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

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

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

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

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

Covariance

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