Parareal algorithm for stochastic differential equations

Arthur Arnoult, Frédéric Legoll, Tony Lelièvre, Yvon Maday

CANUM 2026
4 juin 2026 – Saint-Jacut-de-la-mer

SORBONNE UNIVERSITÉ

EMC²
Extreme-scale Mathematically-based Computational Chemistry

Stochastic differential equation

We consider the following overdamped Langevin equation:

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t \qquad \text{for all } t \in [0,T]$$

with a given drift b, diffusion σ and a Brownian motion W_t.
Numerical scheme: Euler Maruyama

$$X_{n+1}^F = X_n^F + \Delta t\, b(X_n^F) + \sqrt{\Delta t}\,\sigma(X_n^F)\,G_{n+1}$$

- $G_n \sim \mathcal{N}(0,1)$
- We assume that b and/or sigma have a high computation cost
- We use coarse drift and diffusion

$$X_{n+1}^C = X_n^C + \Delta t\, b_C(X_n^C) + \sqrt{\Delta t}\,\sigma_C(X_n^C)\,G_{n+1},$$

How can this rough model be used to calculate the fine solution?

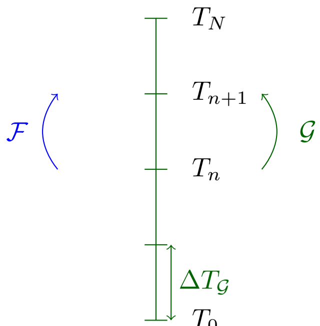
Parareal for ODEs

$\longrightarrow$ J. L. Lions, Y. Maday, G. Turinici,
Résolution d'EDP par un schéma en temps "Pararéel", 2001

We have two propagators:

- Coarse propagator $\mathcal{G}(\mathcal{T}_n, M)$
  Roughly solves the equation on $\mathcal{T}_n$ with initial condition M
  $\longrightarrow$ cheap
  $\longrightarrow$ performed sequentially
- Fine propagator $\mathcal{F}(\mathcal{T}_n, M)$
  Precisely solves the equation on $\mathcal{T}_n$ with initial condition M
  $\longrightarrow$ performed in parallel
  Correction step :

$$X_{n+1}^{k+1} = \underbrace{\mathcal{F}(\mathcal{T}_n, X_n^k)}_{\text{in parallel}} + \mathcal{G}(\mathcal{T}_n, X_n^{k+1}) - \mathcal{G}(\mathcal{T}_n, X_n^k)$$



Studied quantity

- Parareal for trajectories?

$\longrightarrow$ We are interested in the mean value of a given function ψ

- $\psi(X)$: position at final time
- $\psi(X)$: time to cross a potential barrier

$$\mathbb{E}(\Psi(X)) \quad ?$$

Importance sampling

- Monte-Carlo
- Important Sampling : $\mathbb{E}(\psi(X)) = \mathbb{E}(\psi(Y)w(Y))$
- w is a weight function which illustrates how the distribution of Y relates to the distribution of X
- Instead of simulating X, we simulate Y, e.g. if Y is easier to compute

- $X = X^F$, $Y = X^k$

$$\mathbb{E}(\psi(X^F)) = \mathbb{E}(\psi(Y^k)w(Y^k))$$

Importance sampling for Parareal

Theorem

If $k \geq 0$, we have an explicit formula for w (here for $\sigma = \sigma_C$):

$$\mathbb{E}(\psi(X^F)) = \mathbb{E}\left(\psi(X^k)\exp\big(-(\|\mathcal{G}_N^k\|^2 - \|\mathbf{G}_N\|^2)/2\big)\right)$$

with

$$\mathcal{G}_N^k = (\mathcal{S}^F)^{-1}(X^k)$$

- Calculation of $\mathcal{G}_N^k$ imply the inversion $(\mathcal{S}^F)^{-1}$ but can be performed in parallel

$\longrightarrow$ We can correct the error from parareal which is now exact!

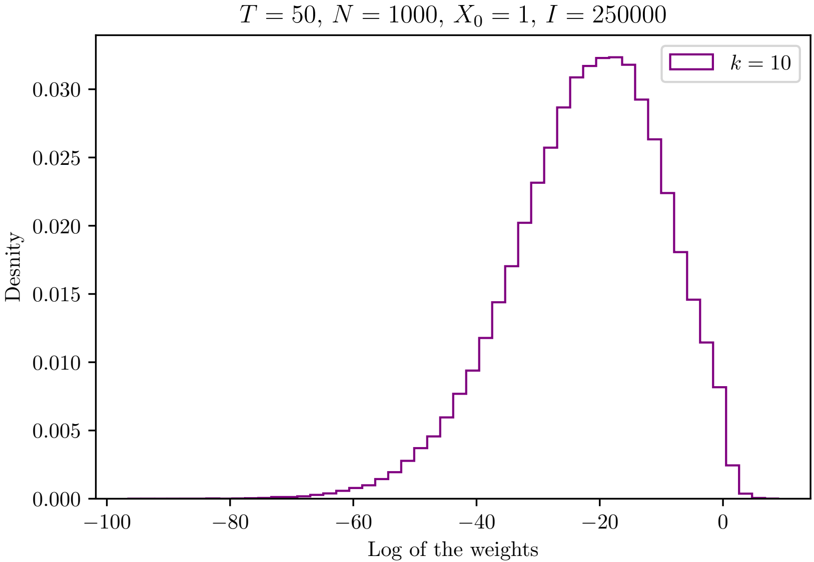
Some remarks. . .

$$\mathbb{E}(\psi(X^F)) = \mathbb{E}(\psi(Y^k)w(Y^k))$$

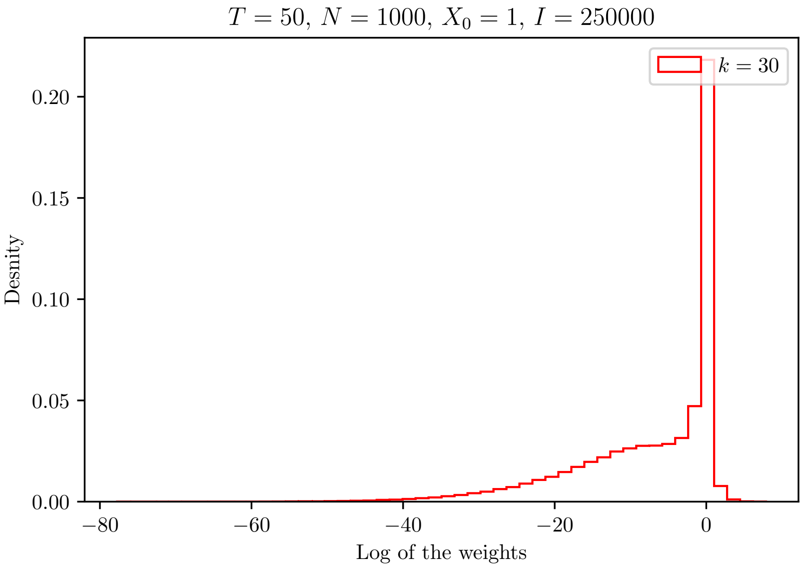
- The correction formula is true for all k
- If $k = 0$, we just replace the fine solver by a coarse solver
- Then what changes if k increases ?

$\longrightarrow$ We must balance between computation time (which increases with k) and the convergence rate of the Monte Carlo method (which decreases with k)

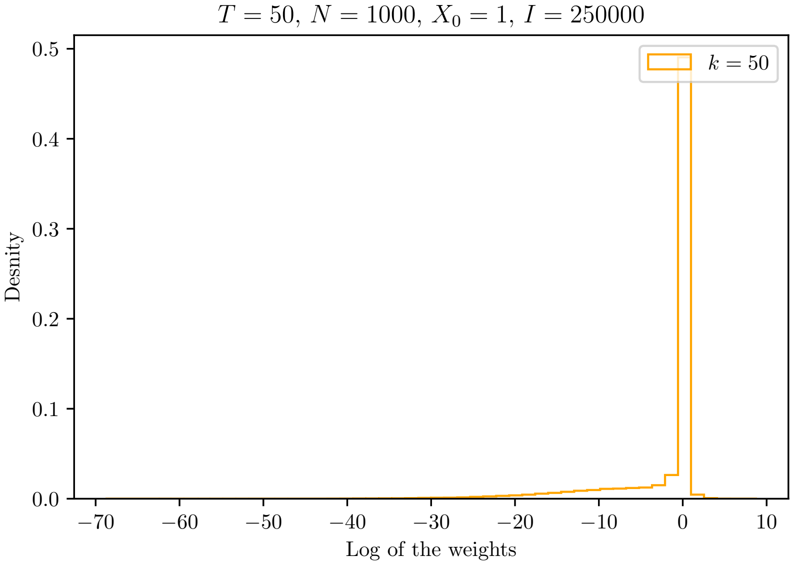
Weight distributions



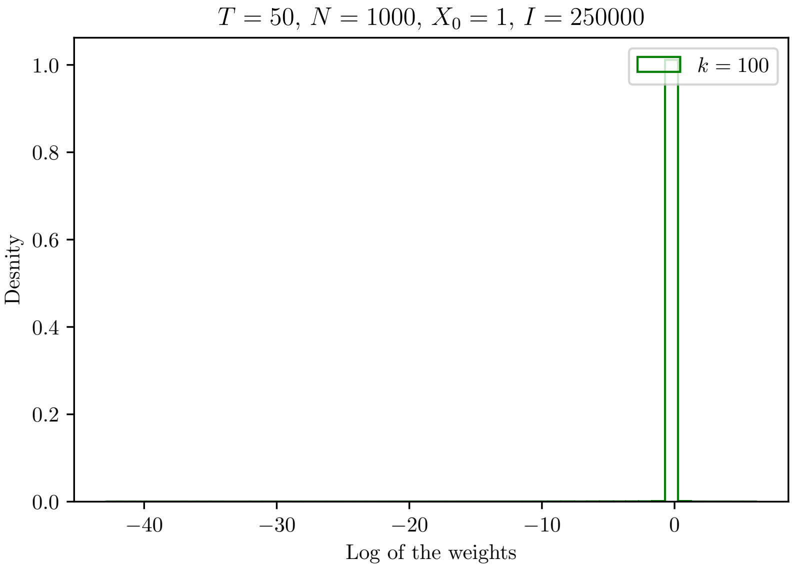
Weight distributions



Weight distributions



Weight distributions



Some remarks...

- Weight captures the difference between the fine solver and the coarse solver, with a mean value of 1
- But high variance (due to the exponential term?) $\longrightarrow$ It must be minimized as much as possible
- If k is large, $w \sim 1$ and correction is useless

Stopping time

$$w = w_1 w_2 \dots w_N$$

with

$$w_i = \exp\left(-\frac{\mathcal{G}_i^2(G_1, \dots, G_i) - G_i^2}{2}\right)$$

- The larger N is, the more random terms there are
- If the studied phenomenon is the transition from one well to another, which occurs after 100 time steps, we would need to consider weights $w_1 w_2 \dots w_{100}$ and not up to N

Stopping time

Theorem

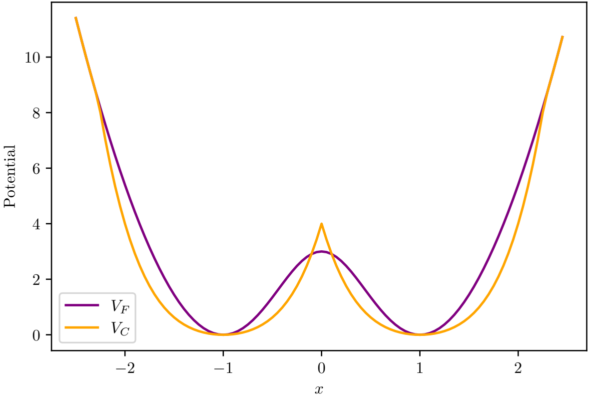
Let $k \in [\![0, K]\!]$. If there exists a stopping time $N_k^\star$ such that $\Psi(x_1, \ldots, x_N) = \Psi(x_1, \ldots, x_{N^\star})$ with $N^\star \leq N$, then, we have

$$\mathbb{E}\left(\Psi(X^F)\right) = \mathbb{E}\left(\Psi(X^k_{N_k^\star})w(X^k_{N_k^\star})\right),$$

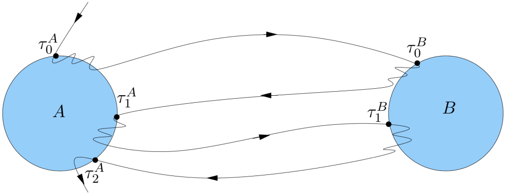
where $X^k_{N_k^\star}$ denotes the vector of the $N_k^\star$ first elements of X^k.

$\longrightarrow$ We reduce the length of the trajectories included in the weights, and we avoid the uncertainty caused by large times

1D test case : Potential curve



Hill formula

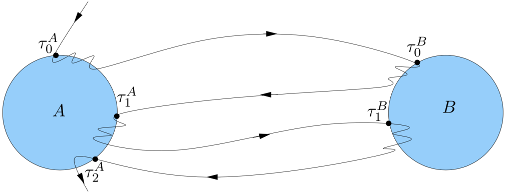


$$T_{AB} = \tau_A \left(\frac{1}{\mathbb{P}_B} - 1 \right) + \tau_B$$

with

- T_{AB} : time from A to B
- τ_A : time from A to A without B
- τ_B : time from A to B without A
- $\mathbb{P}_B$: probability of B from A

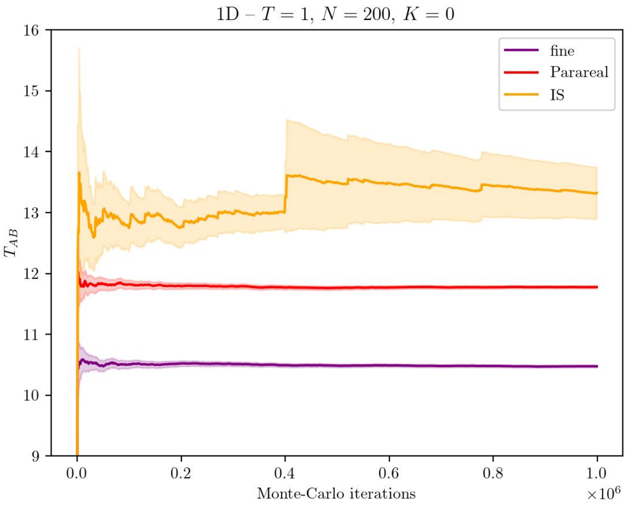
Hill formula



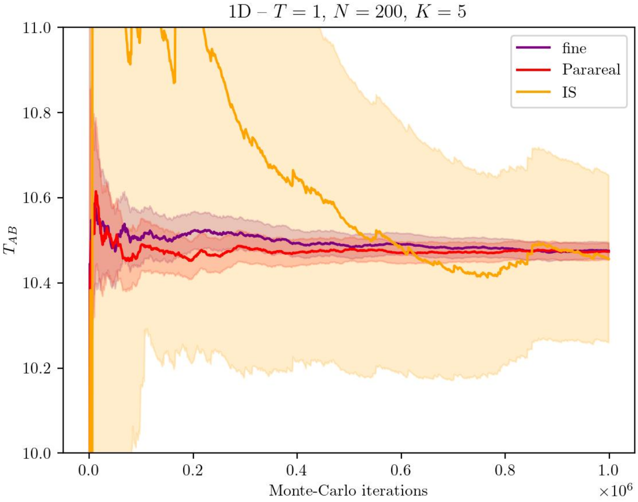
$$T_{AB} = \tau_A \left(\frac{1}{\mathbb{P}_B} - 1 \right) + \tau_B$$

- Only short trajectories : A to B and A to A is fast : N can be small
- It doesn't save time in the simulation, but it greatly reduces the length of the trajectories, so pararel is more efficient
- This greatly reduces the variance in the weights

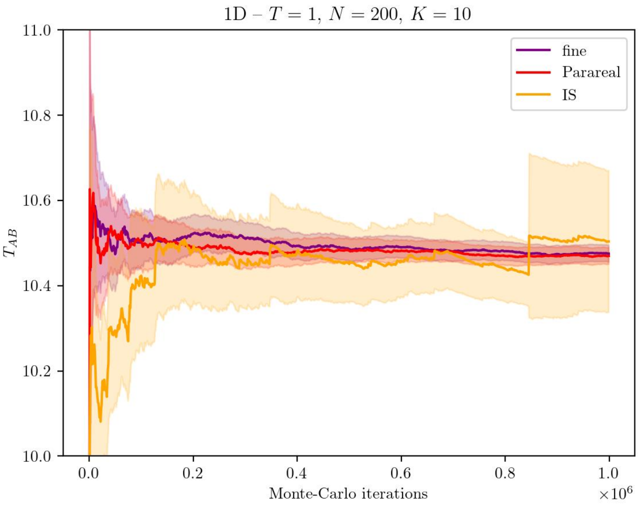
1D test case : Monte-Carlo convergence



1D test case : Monte-Carlo convergence



1D test case : Monte-Carlo convergence



Conclusion & perspectives

- Parareal for stochastic trajectories
- Importance sampling for error correction: calculation of weights
- Random stopping time: weights are considered only at this time step
- Reducing the length of trajectories using Hill formula

- Unsatisfactory numerical results: it is difficult to find cases where parareal performs poorly but IS is more effective in a small number of simulations
- Confidence interval with IS

Conclusion & perspectives

- Parareal for stochastic trajectories
- Importance sampling for error correction: calculation of weights
- Random stopping time: weights are considered only at this time step
- Reducing the length of trajectories using Hill formula

- Unsatisfactory numerical results: it is difficult to find cases where parareal performs poorly but IS is more effective in a small number of simulations
- Confidence interval with IS

- Should we use weights or IS as an indicator of convergence?
- Situations that are less favorable to parareal? For example, by increasing the importance of brownian motion

Thank you for your attention!

SORBONNE UNIVERSITÉ

EMC²

Extreme-scale
Mathematically-based
Computational
Chemistry