Forward Event-Chain Monte Carlo

Fast sampling by randomness control in irreversible Markov chains

Manon Michel and Stéphane Sénécal – Orange Labs, France.

{manon.michel,stephane.senecal}@orange.com

Objective: balancing determinism and randomness to sample efficiently from a target distribution

Efficient sampling by irreversible Markov chains, a recent research field in Bayesian statistics.

- First popular in Physics with **event-chain** (EC) methods [2, 3]: **exploration of space by a fixed direction e** over long displacement. **No rejection** (changed into direction updates) and **no critical parameter fine-tuning**.
- EC and similar methods (e.g. BPS [6]) shown to **outperform state-of-the-art Hamiltonian/Hybrid MC methods** and to provide **subsampling schemes thanks to the factorized filter**.
- However, as in HMC, energy/likelihood gradient information only used to **follow iso-energetic lines**. Resampling of direction e needed to recover ergodicity, introducing extra randomization and a resampling rate to tune.

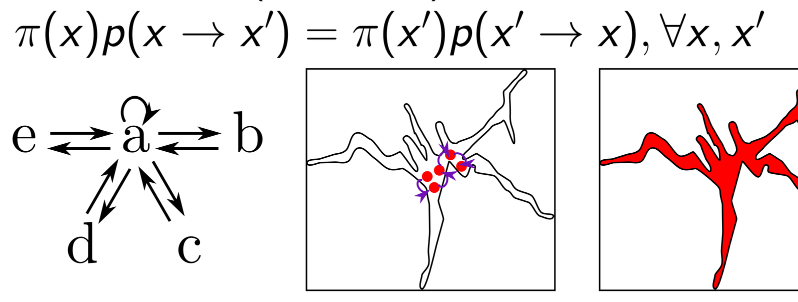
Forward event-chain: replacing randomness by knowledge

- Forward event-chain [1]: irreversible MCMC algorithm that picks **efficient exploration directions** by using gradient information.
- **Accelerations of several orders of magnitude** that increase with dimension (compared to state-of-the-art).
- **No additional complexity** in comparison to state-of-the-art MCMC schemes.

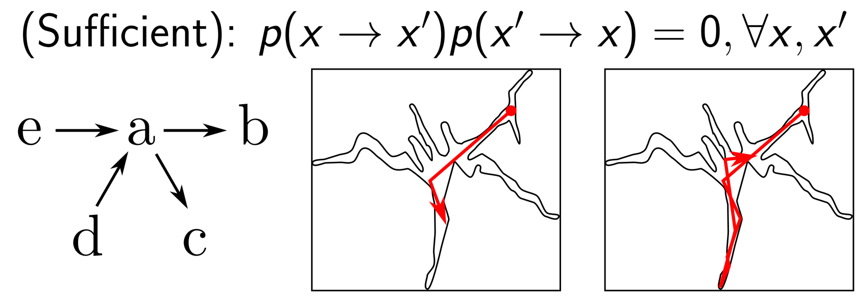
Introduction to Markov chain Monte Carlo (MCMC) methods

- Goal: sample from a target distribution $\pi(x) = \exp(-E(x))$, with $E \sim$ energy.
- Run a Markov chain with transition rates $p(x \to x')$ so that the stationary distribution $= \pi$.
- Global balance (necessary): $\int \pi(x)p(x \to x')dx' = \int \pi(x')p(x' \to x)dx' = \pi(x)$.
- Metropolis filter [4]: $p_{\text{Metro}}(x \to x') = \min(1, \pi(x')/\pi(x)) = \exp(-[\Delta E]^+)$ ($[\Delta E]^+ = \max(0, \Delta E)$).
- Factorized Metropolis filter [3]: $p_{\text{Fact}}(x \to x') = \exp(-\sum_i [\Delta_i E]^+)$, with $E(x) = \sum_i E_i(x)$.

Reversibility (sufficient): Detailed balance
$\pi(x)p(x \to x') = \pi(x')p(x' \to x), \forall x, x'$



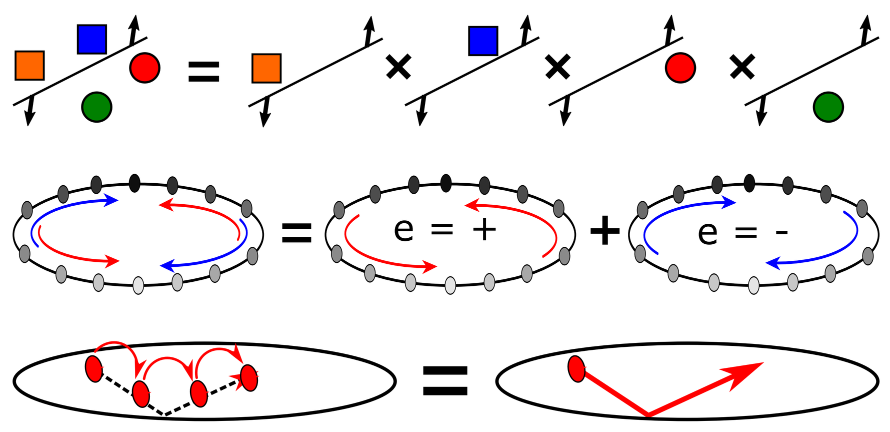
Irreversibility: Maximal global balance
(Sufficient): $p(x \to x')p(x' \to x) = 0, \forall x, x'$



EC method: irreversibility through consistant direction

Key concepts

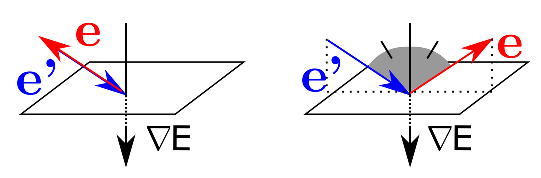
- **Factorized Metropolis filter**
  → Collection of independent 1D problems
- **Lifting [5]**
  → Persistent moves by extending the state space $\pi(x) \to \pi(x, e) = \pi(x)f(e)$
- **Infinitesimal steps/ Continuous time**
  → Infinite number of samples



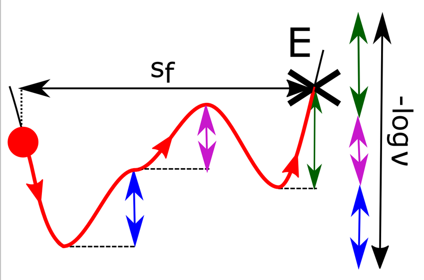
Global balance enforced by turning former rejections into direction changes

- Lifted global balance: $\int \rho_x(-e'|e)f(-e)[-\nabla E \cdot e]^+ de' = \int \rho_x(e|e')f(e')[\nabla E \cdot e']^+ de'$.
- Standard EC implementations, $f(\cdot)$ uniform on S^N and

$$\rho_x(e|e') = \begin{cases} \delta(e - (-e')) & \text{Straight EC} \\ \delta(e - R(e', \nabla E(x))) & \text{Bouncy EC} \end{cases}$$



Implementation: inhomogeneous Poisson process (PP)

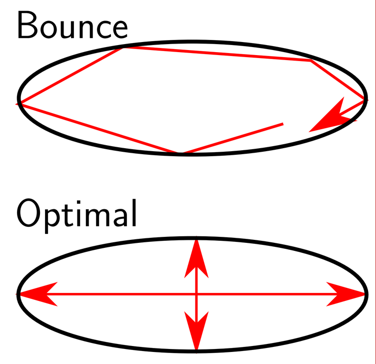


- From state (x, e), sample for each factor next lifting event, ruled by the PP: $p_{\text{event}}(s_f) = \exp\left(-\int_0^{s_f}[dE(x + s \cdot e)]^+\right) = \nu \sim \mathcal{U}[0, 1]$.
- Total cumulative increase in energy $= -\log \nu = \Delta E^*$.
- Update $x \to x + s_f \cdot e$ (If there are several factors, update to the smallest s_f). Then, update e according to $\rho_x(\cdot|e)$.
- Sampling uniformly along the trajectory (e.g. after a fixed length ℓ or at a PP's arrivals).

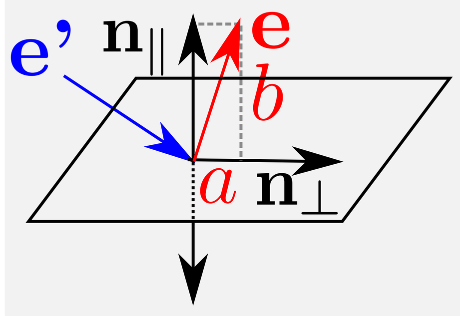
Forward EC: Breaking free from iso-energetic lines by using gradient information – No more resampling

Standard EC Limitations

- No symmetry ($\neq$ Physics), no clear optimal directions: implementation of bouncy events but, as HMC, leads to roughly iso-energetic trajectories without direction resampling.
- Not irreversible in terms of energy: at each lifting event, the previous increase in energy is exactly undone.



Solution: Direct direction pick [1]



- Transition probability from in vector to out vector $e' \to e$
  $\rho(e' \to e) \propto \frac{[-\nabla E \cdot e]^+}{||\nabla E||}$ = Direct sampling
- Thanks to the normalization,
  $\int_{e' IN} [\nabla E \cdot e']^+ = \int_{e OUT} [-\nabla E \cdot e]^+ \propto ||\nabla E||$

Implementation: Update of the lifting variable e

- Direct sampling of $e = a n_\perp + b n_\parallel$ with $\nu \sim \mathcal{U}[0, 1]$, N the dimension and $a^{N-1} = \nu$ and $b = \sqrt{1 - a^2}$ with $n_\parallel = \frac{-\nabla E}{||\nabla E||}$ and $n_\perp = \frac{e' - n_\parallel \cdot e' n_\parallel}{||e' - n_\parallel \cdot e' n_\parallel||}$
- Ergodicity: orthogonal switch of two random components of $n_\perp$: $(n_{\perp,i}^{\text{Switch}}, n_{\perp,j}^{\text{Switch}}) = (-n_{\perp,j}, n_{\perp,i})$

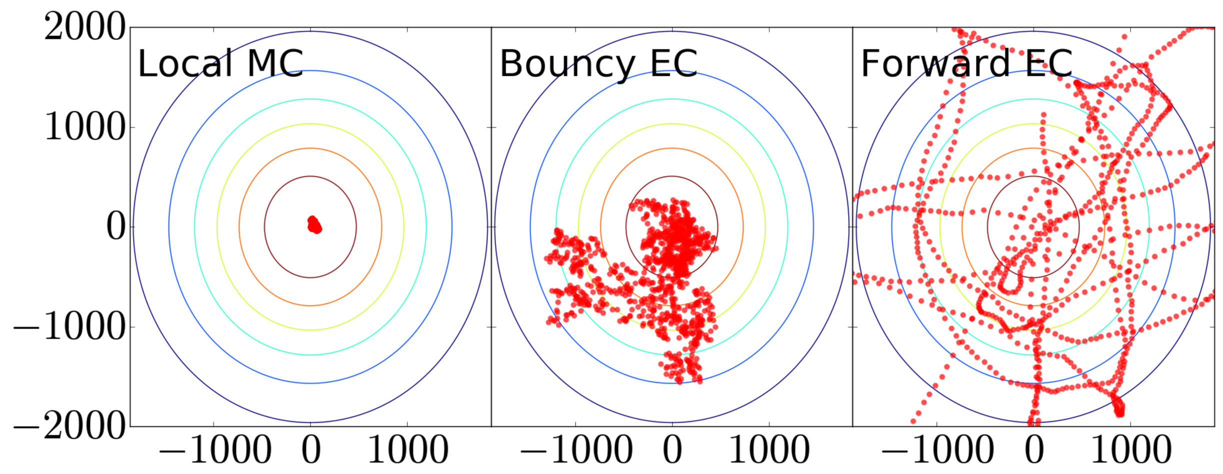
Performance: Forward EC brings accelerations up to several orders of magnitude in comparison to standard EC/HMC

- Ill-conditioned Gaussian distribution of dimension N
- $N = 25, 50, 100, 200, 400$
- $\pi(x) \propto \exp(-E)$, E = energy $= \sum_{i=0}^N x_i^2/\sigma_i^2$
- σ_i^2 log-linearly distributed between 1 and 10^6.

- Local MC = Metropolis;
- Bouncy EC = EC with bounce events
- Optimized EC = EC with straight events on optimal directions

- Forward EC = Forward EC with tuned switch rate
- Forward Switch EC = Forward EC with switch at each event
- Forward Refresh EC = Forward EC with no switch but a direction resampling

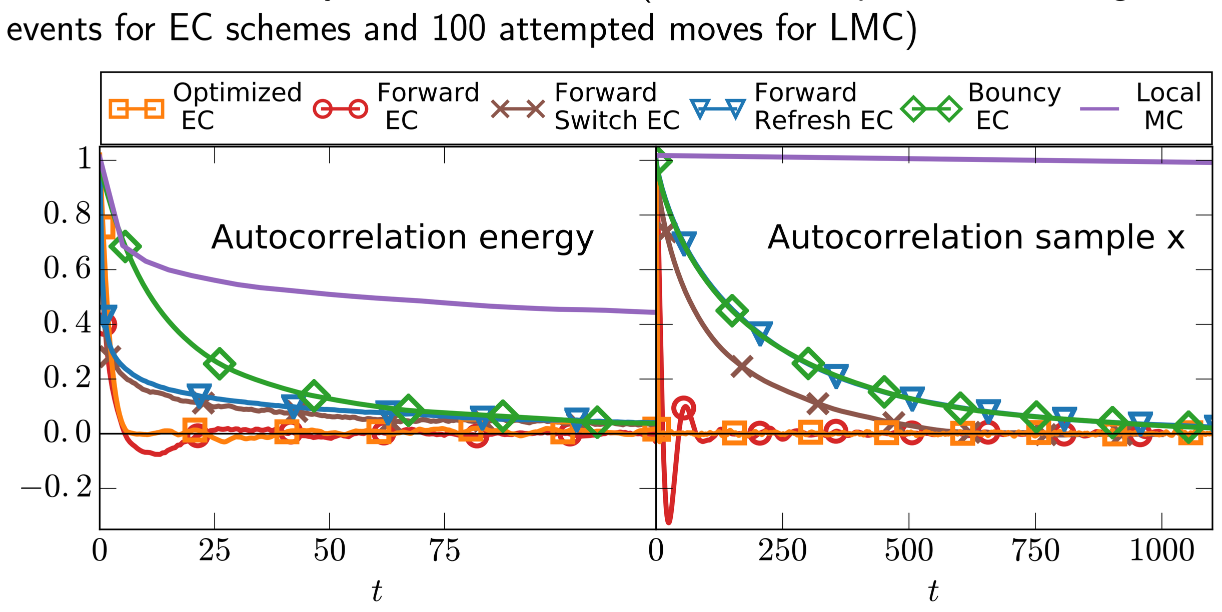
1000 samples (∼ same CPU time) for a distribution of dimension 100 (section on last two dimensions)



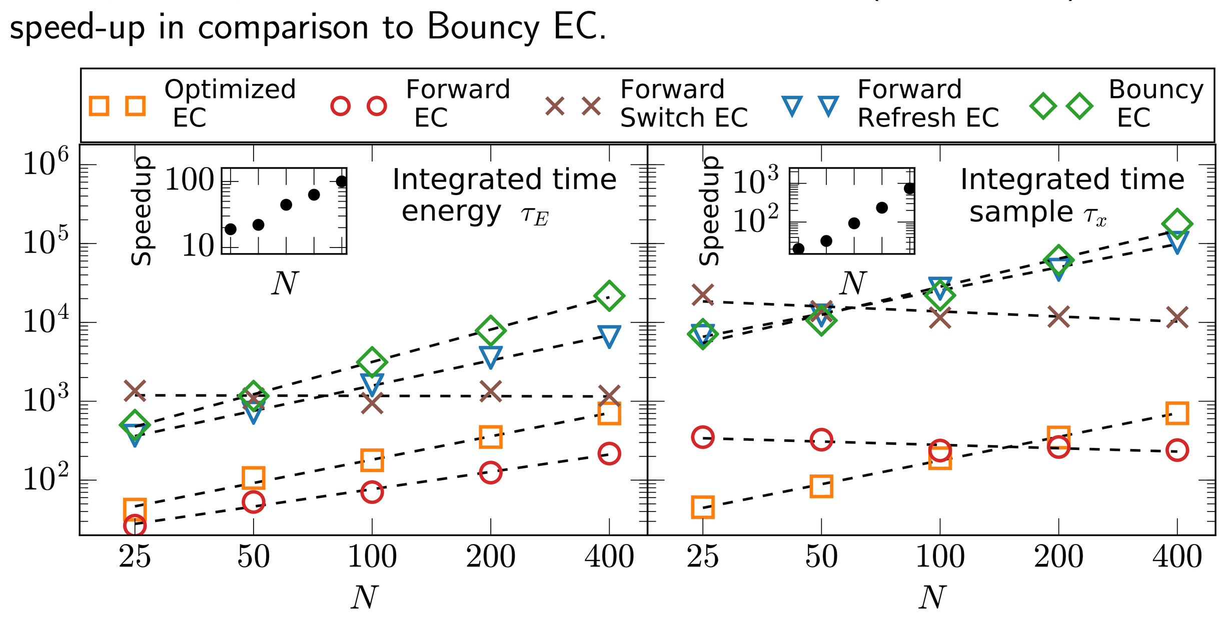
Speed-up in terms of ESS

	Local MC		Bouncy EC		Forward Refresh EC		Optimized EC	
N	E	x	E	x	E	x	E	x
25	7×10^3	8×10^3	19	2.0×10	14	1.9×10	1.6	0.13
50	5×10^3	2×10^4	22	3.2×10	14	3.8×10	2.0	0.26
100	8×10^3	9×10^4	44	9.3×10	23	1.1×10^2	2.5	0.80
200	–	–	63	2.3×10^2	29	1.8×10^2	2.8	1.3
400	–	–	100	7.4×10^2	30	4.3×10^2	3.2	2.9

Autocorrelation decay for dimension 100 (Unit: 100 steps, i.e. 100 lifting events for EC schemes and 100 attempted moves for LMC)



Scaling of integrated autocorrelation time with dim. N (Unit: 1 step). Inset: speed-up in comparison to Bouncy EC.



Forward EC and EC pseudocodes

Input: $E(.)$ target energy, ℓ chain length, N_S final number of samples
Initialize: $x^\star \sim p_0(.)$, direction e randomly in the unit sphere S^N, current chain length $\ell_c \leftarrow \ell$, $k \leftarrow 0$
while $k < N_S$ **do**
  Set initial sample as $x \leftarrow x^\star$
  Sample $\Delta E^* = -\log(\nu)$ with $\nu \sim \mathcal{U}[0, 1]$
  Compute s_f – *displacement before next lifting event*
  if $\ell_c < s_f$ **then**
    Compute $x^\star = x + \ell_c e$
    Set new sample $x^{(k)} \leftarrow x^\star$ and $k \leftarrow k + 1$
    Update chain length to $\ell_c \leftarrow \ell_c + \ell - s_f$
    if Not Forward EC **then**
      Resample e randomly in the unit sphere S^N
    end if
  end if
  Compute $x^\star = x + s_f e$
  Update chain length to $\ell_c \leftarrow \ell_c - s_f$
  Pick a new direction e according to $\rho_x(\cdot|e)$
end while

References

[1] M. Michel and S. Sénécal, *arXiv:1702.08397* (2017).
[2] M. Michel, *Irreversible Markov chains by the factorized filter*, thesis tel.archives-ouvertes.fr/tel-01394204
[3] M. Michel, S. C. Kapfer and W. Krauth, *J. Chem. Phys.* **140**, 54116 (2014).
[4] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
[5] P. Diaconis, S. Holmes, and R. M. Neal, *Annals of Applied Probability* **10**(3):726–752 (2000).
[6] A. Bouchard-Côté, S. J. Vollmer and A. Doucet, *JASA* (2017).