Advanced topics in Markov-chain Monte Carlo

Lecture 2:
Diameters and conductances, liftings, path graph
Part 2/2: Lifting / Examples

Werner Krauth

ICFP -Master Course Ecole Normale Supérieure, Paris, France

26 January 2022

References

- F. Chen, L. Lovasz, I. Pak, **Lifting Markov Chains to Speed up Mixing.** Proceedings of the 17th Annual ACM Symposium on Theory of Computing, 275 (1999) `http://www.math.ucla.edu/~pak/papers/stoc2.pdf`
- P. Diaconis, S. Holmes, R. M. Neal, **Analysis of a nonreversible Markov chain sampler** Ann. Appl. Probab. 10, 726–752 (2000) `https://projecteuclid.org/download/pdf_1/euclid.aoap/1019487508`
- M. Hildebrand, **Rates of convergence of the Diaconis-Holmes-Neal Markov chain sampler with a V-shaped stationary probability**, Markov Proc. Rel. Fields 10, 687–704 (2004)
- W. Krauth, **Event-Chain Monte Carlo: Foundations, Applications, and Prospects**, Front. Phys. 9:663457. `https://www.frontiersin.org/article/10.3389/fphy.2021.663457`

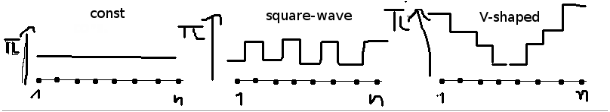
Metropolis algorithm on path graph (1/4)

- path graph $P_n = (\Omega_n, E_n) \ldots$
- vertices $\Leftrightarrow$ sample space $\Omega_n = \{1, \ldots, n\}$
- edges $E_n = \{(1,2), \ldots, (n-1, n)\} \Leftrightarrow$ non-zero P_{ij}
- one-d n-site lattice without pbc.
- stationary distribution $\pi = \{\pi_1, \ldots, \pi_n\} \Leftrightarrow$ GBC.
- Phantom vertices 0 and $n+1$ with $\pi_0 = \pi_{n+1} = 0$, and
- Phantom edges $(0, 1)$ and $(n, n+1)$.

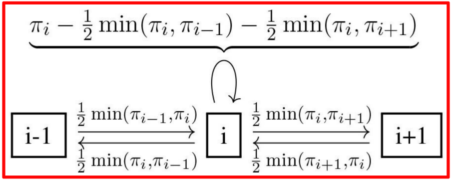
Metropolis algorithm:

1. From vertex $i \in \Omega$ choose $j = i \pm 1$ with probability $1/2$
2. accept $i \to j$ with probability $\min(1, \pi_j/\pi_i)$
3. else stay at i.
4. NB: Phantoms are nice

Metropolis algorithm on path graph (2/4)

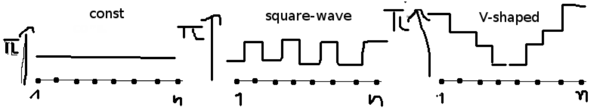


- Checking detailed balance ($\pi_i P_{ij} = \pi_j P_{ji}$):
- Checking global balance ($\sum_i \pi_i P_{ij} = \pi_j$):



- Irreducibility OK
- Aperiodicity OK, thanks to boundaries

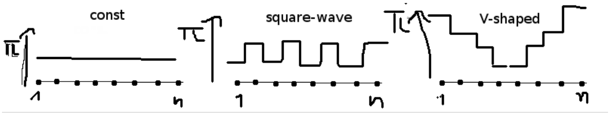
Metropolis algorithm on path graph (3/4)



- Constant: $\pi_i = \frac{1}{n} \; \forall i \in \Omega$.
  Conductance $\Phi = \mathcal{O}(1/n)$.
- Square wave: $\pi_{2k-1} = \frac{2}{3n}$, $\pi_{2k} = \frac{4}{3n}$.
  Conductance $\Phi = \frac{2}{3n}$ (for $n \to \infty$).
- V-shaped: $\pi_i = \text{const}|\frac{n+1}{2} - i| \; \forall i \in \Omega$, where const $= \frac{4}{n^2}$.
  Conductance $\Phi = \frac{2}{n^2}$.

NB: Graph diameter n.
NNB: π normalized.
NNNB: Bottleneck between $i = \frac{n}{2}$ and $j = \frac{n}{2} + 1$.

Metropolis algorithm on path graph (4/4)



- Constant: $\pi_i = \frac{1}{n} \; \forall i \in \Omega$.
  Conductance $\Phi = \mathcal{O}(1/n)$.
  Mixing time $\mathcal{O}\left(n^2\right)$. Markov chain is transport-limited.
- Square wave: $\pi_{2k-1} = \frac{2}{3n}$, $\pi_{2k} = \frac{4}{3n}$.
  Conductance $\Phi = \frac{2}{3n}$ (for $n \to \infty$).
  Mixing time $\mathcal{O}\left(n^2\right)$. Markov chain is transport-limited.
- V-shaped: $\pi_i = \text{const}|\frac{n+1}{2} - i| \; \forall i \in \Omega$, where const $= \frac{4}{n^2}$.
  Conductance $\Phi = \frac{2}{n^2}$.
  Mixing time $\mathcal{O}\left(n^2 \log n\right)$. Markov chain is conductance-limited (up to a log).
  NB: Mixing time in $\mathcal{S}$ is $\mathcal{O}\left(n^2\right)$.

Lifting (Chen et al (1999)) (1/2)

- Markov chain $\Pi \Leftrightarrow$ Lifted Markov chain $\hat{\Pi}$
- Ω (sample space) $\Leftrightarrow$ $\hat{\Omega}$ (lifted sample space)
- P (transition matrix) $\Leftrightarrow$ $\hat{P}$ (lifted transition matrix)
- **Condition 1**: sample space is copied ("lifted"), π preserved

$$\pi_v = \hat{\pi}\left[f^{-1}(v)\right] = \sum_{i \in f^{-1}(v)} \hat{\pi}_i,$$

- **Condition 2**: flows are preserved

$$\underbrace{\pi_v P_{vu}}_{\text{collapsed flow}} = \sum_{i \in f^{-1}(v), j \in f^{-1}(u)} \overbrace{\hat{\pi}_i \hat{P}_{ij}}^{\text{lifted flow}} .$$

- Usually: $\hat{\Omega} = \Omega \times \mathcal{L}$, with $\mathcal{L}$ a set of lifting variables σ

Lifting (Chen et al (1999)) (2/2)

- **Condition 1**: sample space is copied ("lifted"), π preserved

$$\pi_v = \hat{\pi}\left[f^{-1}(v)\right] = \sum_{i \in f^{-1}(v)} \hat{\pi}_i,$$

- **Condition 2**: flows are preserved

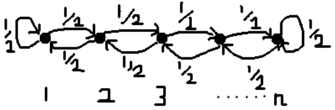
$$\underbrace{\pi_v P_{vu}}_{\text{collapsed flow}} = \sum_{i \in f^{-1}(v), j \in f^{-1}(u)} \overbrace{\hat{\pi}_i \hat{P}_{ij}}^{\text{lifted flow}} .$$

- Lifting does not increase the conductance (TD).

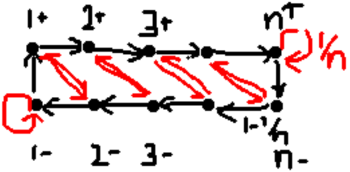
Lifting on the path graph (1/4)

Probability distribution $\pi = (1/n, \ldots, 1/n)$ (Diaconis et al. 2000)

- “Collapsed” Markov chain:

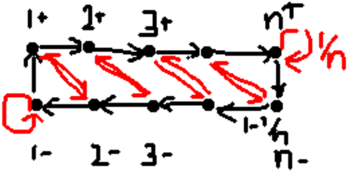


- “Lifted” Markov chain $\hat{\Omega} = \Omega \times \{-, +\}$:

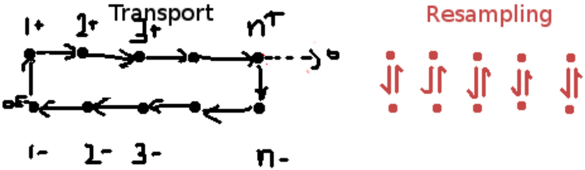


Probability distribution $\pi = (1/n, \ldots, 1/n)$ (Diaconis et al. 2000)

- "Lifted" Markov chain $\hat{\Omega} = \Omega \times \{-, +\}$:



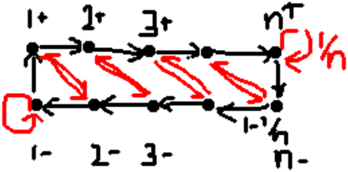
- Two-step version



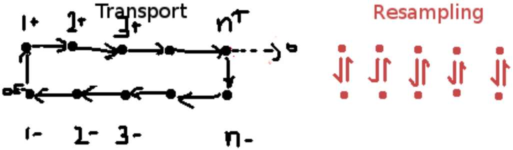
Lifting on the path graph (3/4)

Probability distribution $\pi = (1/n, \ldots, 1/n)$ (Diaconis et al. 2000)

- “Lifted” Markov chain $\hat{\Omega} = \Omega \times \{-, +\}$:



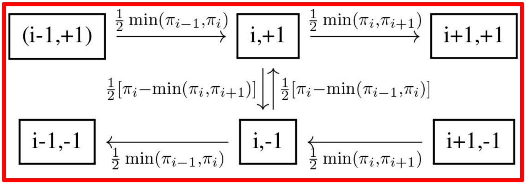
- Two-step version



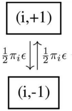
- Analyze, then generalize the behavior at the boundaries

Lifting on the path graph (4/4)

- “Lifted Markov chain: Transport”

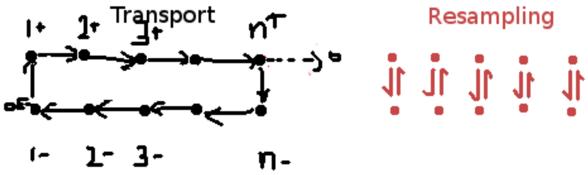


- “Lifted Markov chain: Resampling”



- Resampling can often be dropped

Lifting and global balance



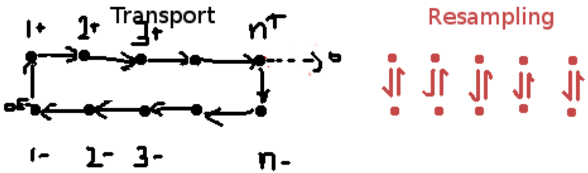
- Flow into configuration $(x, +)$ (transport):

$$(x-1,+) \to (x,+) \quad \mathcal{A}_x^+ = \min(\pi_{x-1}^+, \pi_x^+)$$
$$(x,-) \to (x,+) \quad \mathcal{R}_x^+ = \pi_x^- - \min(\pi_{x-1}^-, \pi_x^-)$$

- Flow into configuration (resampling):

$$(x,+) \to (x,+) \quad \mathcal{L}_x^{++} = (1-\lambda)\pi_x^+ = \tfrac{1}{2}(1-\lambda)\pi(x)$$
$$(x,-) \to (x,+) \quad \mathcal{L}_x^{-+} = \lambda\pi_x^- = \tfrac{1}{2}\lambda\pi(x)$$

Lifting and mixing



- The V-shaped stationary distribution is an ideal model to test the lifted Metropolis algorithm.
- It has conductance $\mathcal{O}\left(2/n^2\right)$, so mixing at least $\sim n^2$
- Collapsed Metropolis: $t_{\text{mix}} = \mathcal{O}\left(n^2 \log n\right)$
- Lifted Metropolis: $t_{\text{mix}} = \mathcal{O}\left(n^2\right)$
- Lifted Metropolis (restricted to $S = \{1, \ldots, \frac{n}{2}\}$) $t_{\text{mix}}^{\text{restricted}} = \mathcal{O}\left(n\right)$ with TVD decreasing to $\frac{1}{2}$.