Solutions 10

1. a) Starting from the base chain whose transition probabilities are

$$\psi_{xy} = \begin{cases} 1/d & \text{if } y \sim x \\ 0 & \text{otherwise} \end{cases}$$

we set out to sample (approximately) from the distribution $\pi_\beta(x) = \exp(-\beta f(x))/Z_\beta$, where Z_β is the normalization constant.

The acceptance probabilities are given by

$$a_{xy} = \min\{1, \exp(-\beta(f(y) - f(x))\}, \quad (\text{for } y \sim x)$$

so the transition probabilities of the Metropolis chain are given by

$$p_{xy} = \begin{cases} (1/d)\ \min\{1, \exp(-\beta(f(y) - f(x))\} & \text{if } y \sim x \\ (1/d) \sum_{z \sim x} (1 - \min\{1, \exp(-\beta(f(z) - f(x))\}) & \text{if } y = x \\ 0 & \text{otherwise} \end{cases}$$

Running then the Metropolis chain for a carefully chosen duration and sampling from it, we obtain an approximate sample x of π_β which, for a carefully chosen value of β, should give a reasonably low value of $f(x)$.

b) In this particular case (which by the way does not really need the Metropolis algorithm, as the function f is quite easy to minimize here!), we obtain

$$a_{xy} = \begin{cases} \exp(-\beta) & \text{if } |y| = |x| + 1 \\ 1 & \text{if } |y| = |x| - 1 \\ 0 & \text{otherwise} \end{cases}$$

and therefore

$$p_{xy} = \begin{cases} (1/d)\ \exp(-\beta) & \text{if } |y| = |x| + 1 \\ 1/d & \text{if } |y| = |x| - 1 \\ (1/d) \sum_{z : |z| = |x|+1} (1 - \exp(-\beta)) = \frac{d - |x|}{d} (1 - \exp(-\beta)) & \text{if } y = x \\ 0 & \text{otherwise} \end{cases}$$

c) In this case, π_β reads

$$\pi_\beta(x) = \exp(-\beta |x|)/Z_\beta$$

with normalization constant

$$Z_\beta = \sum_{x \in S} \exp(-\beta |x|) = \sum_{x_1, \ldots, x_d \in \{0,1\}} \exp(-\beta(x_1 + \ldots + x_d)) = \left(\sum_{x_1 \in \{0,1\}} \exp(-\beta x_1) \right)^d = (1 + e^{-\beta})^d$$

2. a) The transition probabilities are given by

$$\begin{aligned} p_{01} &= \psi_{01}\min\left(1, \tfrac{\psi_{10}\pi_1}{\psi_{01}\pi_0}\right) = \tfrac{e^{-2\beta}}{2} & p_{21} &= \psi_{21}\min\left(1, \tfrac{\psi_{12}\pi_1}{\psi_{21}\pi_2}\right) = \tfrac{e^{-\beta}}{2} \\ p_{10} &= \psi_{10}\min\left(1, \tfrac{\psi_{01}\pi_0}{\psi_{10}\pi_1}\right) = \tfrac{1}{2} & p_{12} &= \psi_{12}\min\left(1, \tfrac{\psi_{21}\pi_2}{\psi_{12}\pi_1}\right) = \tfrac{1}{2} \\ p_{02} &= p_{20} = p_{11} = 0 & p_{00} &= 1 - \tfrac{e^{-2\beta}}{2} \qquad p_{22} = 1 - \tfrac{e^{-\beta}}{2} \end{aligned} \tag{1}$$

b) Let us now check that the detailed balance equation is satisfied:

$$\begin{aligned} p_{01}\pi_0 &= \frac{1}{2}e^{-2\beta} = p_{10}\pi_1 \\ p_{02}\pi_0 &= 0 = p_{20}\pi_2 \\ p_{12}\pi_1 &= \frac{1}{2}e^{-2\beta} = p_{21}\pi_2. \end{aligned}$$

c) As usual, there are several methods to compute the eigenvalues. For example, one can find the three solutions λ_0, λ_1, and λ_2 to the equation

$$\det(P - \lambda I) = 0, \tag{2}$$

where I is the 3×3 identity matrix and P the matrix of the transition probabilites computed in (1).

Another (perhaps even simpler method) method is to solve the following system of equations:

$$\begin{cases} \lambda_0 = 1 \\ \lambda_0 + \lambda_1 + \lambda_2 = \mathrm{tr(P)} \\ \lambda_0 \cdot \lambda_1 \cdot \lambda_2 = \det(P) \end{cases},$$

as we know that the largest eigenvalue is 1, the sum of the eigenvalues equals the trace of P, and their product equals the determinant of P.

Consequently, we obtain

$$\begin{cases} \lambda_0 = 1 \\ \\ \lambda_1 = -\dfrac{e^{-2\beta}}{4} - \dfrac{e^{-\beta}}{4} + \dfrac{1}{2} + \dfrac{1}{4}\sqrt{e^{-4\beta} - 2e^{-3\beta} + e^{-2\beta} + 4} \\ \\ \lambda_2 = -\dfrac{e^{-2\beta}}{4} - \dfrac{e^{-\beta}}{4} + \dfrac{1}{2} - \dfrac{1}{4}\sqrt{e^{-4\beta} - 2e^{-3\beta} + e^{-2\beta} + 4} \end{cases},$$

d) The spectral gap is given by

$$\gamma = 1 - \lambda_1 = \frac{1}{2} + \frac{e^{-2\beta}}{4} + \frac{e^{-\beta}}{4} - \frac{1}{4}\sqrt{e^{-4\beta} - 2e^{-3\beta} + e^{-2\beta} + 4}. \tag{3}$$

Therefore, when β is large, we have

$$\gamma \approx \frac{1}{4}e^{-\beta}. \tag{4}$$

Remark. The value of β has to be tuned carefully and there is an inherent trade-off in its choice. If we pick β too large, then the spectral gap is small and the convergence to the global minimum occurs very slowly, which reflects the fact that we might get stuck in the local minimum (=state 2). On the other hand, if we pick β too small, then convergence is fast, but the stationary distribution is close to uniform in this case, so there is no guarantee that we land in the global minimum either!