KU LEUVEN

Stochastic optimization for tokamak fusion reactor divertor design

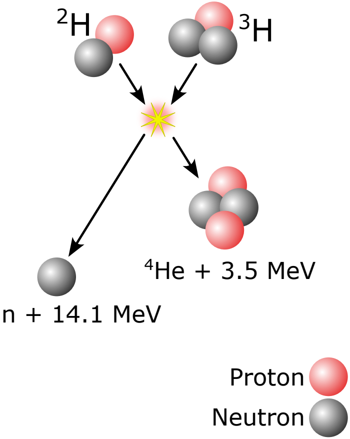
Emil Løvbak, Giovanni Samaey, Stefan Vandewalle
KU Leuven, Department of Computer Science, NUMA Section

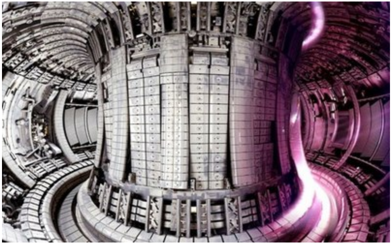
WSC Spring Meeting, 6 May 2022

Research Foundation Flanders
Opening new horizons

NUMA

Application: nuclear fusion in tokamaks



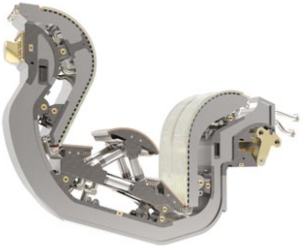


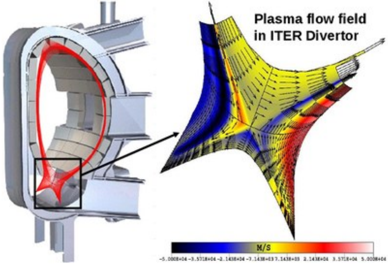
$$Q \geq 10: \quad 50\text{MW} \rightarrow 500\text{MW}$$

iter
china eu india japan korea russia usa

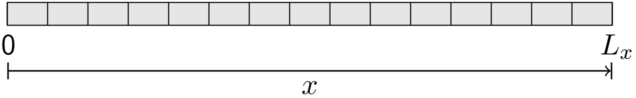
Divertor

- Confine plasma in reactor
- Regulate heat
- Remove waste products: “Ash”, Helium
- Recycle recombined ions as neutrals





A simplified 1D problem[1]



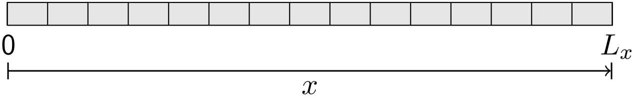
$$J(L_x, \mathbf{q}) = \frac{1}{2} \left(\Gamma - \Gamma_d\right)^2 \Big|_{L_x} + \frac{\lambda}{2} \left(L_x - L_0\right)^2$$

$$\mathbf{q} = \left(\rho, u, f_n\right)^T$$

ρ	plasma density
u	plasma velocity
f_n	neutral position-velocity

1: W. Dekeyser, Optimal Plasma Edge Configurations for Next-Step Fusion Reactors. *PhD thesis*, 2014

A simplified 1D problem[1]



$$J(L_x, \mathbf{q}) = \frac{1}{2}\left(\rho b u - \Gamma_d\right)^2\Big|_{L_x} + \frac{\lambda}{2}\left(L_x - L_0\right)^2$$

$$\mathbf{q} = \left(\rho, u, f_n\right)^T$$

ρ	plasma density
u	plasma velocity
f_n	neutral position-velocity

1: W. Dekeyser, Optimal Plasma Edge Configurations for Next-Step Fusion Reactors. *PhD thesis*, 2014

Forward simulation

- Plasma $\Rightarrow$ Finite volume

$$\frac{\partial}{\partial x}(\rho b u) = S_\rho - C_n \rho$$
$$\frac{\partial}{\partial x}\left(m\rho b u^2 - \eta^i \frac{\partial u}{\partial x}\right) = S_u - b\frac{\partial p}{\partial x}$$

$$S_\psi = \int f_n(x,v)\Psi(x,v)\mathrm{d}v, \quad \psi \in \{\rho/u\} \quad \Rightarrow \quad \text{Low-dimensinal}$$

- Neutrals $\Rightarrow$ Monte Carlo

$$v\frac{\partial}{\partial x} f_n(x,v) + R_i f_n(x,v) = S_{f_n}(\rho, u) + R_{cx}\int f_n(x,v')C(v' \to v)\mathrm{d}v'$$

Computational challenges

High collision rates $R_{cx} \Rightarrow$ Fast timescales

$$\Downarrow$$

Monte Carlo = Time dependent simulation

$$\Downarrow$$

High computational cost?

Computational challenges

High collision rates $R_{cx} \Rightarrow$ Fast timescales

$$\Downarrow$$

Monte Carlo = Time dependent simulation

$$\Downarrow$$

High computational cost?

- Solutions?
  - Monte Carlo = Trivially parallel $\Rightarrow$ More computing power
  - Approximate collision rate with $\infty \Rightarrow$ Diffusion [2]

2: P. A. Markowich, C. A. Ringhofer, and C. Schmeiser, *Semiconductor Equations*. Springer-Verlag, 1990

Outline

1. Asymptotic-preserving multilevel Monte Carlo
2. Gradient computation through discrete adjoint

Research Foundation Flanders
Opening new horizons
NUMA

Outline

1. Asymptotic-preserving multilevel Monte Carlo
2. Gradient computation through discrete adjoint

Rewriting neutral model using AP scheme[3]

$$v\frac{\partial}{\partial x}f_n(x,v) + R_i f_n(x,v) = S_{f_n}(\rho,u) + R_{cx}\int f_n\left(x,v'\right)C\left(v'\to v\right)\mathrm{d}v'$$

- High collision rate R_{cx} & characteristic velocity $\tilde{v}$
- Monte Carlo particle scheme $\Delta t \ll R_{cx}^{-1}$
- $R_{cx}, \tilde{v} \to \infty \Rightarrow$ diffusion

3: G. Dimarco, L. Pareschi, G. Samaey, Asymptotic-Preserving Monte Carlo methods for transport equations in the diffusive limit. *SIAM Journal on Scientific Computing* 40, 2018

Rewriting neutral model using AP scheme[3]

$$v\frac{\partial}{\partial x}f_n(x,v) + R_i f_n(x,v) = S_{f_n}(\rho,u) + R_{cx}\int f_n\left(x,v'\right)C\left(v'\to v\right)\mathrm{d}v'$$

$$\Downarrow$$

$$\frac{v}{1+\Delta t R_{cx}}\frac{\partial}{\partial x}f_n(x,v) + R_i f_n(x,v) = \frac{\Delta t\tilde{v}}{1+\Delta t R_{cx}}\frac{\partial}{\partial x}\left(\tilde{v}\frac{\partial}{\partial x}f_n(x,v)\right) + S_{f_n}(\rho,u) + \frac{R_{cx}}{1+\Delta t R_{cx}}\int f_n\left(x,v'\right)C\left(v'\to v\right)\mathrm{d}v'$$

3: G. Dimarco, L. Pareschi, G. Samaey, Asymptotic-Preserving Monte Carlo methods for transport equations in the diffusive limit. *SIAM Journal on Scientific Computing* 40, 2018

Monte Carlo particle scheme (operator splitting)

- Transport-diffusion:

$$\frac{v}{1+\Delta t R_{cx}}\frac{\partial}{\partial x}f_n(x,v) = \frac{\Delta t \tilde{v}}{1+\Delta t R_{cx}}\frac{\partial}{\partial x}\left(\tilde{v}\frac{\partial}{\partial x}f_n(x,v)\right)$$

$$\Downarrow$$

$$X^{n+1} = X^n + \frac{V^n \Delta t}{1+\Delta t R_{cx}} + \sqrt{2\Delta t}\sqrt{\frac{\Delta t \tilde{v}^2}{1+\Delta t R_{cx}}}\xi^n, \quad \xi^n \sim \mathcal{N}(0,1)$$

- Collision/ionization step

$$R_i f_n(x,v) = \frac{R_{cx}}{1+\Delta t R_{cx}}\int f_n\left(x,v'\right)C\left(v' \to v\right)\mathrm{d}v'$$

Respective event probabilities:

$$1-\exp\left(-\frac{R_{cx}\Delta t}{1+R_{cx}\Delta t}\right), \quad 1-\exp\left(-R_i \Delta t\right)$$

Multilevel Monte Carlo estimator[4]

- Estimator with independent particles:

$$S_\psi = \frac{1}{P} \sum_{p=1}^{P} \sum_{n=0}^{N_p} \int_{X_p^n}^{X_p^{n+1}} g\left(x, V_p^n\right) \mathsf{d}x \prod_{k=1}^{n+1} W_\psi\left(X_{p,\Delta t}^k\right), \ \psi \in \{\rho/u\}$$

- Cost proportional to $P \times N$

4: M.B. Giles, Multilevel Monte Carlo Path Simulation. *Operations Research* 56(3), 2008

Multilevel Monte Carlo estimator[4]

- Estimator with independent particles:

$$S_{\psi,0} = \frac{1}{P_0} \sum_{p=1}^{P_0} \sum_{n=0}^{N_{p,0}} \int_{X_{p,0}^{n}}^{X_{p,0}^{n+1}} g\left(x, V_{p,0}^{n}\right) \mathsf{d}x \prod_{k=1}^{n+1} W_\psi\left(X_{p,0}^{k}\right)$$

- Cost proportional to $P \times N$
- Difference estimator, $\ell > 0$:

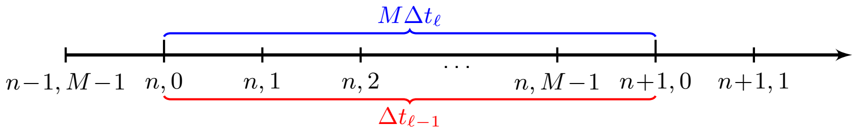
$$S_{\psi,\ell} = \frac{1}{P_\ell} \sum_{p=1}^{P_\ell} \left(\sum_{n=0}^{N_{p,\ell}} \int_{X_{p,\ell}^{n}}^{X_{p,\ell}^{n+1}} g\left(x, V_{p,\ell}^{n}\right) \mathsf{d}x \prod_{k=1}^{n+1} W_\psi\left(X_{p,\ell}^{k}\right) \right.$$
$$\left. - \sum_{n=0}^{N_{p,\ell-1}} \int_{X_{p,\ell-1}^{n}}^{X_{p,\ell-1}^{n+1}} g\left(x, V_{p,\ell-1}^{n}\right) \mathsf{d}x \prod_{k=1}^{n+1} W_\psi\left(X_{p,\ell-1}^{k}\right) \right)$$

4: M.B. Giles, Multilevel Monte Carlo Path Simulation. *Operations Research* 56(3), 2008

Correlating Particle Pairs[5]

$$X_{p,\ell-1}^{n+1} = X_{p,\ell-1}^{n} + \frac{V_{p,\ell-1}^{n}\Delta t_{\ell-1}}{1+\Delta t_{\ell-1}R_{cx}} + \sqrt{2\Delta t_{\ell-1}}\sqrt{\frac{\Delta t_{\ell-1}\left(V_{p,\ell-1}^{n}\right)^2}{1+\Delta t_{\ell-1}R_{cx}}}\xi_{p,\ell-1}^{n}$$

$$X_{p,\ell}^{n+1,0} = X_{p,\ell}^{n,0} + \sum_{m=0}^{M-1}\left(\frac{V_{p,\ell}^{n,m}\Delta t_{\ell}}{1+\Delta t_{\ell}R_{cx}} + \sqrt{2\Delta t_{\ell}}\sqrt{\frac{\Delta t_{\ell}\left(V_{p,\ell}^{n,m}\right)^2}{1+\Delta t_{\ell}R_{cx}}}\xi_{p,\ell}^{n,m}\right)$$



$$\xi_{\ell-1}^{n}, \xi_{\ell}^{n,m} \sim \mathcal{N}(0,1) \quad V_{\ell-1}^{n}, V_{\ell}^{n,m} \sim C(v' \to v)$$

5: E. Løvbak, G. Samaey and S. Vandewalle, A multilevel Monte Carlo method for asymptotic-preserving particle schemes in the diffusive limit. *Numerische Mathematik* 148, 2021

Estimating squared particle displacement[6]

- $R_{cx} = 100$
- $\Delta t_0 = t^* = 0.5$
- $C(v' \to v) = \mathcal{N}(0,1)$

ℓ	Δt_ℓ	Result	Variance	Samples	Cost
0	0.5×10^{0}	9.90×10^{-1}	1.96×10^{0}	3.8×10^{8}	7.7×10^{6}
1	7.8×10^{-5}	-1.39×10^{-2}	2.70×10^{-1}	1.8×10^{6}	2.3×10^{8}
2	3.9×10^{-5}	2.23×10^{-3}	1.69×10^{-2}	2.6×10^{5}	9.8×10^{7}
3	2.0×10^{-5}	1.14×10^{-3}	9.17×10^{-3}	1.3×10^{5}	9.9×10^{7}
4	9.8×10^{-6}	3.79×10^{-4}	4.80×10^{-3}	1.0×10^{5}	1.6×10^{8}
5	4.9×10^{-6}	2.07×10^{-4}	2.03×10^{-3}	6.9×10^{4}	2.1×10^{8}
6	2.4×10^{-6}	-4.74×10^{-4}	2.18×10^{-4}	1.0×10^{3}	6.1×10^{6}
$\sum$		9.80×10^{-1}			8.1×10^{8}

6: E. Løvbak, B. Mortier, G. Samaey and S. Vandewalle, Multilevel Monte Carlo with improved correlation for kinetic equations in the diffusive scaling. V. Krzhizhanovskaya et al. (eds.) *LNCS – ICCS 2020*, 2020

Estimating squared particle displacement[6]

- $R_{cx} = 100$
- $\Delta t_0 = t^* = 0.5$
- $C(v' \to v) = \mathcal{N}(0,1)$

ℓ	Δt_ℓ	Result	Variance	Samples	Cost
0	0.5×10^{0}	9.90×10^{-1}	1.96×10^{0}	3.8×10^{8}	7.7×10^{6}
1	7.8×10^{-5}	-1.39×10^{-2}	2.70×10^{-1}	1.8×10^{6}	2.3×10^{8}
2	3.9×10^{-5}	2.23×10^{-3}	1.69×10^{-2}	2.6×10^{5}	9.8×10^{7}
3	2.0×10^{-5}	1.14×10^{-3}	9.17×10^{-3}	1.3×10^{5}	9.9×10^{7}
4	9.8×10^{-6}	3.79×10^{-4}	4.80×10^{-3}	1.0×10^{5}	1.6×10^{8}
5	4.9×10^{-6}	2.07×10^{-4}	2.03×10^{-3}	6.9×10^{4}	2.1×10^{8}
6	2.4×10^{-6}	-4.74×10^{-4}	2.18×10^{-4}	1.0×10^{3}	6.1×10^{6}
$\sum$		9.80×10^{-1}			8.1×10^{8}

- Single level: 3.8×10^{8} samples with $\Delta t_6 \Rightarrow$ Cost: 1.6×10^{12}

6: E. Løvbak, B. Mortier, G. Samaey and S. Vandewalle, Multilevel Monte Carlo with improved correlation for kinetic equations in the diffusive scaling. V. Krzhizhanovskaya et al. (eds.) *LNCS – ICCS 2020*, 2020

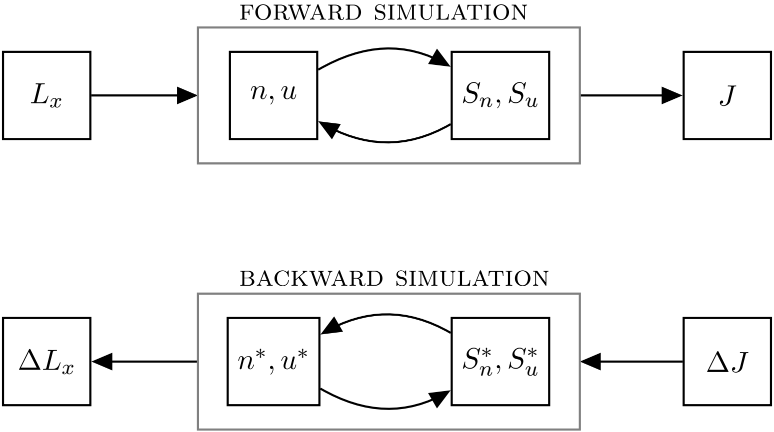
Outline

1. Asymptotic-preserving multilevel Monte Carlo

2. Gradient computation through discrete adjoint

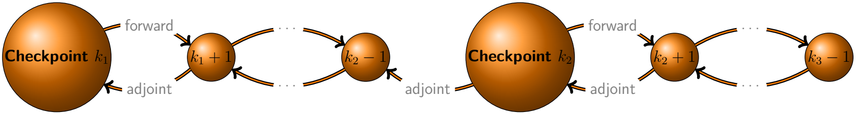
Research Foundation Flanders
Opening new horizons
NUMA

Forward simulation

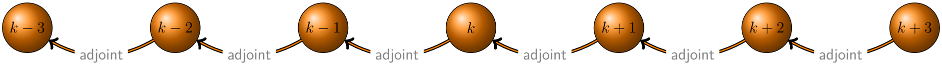


Matching forward and adjoint simulations

- Same paths in forward/backward simulation
- Challenge: $P \times N$ large
- Solutions:
  - Checkpointing: 2 forward simulations + backward simulation



  - Generating the paths in reverse (this work)



Reversing a random number generator

- PCG: permuted congruential generator[7]
  - d-dimensional internal state η_k and constant vectors a, c, m

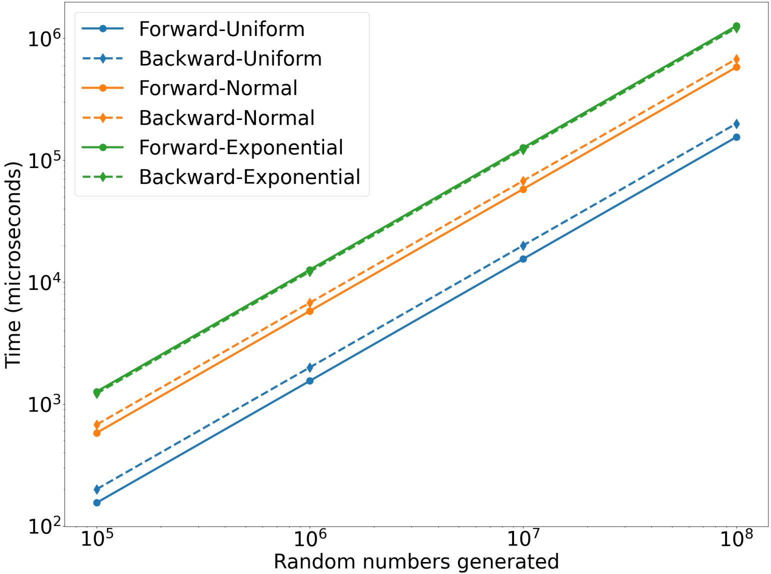
$$\eta_{k+1} = (a \odot \eta_k + c) \text{ mod } m$$

  - 1-way (permutation) function generates output from η_k
  - Passes `TestU01` with flying colors
- Reversing modular operations $\rightarrow$ reversed uniform sequence
- Normal distribution through reversed Ziggurat algorithm
- Exponential distribution through inverse transform:

$$u \sim \mathcal{U}([0,1]) \Rightarrow -\lambda \ln(1-u) \sim \mathcal{E}(\lambda)$$

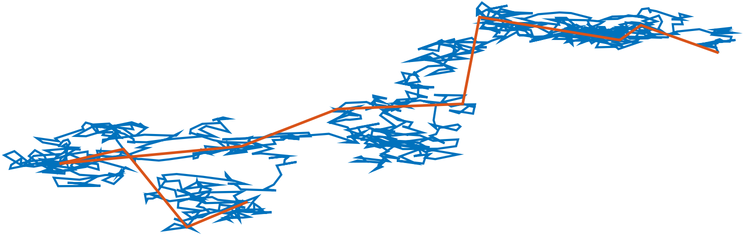
7: M.E. O'Neill, PCG: A Family of Simple Fast Space-Efficient Statistically Good Algorithms for Random Number Generation, technical report HMC-CS-2014-0905, Harvey Mudd College, 2014

Reversable PCG



Outlook and challenges

- Alternative APMLMC-scheme[8]
- Boundaries for combined transport/diffusion
- Scoring diffusive particles in 2D/3D[9]



Thank you for your attention!

8: B. Mortier, P. Robbe, M. Baelmans, G. Samaey, Multilevel asymptotic-preserving Monte Carlo for kinetic-diffusive particle simulations of the Boltzmann-BGK equation. *Journal of Computational Physics* 450, pp. 110736, 2022

9: B. Mortier, M. Baelmans, G. Samaey, A comparison of source term estimators in coupled finite-volume/Monte-Carlo methods with applications to plasma edge simulations in nuclear fusion, arXiv:2012.08981. 2020

Research Foundation Flanders
Opening new horizons
NUMA

References

1 W. Dekeyser, Optimal Plasma Edge Configurations for Next-Step Fusion Reactors. *PhD thesis*, 2014

2 P. A. Markowich, C. A. Ringhofer, and C. Schmeiser, *Semiconductor Equations*. Springer-Verlag, 1990

3 G. Dimarco, L. Pareschi, G. Samaey, Asymptotic-Preserving Monte Carlo methods for transport equations in the diffusive limit. *SIAM Journal on Scientific Computing* 40, pp. A504–A528, 2018

4 M.B. Giles, Multilevel Monte Carlo Path Simulation. *Operations Research* 56(3), 2008

5 **E. Løvbak**, G. Samaey and S. Vandewalle, A multilevel Monte Carlo method for asymptotic-preserving particle schemes in the diffusive limit. *Numerische Mathematik* 148, pp. 141–186, 2021

6 **E. Løvbak**, B. Mortier, G. Samaey and S. Vandewalle, Multilevel Monte Carlo with improved correlation for kinetic equations in the diffusive scaling. V. Krzhizhanovskaya et al. (eds.) *LNCS – ICCS 2020*, pp. 374–388, 2020

7 M.E. O'Neill, PCG: A Family of Simple Fast Space-Efficient Statistically Good Algorithms for Random Number Generation, technical report HMC-CS-2014-0905, Harvey Mudd College, 2014

8 B. Mortier, P. Robbe, M. Baelmans, G. Samaey, Multilevel asymptotic-preserving Monte Carlo for kinetic-diffusive particle simulations of the Boltzmann-BGK equation. *Journal of Computational Physics* 450, pp. 110736, 2022

9 B. Mortier, M. Baelmans, G. Samaey, A comparison of source term estimators in coupled finite-volume/Monte-Carlo methods with applications to plasma edge simulations in nuclear fusion. arXiv:2012.08981, 2020

Correlating Particle Pairs: Diffusion[10]

- Coarse diffusion = weighted sum of fine contributions:

$$\xi^n_{\ell-1} = \sqrt{\theta_\ell}\,\xi^n_{\ell-1,\text{diffusion}} + \sqrt{1-\theta_\ell}\,\xi^n_{\ell-1,\text{velocities}}$$

$$\Downarrow$$

Weak Brownian motion

- Fine diffusion: $\xi^n_{\ell-1,\text{diffusion}} = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} \xi^{n,m}_\ell$
- Fine velocities: $\xi^n_{\ell-1,\text{velocities}} = \left(\mathbb{V}\left[\sum_{m=0}^{M-1} V^{n,m}_\ell\right]\right)^{-\frac{1}{2}} \sum_{m=0}^{M-1} V^{n,m}_\ell$

5: E. Løvbak, B. Mortier, G. Samaey and S. Vandewalle, Multilevel Monte Carlo with improved correlation for kinetic equations in the diffusive scaling. V. Krzhizhanovskaya et al. (eds.) *LNCS – ICCS 2020*, 2020

Correlating Particle Pairs: Collision/absorption[11]

- Event probability p_e is different for different Δt
- Implementation:

$$u_\ell^n > 1 - p_e,\ u_\ell^n \sim \mathcal{U}[0,1] \quad \Rightarrow \quad \text{simulate event}$$

- Generate uniformly distributed $u_{\ell-1}^n$ from $u_{\text{max}} = \max_m u_\ell^{n,m}$:

$$u_{\ell-1}^n = u_{\text{max}}^M \sim \mathcal{U}([0,1])$$

- Collision: new velocity: $V_{\ell-1}^{n+1} = V_\ell^{n+1,0}$
- Absorption: Re-weight/kill particle

6: E. Løvbak, G. Samaey and S. Vandewalle, A multilevel Monte Carlo method for asymptotic-preserving particle schemes in the diffusive limit. *Numerische Mathematik* 148, 2021

Continuous vs. discrete adjoints

$$
\begin{tikzcd}[column sep=huge, row sep=huge]
\mathcal{L}[q] \arrow[r, leftrightarrow, "\text{Adjoint}"] \arrow[d, "\text{Discretize}"] & \mathcal{L}^*[q^*] \arrow[d, "\text{Discretize}"] \\
\hat{\mathcal{L}}[\hat{q}] & \hat{\mathcal{L}}^*[\hat{q}^*]
\end{tikzcd}
$$

$$S_\psi^* = \sum_k \int f_n^*(x,v)\Psi_k^*(x,v) \int f_n(x,v')\Psi_k(x,v',v)\mathrm{d}\mathbf{v}'\,\mathrm{d}v,\;\; \psi \in \{\rho/u\}$$

Continuous vs. discrete adjoints

$$\begin{tikzcd} \mathcal{L}[q] \arrow[d, "\text{Discretize}"] & \mathcal{L}^*[q^*] \\ \hat{\mathcal{L}}[\hat{q}] \arrow[r, leftrightarrow, "\text{Adjoint}"'] & \hat{\mathcal{L}}^*[\hat{q}^*] \end{tikzcd}$$

$$S_\psi = \frac{1}{P} \sum_{p=1}^{P} \sum_{n=0}^{N_p} \int_{X_p^n}^{X_p^{n+1}} g\left(x, V_p^n\right) \mathsf{d}x \prod_{k=1}^{n+1} W_\psi\left(X_p^k\right), \ \psi \in \{\rho/mu\}$$

Discrete adjoint formulation

- Forward scoring

$$S_\psi = \frac{1}{P}\sum_{p=1}^{P}\sum_{n=0}^{N_p}\int_{X_p^n}^{X_p^{n+1}} g\left(x, V_p^n\right) \mathsf{d}x \prod_{k=1}^{n+1} W_\psi\left(X_p^k\right), \ \psi \in \{\rho/mu\}$$

- Adjoint scoring

$$S_\psi^* = \frac{1}{P}\sum_{p=1}^{P}\sum_{n=0}^{N_p}\int_{X_p^n}^{X_p^{n+1}} \frac{\partial}{\partial\psi} g\left(x, V_p^n\right) \mathsf{d}x \prod_{k=1}^{n+1} W_\psi\left(X_p^k\right)$$
$$+ \frac{1}{P}\sum_{p=1}^{P}\sum_{n=0}^{N_p}\int_{X_p^n}^{X_p^{n+1}} g\left(x, V_p^n\right) \mathsf{d}x \prod_{k=1}^{n+1} W_\psi\left(X_p^k\right) \sum_{l=1}^{n+1} \frac{\frac{\partial W_\psi(X_p^l)}{\partial\psi}}{W_\psi\left(X_p^k\right)}$$

- Adjoint simulation runs backward in time

Convergence of stochastic gradient descent[12]

- Given
  - Objective function $J(\Omega, \mathbf{q})$ and Lipschitz gradient $\nabla J(\Omega, \mathbf{q})$
  - Sequence of estimates $\Omega_k, \mathbf{q}_k$ for which $J(\Omega_k, \mathbf{q}_k) \geq J^*$
  - Stochastic gradient estimate $\nabla \hat{J}_k(\Omega_k, \mathbf{q}_k, \eta_k)$ with noise η_k
  - Constants $\mu_G > \mu > 0$ and $M, M_V \geq 0$ so

$$\nabla J(\Omega_k, \mathbf{q}_k)^T \mathbb{E}_{\eta_k}\left[\nabla \hat{J}_k(\Omega_k, \mathbf{q}_k, \eta_k)\right] \geq \mu \left\|\nabla J(\Omega_k, \mathbf{q}_k)\right\|_2^2 \quad \text{(descending)}$$

$$\left\|\mathbb{E}_{\eta_k}\left[\nabla \hat{J}_k(\Omega_k, \mathbf{q}_k, \eta_k)\right]\right\|_2 \leq \mu_G \left\|\nabla J(\Omega_k, \mathbf{q}_k)\right\|_2 \quad \text{(bounded)}$$

$$\mathbb{V}_{\eta_k}\left[\nabla \hat{J}_k(\Omega_k, \mathbf{q}_k, \eta_k)\right] < M + M_V \left\|\nabla J(\Omega_k, \mathbf{q}_k)\right\|_2^2 \quad \text{(variance)}$$

- Stochastic gradient descent with step size α:

$$\mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K} ||\nabla J(\Omega_k, \mathbf{q}_k)||_2^2\right] \leq \frac{\alpha L M}{\mu} + \frac{2(J(\Omega_1, \mathbf{q}_1) - J^*)}{K\mu\alpha} \xrightarrow{K \to \infty} \frac{\alpha L M}{\mu}$$

10: L. Bottou, F.E. Curtis and J. Nocedal, Optimization Methods for Large-Scale Machine Learning. *SIAM Review* 60(2), 2018

Convergence of stochastic gradient descent[12]

- Given
  - Objective function $J(\Omega, \mathbf{q})$ and Lipschitz gradient $\nabla J(\Omega, \mathbf{q})$
  - Sequence of estimates $\Omega_k, \mathbf{q}_k$ for which $J(\Omega_k, \mathbf{q}_k) \geq J^*$
  - Stochastic gradient estimate $\nabla \hat{J}_k(\Omega_k, \mathbf{q}_k, \eta_k)$ with noise η_k
  - Constants $\mu_G > \mu > 0$ and $M, M_V \geq 0$ so

$$\nabla J(\Omega_k, \mathbf{q}_k)^T \mathbb{E}_{\eta_k}\left[\nabla \hat{J}_k(\Omega_k, \mathbf{q}_k, \eta_k)\right] \geq \mu \left\|\nabla J(\Omega_k, \mathbf{q}_k)\right\|_2^2 \quad \text{(descending)}$$

$$\left\|\mathbb{E}_{\eta_k}\left[\nabla \hat{J}_k(\Omega_k, \mathbf{q}_k, \eta_k)\right]\right\|_2 \leq \mu_G \left\|\nabla J(\Omega_k, \mathbf{q}_k)\right\|_2 \quad \text{(bounded)}$$

$$\mathbb{V}_{\eta_k}\left[\nabla \hat{J}_k(\Omega_k, \mathbf{q}_k, \eta_k)\right] < M + M_V \left\|\nabla J(\Omega_k, \mathbf{q}_k)\right\|_2^2 \quad \text{(variance)}$$

- Stochastic gradient descent with step size α:

$$\mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K} ||\nabla J(\Omega_k, \mathbf{q}_k)||_2^2\right] \leq \frac{\alpha L M}{\mu} + \frac{2(J(\Omega_1, \mathbf{q}_1) - J^*)}{K\mu\alpha} \xrightarrow{K\to\infty} \frac{\alpha L M}{\mu}$$

10: L. Bottou, F.E. Curtis and J. Nocedal, Optimization Methods for Large-Scale Machine Learning. *SIAM Review* 60(2), 2018