} \left[\partial_t f(t, \mathbf{x}_m; \theta) - \nabla \cdot (A \nabla f(t, \mathbf{x}_m; \theta)) + \mathbf{b} \cdot \nabla f(t, \mathbf{x}_m; \theta) + r f(t, \mathbf{x}_m; \theta) \right]^2 \\ &+ \frac{1}{M_2} \sum_{m=1}^{M_2} \left[f(0, \mathbf{x}_m; \theta) - \Psi(\mathbf{x}_m) \right]^2 . \end{aligned}$$

Algorithm: DGM

Initialize θ_0.
for each sampling stage $n = 1, ..., N$ **do**
    Generate M random points $(t_m, \mathbf{x}_m)$ for training.
    Calculate the cost functional $L(\theta_n; t, \mathbf{x})$ for the selected points.
    Take a descent step $\theta_{n+1} = \theta_n - \alpha \nabla_\theta L(\theta_n; t, \mathbf{x})$.
end for

Time Deep Gradient Flow

$$\frac{\partial u}{\partial t} - \nabla \cdot (A \nabla u) + \mathbf{b} \cdot \nabla u + ru = 0$$
$$u(0, \mathbf{x}) = \Psi(\mathbf{x})$$

Time Deep Gradient Flow

$$\frac{\partial u}{\partial t} - \nabla \cdot (A \nabla u) + \mathbf{b} \cdot \nabla u + ru = 0$$
$$u(0, \mathbf{x}) = \Psi(\mathbf{x})$$

- Divide $[0, T]$ in K intervals $(t_{k-1}, t_k]$ with $h = t_k - t_{k-1}$, $U^0 = \Psi$

Time Deep Gradient Flow

$$\frac{\partial u}{\partial t} - \nabla \cdot (A \nabla u) + \mathbf{b} \cdot \nabla u + ru = 0$$
$$u(0, \mathbf{x}) = \Psi(\mathbf{x})$$

- Divide $[0, T]$ in K intervals $(t_{k-1}, t_k]$ with $h = t_k - t_{k-1}$, $U^0 = \Psi$

$$\frac{U^k - U^{k-1}}{h} \quad -\nabla \cdot \left(A \nabla U^k\right) \quad +\mathbf{b} \cdot \nabla U^{k-1} \quad +rU^k = 0$$

Time Deep Gradient Flow

$$\frac{\partial u}{\partial t} - \nabla \cdot (A \nabla u) + \mathbf{b} \cdot \nabla u + ru = 0$$
$$u(0, \mathbf{x}) = \Psi(\mathbf{x})$$

- Divide $[0, T]$ in K intervals $(t_{k-1}, t_k]$ with $h = t_k - t_{k-1}$, $U^0 = \Psi$

$$\frac{U^k - U^{k-1}}{h} - \nabla \cdot \left(A \nabla U^k\right) + \mathbf{b} \cdot \nabla U^{k-1} + rU^k = 0$$
$$\frac{\frac{3}{2} U^k - 2U^{k-2} + \frac{1}{2} U^{k-1}}{h} - \nabla \cdot \left(A \nabla U^k\right) + \mathbf{b} \cdot \left(2 \nabla U^{k-1} - \nabla U^{k-2}\right) + rU^k = 0$$

Time Deep **Gradient Flow**

$$\frac{U^k - U^{k-1}}{h} - \nabla \cdot \left(A \nabla U^k\right) + \mathbf{b} \cdot \nabla U^{k-1} + rU^k = 0$$

$$\frac{U^k - \frac{4}{3}U^{k-2} + \frac{1}{3}U^{k-1}}{\frac{2}{3}h} - \nabla \cdot \left(A \nabla U^k\right) + \mathbf{b} \cdot \left(2\nabla U^{k-1} - \nabla U^{k-2}\right) + rU^k = 0$$

Time Deep **Gradient Flow**

$$\frac{U^k - U^{k-1}}{h} - \nabla \cdot \left(A \nabla U^k\right) + \mathbf{b} \cdot \nabla U^{k-1} + rU^k = 0$$

$$\frac{U^k - \frac{4}{3}U^{k-2} + \frac{1}{3}U^{k-1}}{\frac{2}{3}h} - \nabla \cdot \left(A \nabla U^k\right) + \mathbf{b} \cdot \left(2\nabla U^{k-1} - \nabla U^{k-2}\right) + rU^k = 0$$

$$I_1^k(u) = \frac{1}{2}\left\| u - U^{k-1} \right\|^2 + h \left(\int \frac{1}{2}\left((\nabla u)^{\mathsf{T}} A \nabla u + ru^2 \right) + \left(\mathbf{b} \cdot \nabla U^{k-1}\right) u \mathrm{d}\mathbf{x} \right),$$

$$I_2^k(u) = \frac{1}{2}\left\| u - \frac{4}{3}U^{k-1} + \frac{1}{3}U^{k-2} \right\|^2 + \frac{2h}{3} \left(\int \frac{1}{2}\left((\nabla u)^{\mathsf{T}} A \nabla u + ru^2 \right) + \mathbf{b} \cdot \left(2\nabla U^{k-1} - \nabla U^{k-2}\right) u \mathrm{d}\mathbf{x} \right),$$

Time **Deep** Gradient Flow

$$I_n^k(u) = \frac{1}{2} \left\| u + \sum_{j=1}^{n} \alpha_n^j U^{k-j} \right\|^2 + \beta_n h \left(\int \frac{1}{2} \left((\nabla u)^\mathsf{T} A \nabla u + r u^2 \right) + \left(\mathbf{b} \cdot \sum_{j=1}^{n} \gamma_n^j \nabla U^{k-j} \right) u \mathrm{d}\mathbf{x} \right)$$

Time **Deep** Gradient Flow

$$I_n^k(u) = \frac{1}{2}\left\| u + \sum_{j=1}^{n} \alpha_n^j U^{k-j} \right\|^2 + \beta_n h \left(\int \frac{1}{2}\left((\nabla u)^\mathsf{T} A \nabla u + r u^2 \right) + \left(\mathbf{b} \cdot \sum_{j=1}^{n} \gamma_n^j \nabla U^{k-j} \right) u \mathrm{d}\mathbf{x} \right)$$

$$\begin{aligned} L_n^k(\theta; \mathbf{x}) = & \frac{1}{2M} \sum_{m=1}^{M} \left(f^k(\mathbf{x}_m; \theta) + \sum_{j=1}^{n} \alpha_n^j f^{k-j}(\mathbf{x}_m) \right)^2 \\ & + \frac{\beta_n h}{M} \sum_{m=1}^{M} \left[\frac{1}{2}\left(\left(\nabla f^k(\mathbf{x}_m; \theta) \right)^\mathsf{T} A \nabla f^k(\mathbf{x}_m; \theta) + r \left(f^k(\mathbf{x}_m; \theta) \right)^2 \right) \right. \\ & \left. + \left(\mathbf{b} \cdot \sum_{j=1}^{n} \gamma_n^j \nabla f^{k-j}(\mathbf{x}_m) \right) f^k(\mathbf{x}_m; \theta) \right]. \end{aligned}$$

Algorithm: TDGF

Initialize θ_0^0.
Set $f^0(\mathbf{x};\theta) = \Psi(\mathbf{x})$.
for each time step $k = 1, \ldots, K$ **do**
    Initialize $\theta_0^k = \theta^{k-1}$.
    for each sampling stage $n = 1, ..., N$ **do**
        Generate M random points $\mathbf{x}_m$ for training.
        Calculate the cost functional $L^k(\theta_n^k;\mathbf{x})$ for the selected points.
        Take a descent step $\theta_{n+1}^k = \theta_n^k - \alpha \nabla_\theta L^k(\theta_n^k;\mathbf{x})$.
    end for
end for

Architecture

$$X^0 = \sigma_1\left(W^0\mathbf{x} + b^0\right),$$

$$X^l = \sigma_1\left(W^l X^{l-1} + b^1\right), \qquad l = 1, \dots, L,$$

$$f(\mathbf{x};\theta) = \left(S - K\mathrm{e}^{-rt}\right)^+ + \sigma_2\left(WX^L + b\right),$$

$$\left(W^1, b^1\right) \in \left(\mathbb{R}^{D\times d}, \mathbb{R}^D\right)$$

$$\left(W^l, b^l\right) \in \left(\mathbb{R}^{D\times D}, \mathbb{R}^D\right)$$

$$\left(W, b\right) \in \left(\mathbb{R}^{1\times D}, \mathbb{R}\right)$$

Parameters

$$\left(W^1, b^1\right) \in \left(\mathbb{R}^{D\times d}, \mathbb{R}^{D}\right)$$

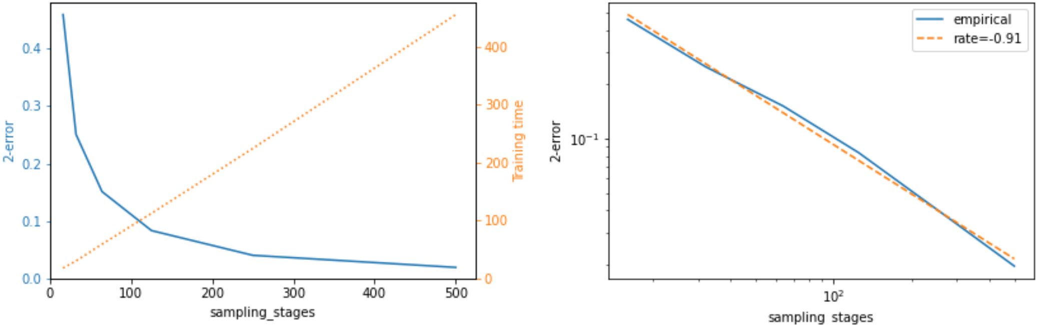
$$\left(W^l, b^l\right) \in \left(\mathbb{R}^{D\times D}, \mathbb{R}^{D}\right)$$

$$\left(W, b\right) \in \left(\mathbb{R}^{1\times D}, \mathbb{R}\right)$$

$$\theta = \left(W^1, b^1, W^l, b^l, W, b\right), \quad l = 1, \ldots, L$$

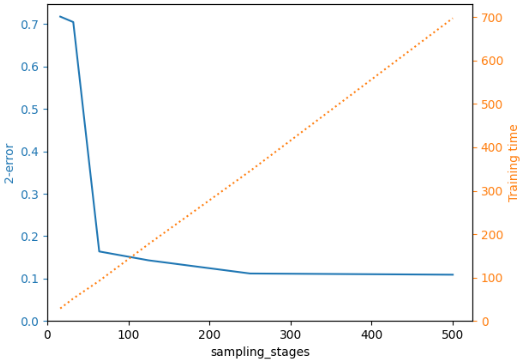
- L: layers
- D: nodes per layer

Sampling stages: TDGF, BS

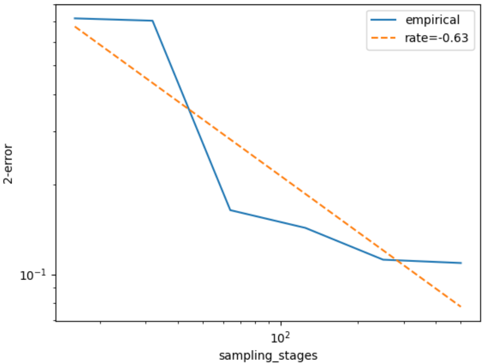


(a) Linear scale (b) Logarithmic scale

Sampling stages: TDGF, Heston

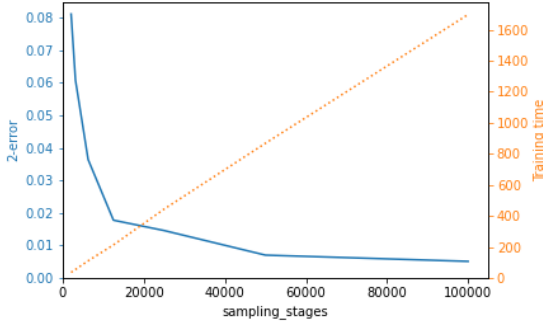


(a) Linear scale

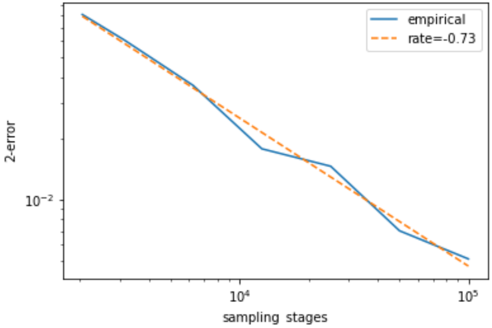


(b) Logarithmic scale

Sampling stages: DGM, BS

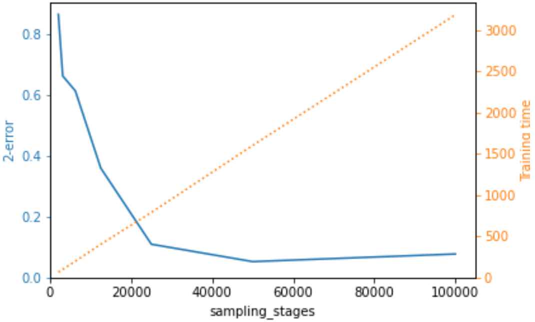


(a) Linear scale

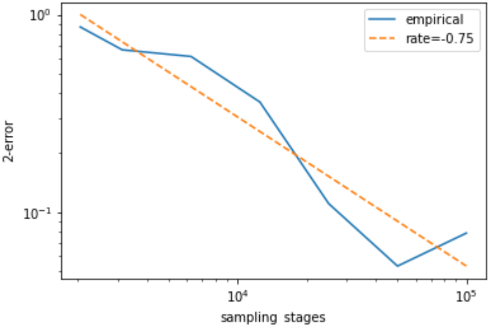


(b) Logarithmic scale

Sampling stages: DGM, Heston

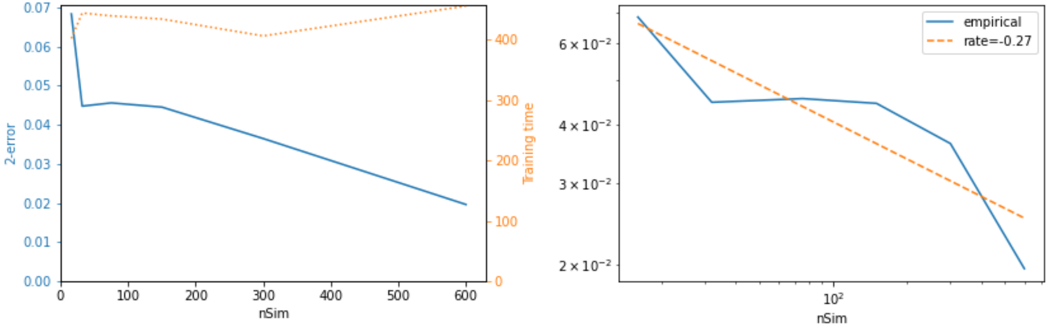


(a) Linear scale



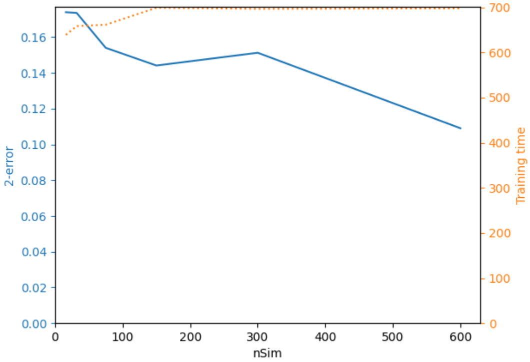
(b) Logarithmic scale

Samples: TDGF, BS

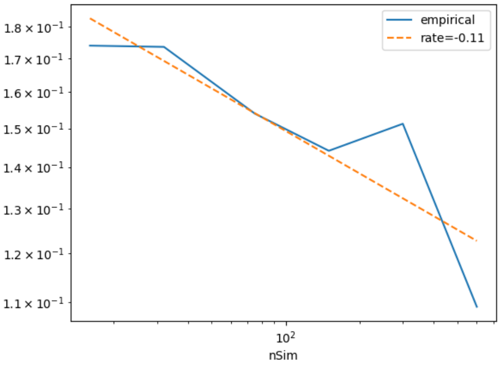


(a) Linear scale (b) Logarithmic scale

Samples: TDGF, Heston

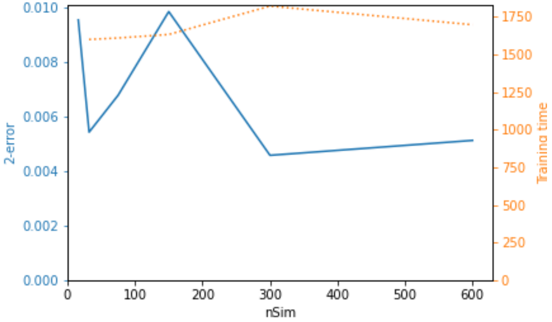


(a) Linear scale

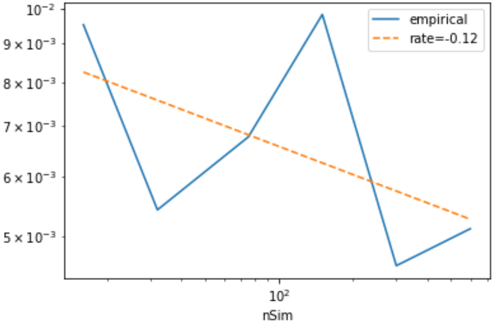


(b) Logarithmic scale

Samples: DGM, BS

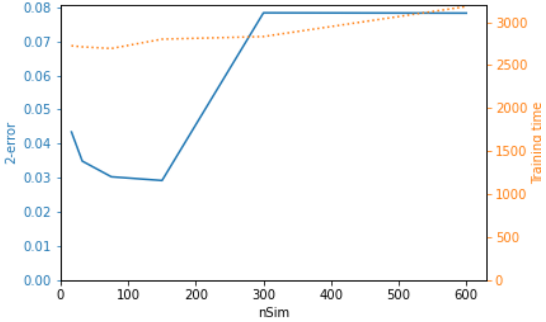


(a) Linear scale

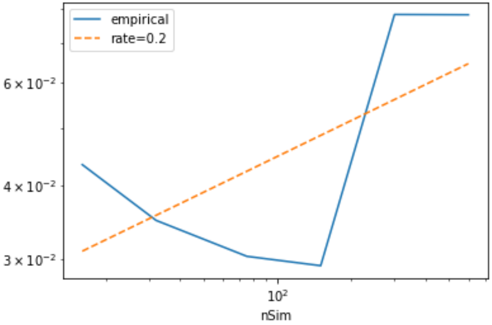


(b) Logarithmic scale

Samples: DGM, Heston

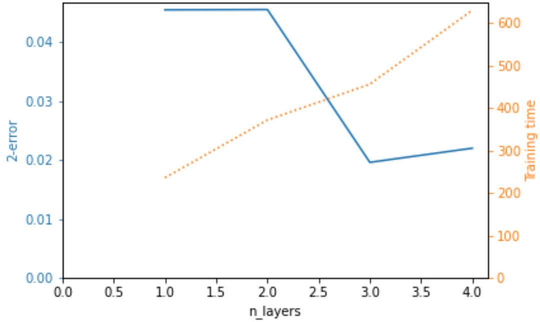


(a) Linear scale

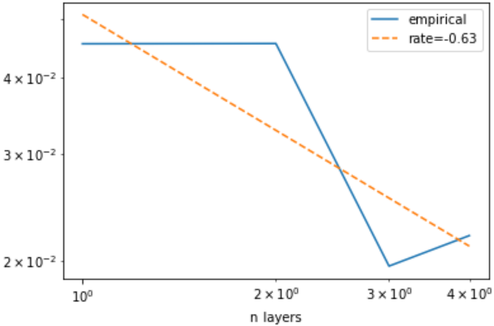


(b) Logarithmic scale

Layers: TDGF, BS

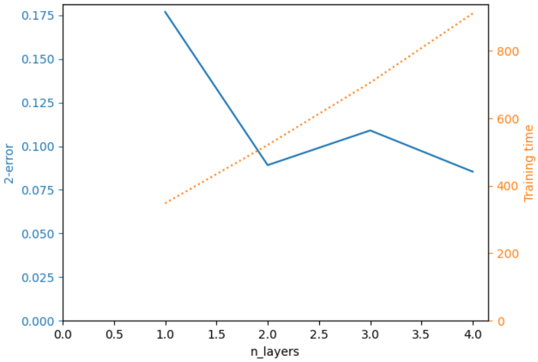


(a) Linear scale

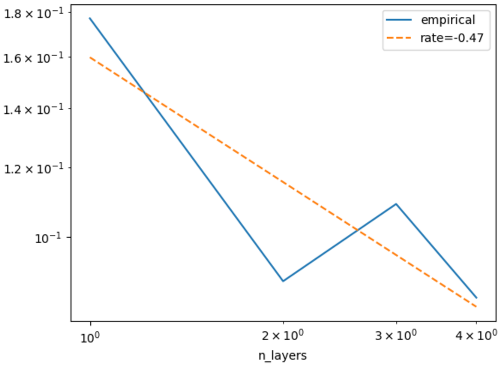


(b) Logarithmic scale

Layers: TDGF, Heston

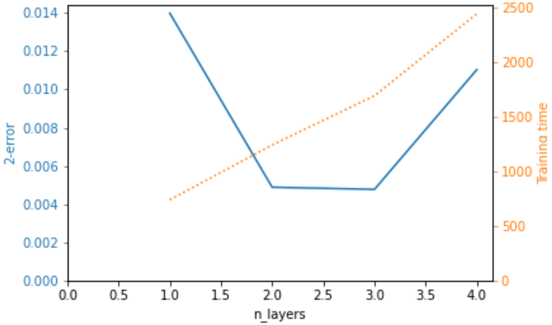


(a) Linear scale

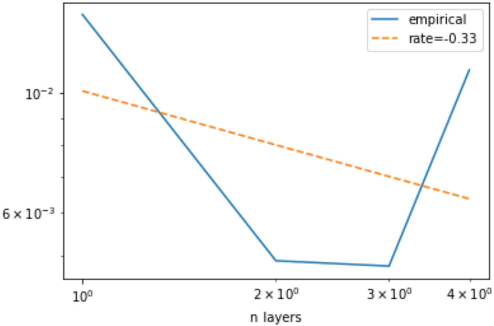


(b) Logarithmic scale

Layers: DGM, BS

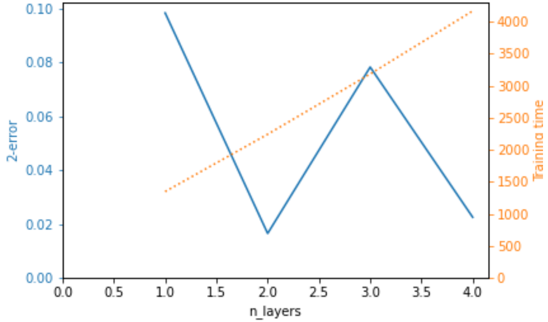


(a) Linear scale

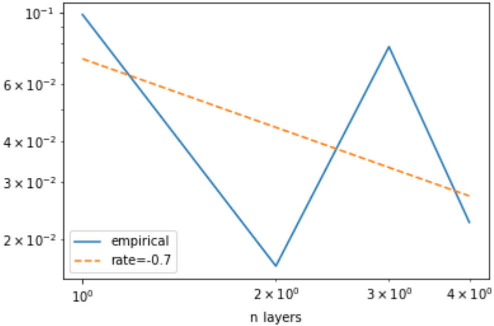


(b) Logarithmic scale

Layers: DGM, Heston

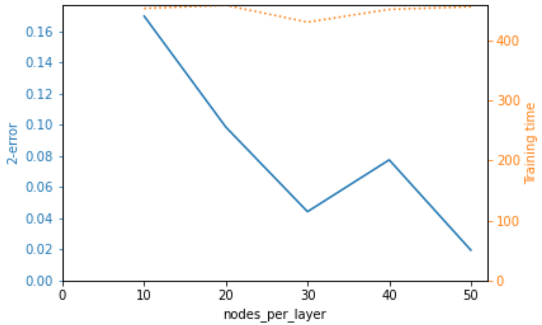


(a) Linear scale

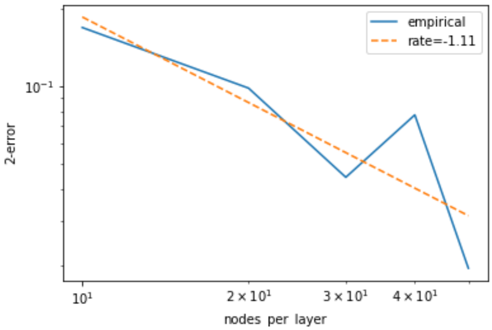


(b) Logarithmic scale

Nodes per layer: TDGF, BS

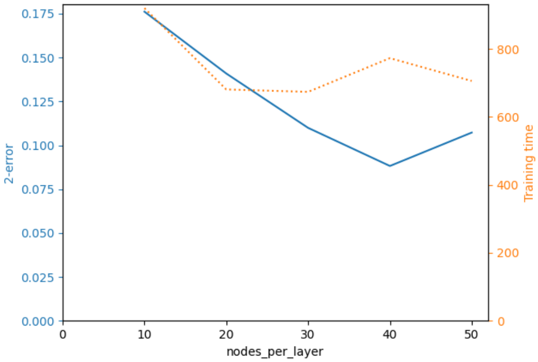


(a) Linear scale

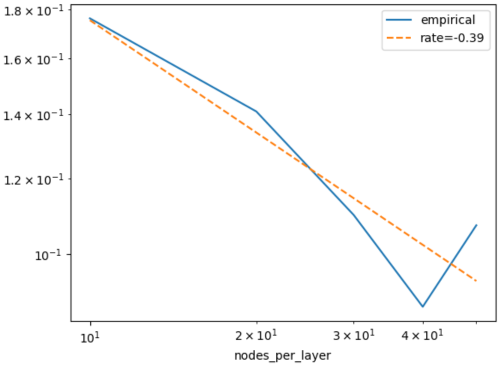


(b) Logarithmic scale

Nodes per layer: TDGF, Heston

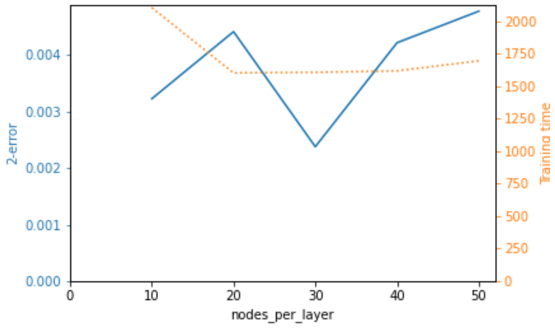


(a) Linear scale

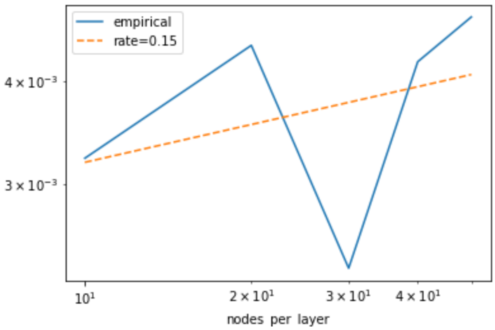


(b) Logarithmic scale

Nodes per layer: DGM, BS

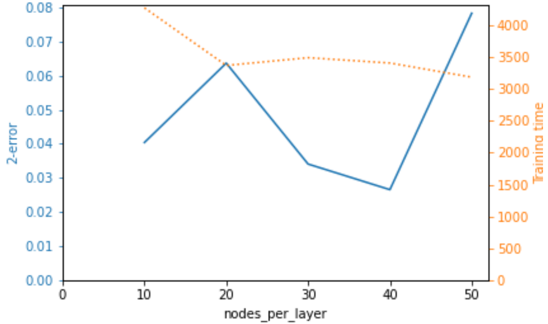


(a) Linear scale

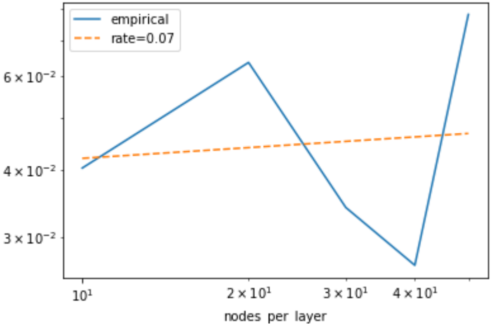


(b) Logarithmic scale

Nodes per layer: DGM, Heston

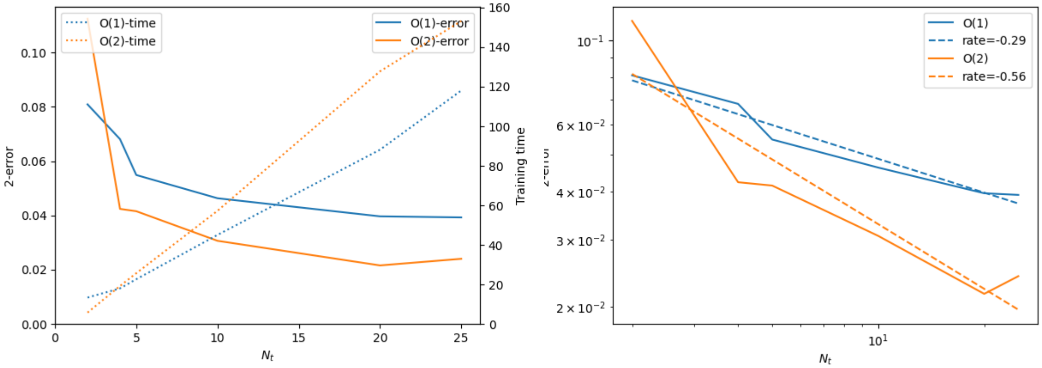


(a) Linear scale



(b) Logarithmic scale

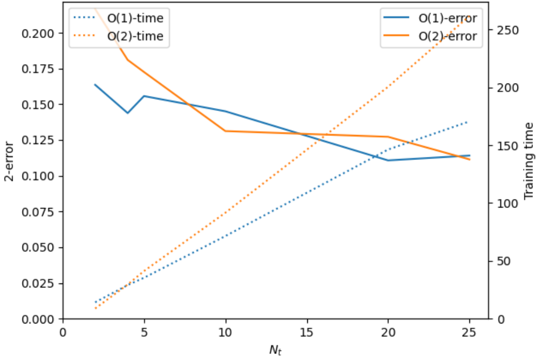
Time steps, BS



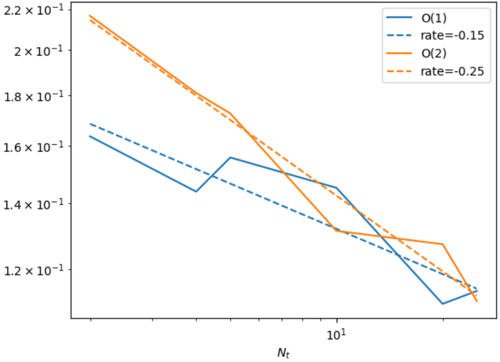
(a) Linear scale

(b) Logarithmic scale

Time steps, BS



(a) Linear scale



(b) Logarithmic scale

Conclusion

Parameter	Accuracy	Training time
Sampling stages	✓	✓
Samples	-	×
Layers	✓	✓
Nodes per layer	-	×
Time steps	✓	✓

Error Analysis of Deep PDE Solvers for Option Pricing

12th General AMaMeF Conference

Jasper Rou

June 23, 2025

j.g.rou@tudelft.nl www.jasperrou.nl