My solutions to

Deep Learning: Foundations and Concepts

Dario Miro Konopatzki

20 Diffusion Models

20.1

Mean

$$\begin{aligned}
\mathbb{E}[Z_t] &= \mathbb{E}\left[\sqrt{1-\beta_t}Z_{t-1} + \sqrt{\beta_t}\mathcal{E}_t\right] \\
&= \sqrt{1-\beta_t}\mathbb{E}[Z_{t-1}] + \sqrt{\beta_t}\mathbb{E}[\mathcal{E}_t] \qquad \text{linearity of } \mathbb{E} \\
&= \sqrt{1-\beta_t}\mathbb{E}[Z_{t-1}] \qquad \mathcal{E}_t \sim \mathcal{N}(0,1) \text{ so i.p. } \mathbb{E}[\mathcal{E}_t] = 0
\end{aligned}$$

$$\begin{aligned}
\|\mathbb{E}[Z_t]\| &= \left\|\sqrt{1-\beta_t}\mathbb{E}[Z_{t-1}]\right\| \\
&= \left|\sqrt{1-\beta_t}\right| \|\mathbb{E}[Z_{t-1}]\| \\
&< \|\mathbb{E}[Z_{t-1}]\| \qquad \left|\sqrt{1-\beta_t}\right| < 1 \text{ since } 0 < \beta_t < 1
\end{aligned}$$

Auxiliary Calculations

$$\begin{aligned}
\mathbb{E}\left[\mathcal{E}_t\mathcal{E}_t^\top\right] &= \mathbb{E}\left[\mathcal{E}_t\mathcal{E}_t^\top\right] - \mathbb{V}[\mathcal{E}_t] + \mathbb{V}[\mathcal{E}_t] \\
&= \mathbb{E}\left[\mathcal{E}_t\mathcal{E}_t^\top\right] - \left(\mathbb{E}\left[\mathcal{E}_t\mathcal{E}_t^\top\right] - \mathbb{E}[\mathcal{E}_t]\mathbb{E}[\mathcal{E}_t]^\top\right) + \mathbb{V}[\mathcal{E}_t] \\
&= \mathbb{E}\left[\mathcal{E}_t\mathcal{E}_t^\top\right] - \mathbb{E}\left[\mathcal{E}_t\mathcal{E}_t^\top\right] + \mathbb{E}[\mathcal{E}_t]\mathbb{E}[\mathcal{E}_t]^\top + \mathbb{V}[\mathcal{E}_t] \\
&= \mathbb{E}[\mathcal{E}_t]\mathbb{E}[\mathcal{E}_t]^\top + \mathbb{V}[\mathcal{E}_t] \\
&= \mathbb{I} \qquad \mathbb{E}[\mathcal{E}_t] = 0,\ \mathbb{V}[\mathcal{E}_t] = \mathbb{I} \text{ since by assumption } \mathcal{E}_t \sim \mathcal{N}(0,\mathbb{I})
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Z_tZ_t^\top] &= \mathbb{E}\left[\left(\sqrt{1-\beta_t}Z_{t-1} + \sqrt{\beta_t}\mathcal{E}_t\right)\left(\sqrt{1-\beta_t}Z_{t-1} + \sqrt{\beta_t}\mathcal{E}_t\right)^\top\right] \\
&= \mathbb{E}\Big[(1-\beta_t)Z_{t-1}Z_{t-1}^\top + \sqrt{1-\beta_t}\sqrt{\beta_t}Z_{t-1}\mathcal{E}_t^\top
\end{aligned}$$

$$
\begin{aligned}
&\quad + \sqrt{\beta_t}\sqrt{1-\beta_t}\mathcal{E}_t Z_{t-1}^\top + \beta_t \mathcal{E}_t \mathcal{E}_t^\top \Big] \\
&= (1-\beta_t)\mathbb{E}\left[Z_{t-1}Z_{t-1}^\top\right] + \sqrt{1-\beta_t}\sqrt{\beta_t}\mathbb{E}\left[Z_{t-1}\mathcal{E}_t^\top\right] \\
&\quad + \sqrt{\beta_t}\sqrt{1-\beta_t}\mathbb{E}\left[\mathcal{E}_t Z_{t-1}^\top\right] + \beta_t\mathbb{E}\left[\mathcal{E}_t\mathcal{E}_t^\top\right] && \mathbb{E} \text{ linear} \\
&= (1-\beta_t)\mathbb{E}\left[Z_{t-1}Z_{t-1}^\top\right] + \sqrt{1-\beta_t}\sqrt{\beta_t}\mathbb{E}\left[Z_{t-1}\right]\mathbb{E}\left[\mathcal{E}_t^\top\right] \\
&\quad + \sqrt{\beta_t}\sqrt{1-\beta_t}\mathbb{E}\left[\mathcal{E}_t\right]\mathbb{E}\left[Z_{t-1}^\top\right] + \beta_t\mathbb{E}\left[\mathcal{E}_t\mathcal{E}_t^\top\right] && Z_{t-1} \perp \mathcal{E}_t \\
&\overset{(\star)}{=} (1-\beta_t)\mathbb{E}\left[Z_{t-1}Z_{t-1}^\top\right] + \beta_t\mathbb{E}\left[\mathcal{E}_t\mathcal{E}_t^\top\right] \\
&= (1-\beta_t)\mathbb{E}\left[Z_{t-1}Z_{t-1}^\top\right] + \beta_t\mathbb{I}
\end{aligned}
$$

$(\star)$ $\mathcal{E}_t \sim \mathcal{N}(0, \mathbb{I})$, i.p. $\mathbb{E}[\mathcal{E}_t] = 0$

Covariance

$$
\begin{aligned}
\|\mathrm{cov}(Z_t) - \mathbb{I}\| &= \|\mathbb{E}\left[Z_t Z_t^\top\right] - \mathbb{E}[Z_t]\mathbb{E}[Z_t]^\top - \mathbb{I}\| \\
&= \|(1-\beta_t)\mathbb{E}\left[Z_{t-1}Z_{t-1}^\top\right] + \beta_t\mathbb{I} - (1-\beta_t)\mathbb{E}[Z_{t-1}]\mathbb{E}[Z_{t-1}]^\top - \mathbb{I}\| \\
&= \|(1-\beta_t)\left(\mathbb{E}\left[Z_{t-1}Z_{t-1}^\top\right] - \mathbb{E}[Z_{t-1}]\mathbb{E}[Z_{t-1}]^\top - \mathbb{I}\right)\| \\
&= \|(1-\beta_t)\left(\mathrm{cov}(Z_{t-1}) - \mathbb{I}\right)\| \\
&= |1-\beta_t|\|\mathrm{cov}(Z_{t-1}) - \mathbb{I}\| \\
&< \|\mathrm{cov}(Z_{t-1}) - \mathbb{I}\| \qquad |1-\beta_t| < 1 \text{ since } 0 < \beta_t < 1
\end{aligned}
$$

20.2

For every x s.t. $q_X(x) \neq 0$:

$$
\begin{aligned}
q_{Z_1|X=x}(z_1) &= \frac{q_{Z_1,X}(z_1, x)}{q_X(x)} \qquad \text{def. of conditional density} \\
&\overset{(\star)}{=} \frac{q_{\frac{1}{\sqrt{\beta_1}}(Z_1 - \sqrt{1-\beta_1}X), X}\left(\frac{1}{\sqrt{\beta_1}}\left(z_1 - \sqrt{1-\beta_1}x\right), x\right)}{q_X(x)} \\
&\quad \cdot \left|\det\begin{pmatrix} \frac{1}{\sqrt{\beta_1}}\mathbb{I}_D & -\frac{\sqrt{1-\beta_1}}{\sqrt{\beta_1}}\mathbb{I}_D \\ 0 & \mathbb{I}_D \end{pmatrix}\right| \\
&\overset{(\dagger)}{=} \frac{q_{\frac{1}{\sqrt{\beta_1}}(\cancel{\sqrt{1-\beta_1}X} + \sqrt{\beta_1}\mathcal{E}_1 \cancel{- \sqrt{1-\beta_1}X}), X}\left(\frac{1}{\sqrt{\beta_1}}\left(z_1 - \sqrt{1-\beta_1}x\right), x\right)\frac{1}{\sqrt{\beta_1}^D}}{q_X(x)} \\
&= \frac{1}{\sqrt{\beta_1^D}}\frac{q_{\mathcal{E}_1,X}\left(\frac{1}{\sqrt{\beta_1}}\left(z_1 - \sqrt{1-\beta_1}x\right), x\right)}{q_X(x)}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{\beta_1^D}} \frac{q_{\mathcal{E}_1}\left(\frac{1}{\sqrt{\beta_1}}\left(z_1 - \sqrt{1-\beta_1}x\right)\right)\cancel{q_X(x)}}{\cancel{q_X(x)}} \qquad \mathcal{E}_1 \perp X \\
&\overset{\ddagger}{=} \frac{1}{\sqrt{\beta_1^D}} \frac{1}{\sqrt{(2\pi)^D \mathbb{I}_D}} \\
&\quad \cdot e^{-\frac{1}{2}\left(\frac{1}{\sqrt{\beta_1}}\left(z_1-\sqrt{1-\beta_1}x\right)-0\right)^\top \mathbb{I}_D^{-1}\left(\frac{1}{\sqrt{\beta_1}}\left(z_1-\sqrt{1-\beta_1}x\right)-0\right)} \\
&= \frac{1}{\sqrt{(2\pi)^D \beta_1^D \mathbb{I}_D}} e^{-\frac{1}{2}\left(z_1-\sqrt{1-\beta_1}x\right)^\top \frac{1}{\beta_1}\mathbb{I}_D\left(z_1-\sqrt{1-\beta_1}x\right)} \\
&= \frac{1}{\sqrt{(2\pi)^D \det(\beta_1 \mathbb{I}_D)}} e^{-\frac{1}{2}\left(z_1-\sqrt{1-\beta_1}x\right)^\top (\beta_1\mathbb{I}_D)^{-1}\left(z_1-\sqrt{1-\beta_1}x\right)}
\end{aligned}
$$

which is density of distribution $\mathcal{N}\left(\sqrt{1-\beta_1}x, \beta_1\mathbb{I}\right)$.

($\star$) Change of variable with $g(u,v) := (\sqrt{\beta_1}u + \sqrt{1-\beta_1}v, v)$ such that $(Z_1, X) = g\left(\frac{1}{\sqrt{\beta_1}}\left(Z_1 - \sqrt{1-\beta_1}X\right), X\right)$

($\dagger$) Definition of Z_1

($\ddagger$) $\mathcal{E}_1 \sim \mathcal{N}(0, \mathbb{I})$

20.5

$$
\begin{aligned}
\operatorname{cov}[A+B] &= \mathbb{E}\left[(A+B)(A+B)^\top\right] - \mathbb{E}[A+B]\mathbb{E}[A+B]^\top \\
&= \mathbb{E}\left[AA^\top + AB^\top + BA^\top + BB^\top\right] \\
&\quad - \mathbb{E}[A+B]\mathbb{E}[A+B]^\top \\
&= \mathbb{E}\left[AA^\top\right] + \mathbb{E}\left[AB^\top\right] + \mathbb{E}\left[BA^\top\right] + \left[BB^\top\right] \\
&\quad - \left(\mathbb{E}[A] + \mathbb{E}[B]\right)\left(\mathbb{E}[A] + \mathbb{E}[B]\right)^\top \\
&= \mathbb{E}\left[AA^\top\right] + \mathbb{E}\left[AB^\top\right] + \mathbb{E}\left[BA^\top\right] + \left[BB^\top\right] \\
&\quad - \mathbb{E}[A]\mathbb{E}\left[A\right]^\top - \mathbb{E}[A]\mathbb{E}\left[B\right]^\top - \mathbb{E}[B]\mathbb{E}\left[A\right]^\top - \mathbb{E}[B]\mathbb{E}[B]^\top \\
&\overset{(\star)}{=} \mathbb{E}\left[AA^\top\right] \cancel{+\mathbb{E}[A]\mathbb{E}\left[B^\top\right]} \cancel{+\mathbb{E}[B]\mathbb{E}\left[A^\top\right]} + \mathbb{E}\left[BB^\top\right] \\
&\quad - \mathbb{E}[A]\mathbb{E}\left[A\right]^\top \cancel{-\mathbb{E}[A]\mathbb{E}\left[B\right]^\top} - \cancel{\mathbb{E}[B]\mathbb{E}\left[A\right]^\top} - \mathbb{E}[B]\mathbb{E}[B]^\top \\
&= \mathbb{E}\left[AA^\top\right] - \mathbb{E}[A]\mathbb{E}\left[A^\top\right] + \mathbb{E}\left[BB^\top\right] - \mathbb{E}[B]\mathbb{E}\left[B^\top\right] \\
&= \operatorname{cov}[A] + \operatorname{cov}[B]
\end{aligned}
$$

($\star$) $A \perp B$

$$
\begin{aligned}
\operatorname{cov}(\lambda A) &= \mathbb{E}\left[\lambda A(\lambda A)^\top\right] - \mathbb{E}\left[\lambda A\right]\mathbb{E}[\lambda A]^\top \\
&= \lambda^2\left(\mathbb{E}\left[AA^\top\right] - \mathbb{E}[A]\mathbb{E}[A]^\top\right) \\
&= \lambda^2 \operatorname{cov}(A)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[Z_t] &= \mathbb{E}\left[\sqrt{1-\beta_t} Z_{t-1} + \sqrt{\beta_t}\mathcal{E}_t\right] \\
&= \sqrt{1-\beta_t}\mathbb{E}[Z_{t-1}] + \sqrt{\beta_t}\mathbb{E}[\mathcal{E}_t] \\
&= \sqrt{1-\beta_t} \cdot 0 + \sqrt{\beta_t} \cdot 0 \qquad \mathcal{E}_t \sim \mathcal{N}(0, \mathbb{I}) \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{cov}[Z_t] &= \operatorname{cov}\left[\sqrt{1-\beta_t} Z_{t-1} + \sqrt{\beta_t}\mathcal{E}_t\right] \\
&\overset{(\dagger)}{=} \left(\sqrt{1-\beta_t}\right)^2 \operatorname{cov}\left[Z_{t-1}\right] + \left(\sqrt{\beta_t}\right)^2 \operatorname{cov}\left[\mathcal{E}_t\right] \\
&= (1-\beta_t)\,\mathbb{I} + \beta_t\mathbb{I} \qquad \mathcal{E}_t \sim \mathcal{N}(0, \mathbb{I}) \\
&= \mathbb{I} - \beta_t\mathbb{I} + \beta_t\mathbb{I} \\
&= \mathbb{I}
\end{aligned}
$$

(†) properties of cov as shown above