NNDL Exercise set 6 – Mathematical exercises

1.

We have $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

a)

For step size μ the Langevin iteration is defined as

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mu\phi_{\mathbf{x}}(\mathbf{x}_t) + \sqrt{2\mu}\mathbf{n}_t, \tag{1}$$

where $\phi_{\mathbf{x}}(\mathbf{x}_t) = -\mathbf{\Sigma}^{-1}\mathbf{x}$ is the score function. Thus we can write the Langevin iteration as

$$\mathbf{x}_{t+1} = \mathbf{M}\mathbf{x}_t + \alpha\mathbf{n}_t, \tag{2}$$

where $\mathbf{M} = \mathbf{I} - \mu\mathbf{\Sigma}^{-1}$ and $\alpha = \sqrt{2\mu}$.

b)

$$\begin{aligned}
\mathbf{x}_1 &= \mathbf{M}\mathbf{x}_0 + \alpha\mathbf{n}_0 \\
\mathbf{x}_2 &= \mathbf{M}\mathbf{x}_1 + \alpha\mathbf{n}_1 = \mathbf{M}^2\mathbf{x}_0 + \alpha\mathbf{M}\mathbf{n}_0 + \alpha\mathbf{n}_1 \\
\mathbf{x}_3 &= \mathbf{M}\mathbf{x}_2 + \alpha\mathbf{n}_2 = \mathbf{M}^3\mathbf{x}_0 + \alpha\mathbf{M}^2\mathbf{n}_0 + \alpha\mathbf{M}\mathbf{n}_1 + \alpha\mathbf{n}_2
\end{aligned} \tag{3}$$

c)

$$\mathbf{x}_T = \mathbf{M}^T\mathbf{x}_0 + \alpha\sum_{i=0}^{T-1}\mathbf{M}^{T-1-i}\mathbf{n}_i \tag{4}$$

d)

We have $\mathbf{\Sigma} = \mathbf{I} \implies \mathbf{M} = (1-\mu)\mathbf{I}$ and thus $\mathbf{M}^T = (1-\mu)^T\mathbf{I}$. Assuming $0 < \mu < 1$, the expression $(1-\mu)^T \to 0$ as $T \to \infty$ and therefore $\mathbf{M}^T\mathbf{x}_0 \to 0$ as $T \to \infty$, so the influence of the initial point reduces to zero.

2.

We have

$$x = az + n, \tag{5}$$

where $z \sim \mathcal{N}(0,1)$ and $n \sim \mathcal{N}(0,\sigma^2)$.

a)

Now

$$p(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right). \tag{6}$$

For $p(x|z)$ we can use the affine transformation rule for Gaussians, since z is now fixed (a constant), i.e. if $X \sim \mathcal{N}(\mu,\sigma^2)$, then $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$. Plugging in $X = n$, $a = 1$ and $b = az$, we obtain

$$p(x|z) = \mathcal{N}(az,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-az)^2}{2\sigma^2}\right). \tag{7}$$

The joint pdf $p(x,z)$ is then

$$p(x,z) = p(x|z)p(z) = \frac{1}{2\pi\sigma} \exp\left(-\frac{(x-az)^2}{2\sigma^2} - \frac{z^2}{2}\right). \tag{8}$$

b)

For $X_i \sim \mathcal{N}(\mu_i,\sigma_i^2)$ we have the general result

$$\sum_{i=1}^{n} a_i X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right) \tag{9}$$

for some constant a_i and assuming X_i to be independent $\forall i$. Since $x = az + n$, where z and n are independent, we obtain

$$p(x) = \mathcal{N}(0, a^2+\sigma^2) = \frac{1}{\sqrt{2\pi(a^2+\sigma^2)}} \exp\left(-\frac{x^2}{2(a^2+\sigma^2)}\right). \tag{10}$$

c)

Given a sample $X = (x_1,\ldots,x_N)$ the log likelihood is

$$\begin{aligned}
\mathcal{L}(a) &= \log p(X) \\
&= \sum_{i=1}^{N} \log p(x_i) \\
&= \sum_{i=1}^{N} \left(-\frac{1}{2}\log(2\pi(a^2+\sigma^2)) - \frac{x_i^2}{2(a^2+\sigma^2)}\right) \\
&= -\frac{N}{2}\log(2\pi(a^2+\sigma^2)) - \frac{1}{2(a^2+\sigma^2)}\sum_{i=1}^{N} x_i^2.
\end{aligned} \tag{11}$$

Proceed by setting the derivative w.r.t. a to zero and solving for a.

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial a} &= -\frac{Na}{(a^2+\sigma^2)} + \frac{a}{(a^2+\sigma^2)^2}\sum_{i=1}^{N} x_i^2 = 0 \\
\implies N(a^2+\sigma^2) &= \sum_{i=1}^{N} x_i^2 \\
\implies a^2 &= \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \sigma^2 \\
\implies \hat{a} &= \pm\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \sigma^2}.
\end{aligned} \tag{12}$$

3.

Starting from the right side and using the general properties $\min -f(x) = -\max f(x)$ and $\max -f(x) = -\min f(x)$ for some function $f(x)$, we obtain

$$\begin{aligned}
-\max_a \min_b -J(a,b) &= -\max_a -\max_b J(a,b) \\
&= \min_a \max_b J(a,b).
\end{aligned} \tag{13}$$

4.

We have $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and a matrix $\mathbf{U}$ for which $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}$. To find the distribution of $\mathbf{y} = \mathbf{U}\mathbf{x}$, we first note that a linear transformation of a Gaussian is also a Gaussian, thus $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, for some $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. To find $\boldsymbol{\mu}$, we compute the expectation

$$\mathbb{E}[\mathbf{U}\mathbf{x}] = \mathbf{U}\underbrace{\mathbb{E}[\mathbf{x}]}_{=\mathbf{0}} = \mathbf{0}. \tag{14}$$

So $\mathbf{y}$ has a mean of $\boldsymbol{\mu} = \mathbf{0}$. The covariance $\boldsymbol{\Sigma}$ of $\mathbf{y}$ is then

$$\begin{aligned}
\boldsymbol{\Sigma} &= \operatorname{cov}(\mathbf{y}) \\
&= \operatorname{cov}(\mathbf{U}\mathbf{x}) \\
&= \mathbf{U}\underbrace{\operatorname{cov}(\mathbf{x})}_{=\mathbf{I}}\mathbf{U}^T \\
&= \mathbf{U}\mathbf{U}^T \\
&= \mathbf{I}.
\end{aligned} \tag{15}$$

Therefore $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and hence is also white.