

My solutions to
Deep Learning: Foundations and Concepts

Dario Miro Konopatzki

20 Diffusion Models

20.1

Mean

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

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

Auxiliary Calculations

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

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

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

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

Covariance

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

20.2

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

$$\begin{aligned}
q_{Z_1|X=x}(z_1) & = \frac{q_{Z_1,X}(z_1, x)}{q_X(x)} \quad \text{def. of conditional density} \\
& \stackrel{(\star)}{=} \frac{q_{\mathcal{E}_1,X} \left(\frac{1}{\sqrt{\beta_1}} (z_1 - \sqrt{1 - \beta_1} x), x \right)}{q_X(x)} \cdot \left| \det \begin{pmatrix} \frac{1}{\sqrt{\beta_1}} \mathbb{I}_D & -\frac{\sqrt{1 - \beta_1}}{\sqrt{\beta_1}} \mathbb{I}_D \\ 0 & \mathbb{I}_D \end{pmatrix} \right| \\
& = \frac{1}{\sqrt{\beta_1^D}} \frac{q_{\mathcal{E}_1,X} \left(\frac{1}{\sqrt{\beta_1}} (z_1 - \sqrt{1 - \beta_1} x), x \right)}{q_X(x)} \\
& = \frac{1}{\sqrt{\beta_1^D}} \frac{q_{\mathcal{E}_1} \left(\frac{1}{\sqrt{\beta_1}} (z_1 - \sqrt{1 - \beta_1} x) \right) \cancel{q_X(x)}}{\cancel{q_X(x)}} \quad \mathcal{E}_1 \perp X \\
& \stackrel{\dagger}{=} \frac{1}{\sqrt{\beta_1^D}} \frac{1}{\sqrt{(2\pi)^D \mathbb{I}_D}}
\end{aligned}$$

$$\begin{aligned}
& \cdot e^{-\frac{1}{2} \left(\frac{1}{\sqrt{\beta_1}} (z_1 - \sqrt{1-\beta_1}x) - 0 \right)^\top \mathbb{I}_D^{-1} \left(\frac{1}{\sqrt{\beta_1}} (z_1 - \sqrt{1-\beta_1}x) - 0 \right)} \\
&= \frac{1}{\sqrt{(2\pi)^D \beta_1^D \mathbb{I}_D}} e^{-\frac{1}{2} (z_1 - \sqrt{1-\beta_1}x)^\top \frac{1}{\beta_1} \mathbb{I}_D (z_1 - \sqrt{1-\beta_1}x)} \\
&= \frac{1}{\sqrt{(2\pi)^D \det(\beta_1 \mathbb{I}_D)}} e^{-\frac{1}{2} (z_1 - \sqrt{1-\beta_1}x)^\top (\beta_1 \mathbb{I}_D)^{-1} (z_1 - \sqrt{1-\beta_1}x)}
\end{aligned}$$

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

(\star) Change of variable with $g(u, v) := (\sqrt{\beta_1}u + \sqrt{1-\beta_1}v, v)$

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

20.3

(\star) Note that if $X \sim \mathcal{N}(\mu, \Sigma)$, $\text{im}(X) \subseteq \mathbb{R}^D$, $a \in \mathbb{R} \setminus \{0\}$, $b \in \mathbb{R}^D$, then

$$\begin{aligned}
p_{aX+b}(y) &= p_X \left(\frac{1}{a}(y-b) \right) \left| \det \left(\frac{1}{a} \mathbb{I}_D \right) \right| \\
&= \frac{1}{\sqrt{(2\pi)^D \det(\Sigma)}} e^{-\frac{1}{2} \left(\frac{1}{a}(y-b) - \mu \right)^\top \Sigma^{-1} \left(\frac{1}{a}(y-b) - \mu \right)} \frac{1}{|a^D|} \\
&= \frac{1}{\sqrt{(2\pi)^D \det(\Sigma) (a^D)^2}} e^{-\frac{1}{2} (y - (a\mu+b))^\top \left(\frac{1}{a^2} \Sigma^{-1} \right) (y - (a\mu+b))} \\
&= \frac{1}{\sqrt{(2\pi)^D \det(\Sigma) \det(a^2 \mathbb{I}_D)}} e^{-\frac{1}{2} (y - (a\mu+b))^\top (a^2 \Sigma)^{-1} (y - (a\mu+b))} \\
&= \frac{1}{\sqrt{(2\pi)^D \det(a^2 \Sigma)}} e^{-\frac{1}{2} (y - (a\mu+b))^\top (a^2 \Sigma)^{-1} (y - (a\mu+b))}
\end{aligned}$$

is density for $\mathcal{N}(a\mu + b, a^2 \Sigma)$.

Induction: $Z_t = \sqrt{\alpha_t}X + \tilde{\mathcal{E}}_t$ with $\tilde{\mathcal{E}}_t \sim \mathcal{N}(0, (1 - \alpha_t)\mathbb{I}_D)$ for all t .

$t = 1$

$$\begin{aligned}
Z_1 &= \sqrt{1-\beta_1}X + \sqrt{\beta_1}\mathcal{E}_1 && \text{by def.} \\
&= \sqrt{\alpha_1}X + \underbrace{\sqrt{1-\alpha_1}\mathcal{E}_1}_{=:\tilde{\mathcal{E}}_1} && \text{def. of } \alpha_1
\end{aligned}$$

where $\tilde{\mathcal{E}}_1 \sim \mathcal{N}(0, (1 - \alpha_1)\mathbb{I}_D)$ holds via (\star) since $\mathcal{E}_1 \sim \mathcal{N}(0, \mathbb{I}_D)$ by assumption.

$t \rightarrow t + 1$

$$\begin{aligned}
Z_{t+1} &= \sqrt{1 - \beta_{t+1}} Z_t + \sqrt{\beta_{t+1}} \mathcal{E}_{t+1} && \text{by def.} \\
&= \sqrt{1 - \beta_{t+1}} \left(\sqrt{\alpha_t} X + \tilde{\mathcal{E}}_t \right) + \sqrt{\beta_{t+1}} \mathcal{E}_{t+1} && \text{ind. hypothesis} \\
&= \sqrt{(1 - \beta_{t+1}) \alpha_t} X + \sqrt{1 - \beta_{t+1}} \tilde{\mathcal{E}}_t + \sqrt{\beta_{t+1}} \mathcal{E}_{t+1} \\
&= \sqrt{\alpha_{t+1}} X + \underbrace{\sqrt{1 - \beta_{t+1}} \tilde{\mathcal{E}}_t + \sqrt{\beta_{t+1}} \mathcal{E}_{t+1}}_{:= \tilde{\mathcal{E}}_{t+1}} && \text{def. of } \alpha_{t+1}
\end{aligned}$$

where $\mathcal{E}_{t+1} \sim \mathcal{N}(0, \mathbb{I}_D)$ by assumption and $\tilde{\mathcal{E}}_t \sim \mathcal{N}(0, (1 - \alpha_t) \mathbb{I}_D)$ by induction hypothesis. Hence via (\star) with (3.212) it holds that

$$\begin{aligned}
\tilde{\mathcal{E}}_{t+1} &\sim \mathcal{N}(0, (1 - \beta_{t+1})(1 - \alpha_t) \mathbb{I}_D + \beta_{t+1} \mathbb{I}_D) \\
&= \mathcal{N}(0, (1 - \beta_{t+1} - \alpha_{t+1}) \mathbb{I}_D + \beta_{t+1} \mathbb{I}_D) \\
&= \mathcal{N}(0, (1 - \alpha_{t+1}) \mathbb{I}_D)
\end{aligned}$$

Conditional density

$$\begin{aligned}
q_{Z_t|X=x}(z_t) &= \frac{q_{\sqrt{\alpha_t} X + \tilde{\mathcal{E}}_t, X}(z_t, x)}{q_X(x)} \\
&\stackrel{(\dagger)}{=} \frac{q_{\tilde{\mathcal{E}}_t, X}(z_t - \sqrt{\alpha_t} x, x)}{p_X(x)} \cdot \underbrace{\left| \det \begin{pmatrix} \mathbb{I}_D & -\sqrt{\alpha_t} \\ 0 & \mathbb{I}_D \end{pmatrix} \right|}_{=1} \\
&\stackrel{(\ddagger)}{=} \frac{q_{\tilde{\mathcal{E}}_t}(z_t - \sqrt{\alpha_t} x) \cancel{q_X(x)}}{\cancel{q_X(x)}} \\
&= \frac{1}{\sqrt{(2\pi)^D \det((1 - \alpha_t) \mathbb{I}_D)}} e^{-\frac{1}{2}(z_t - \sqrt{\alpha_t} x)^\top ((1 - \alpha_t) \mathbb{I}_D)^{-1} (z_t - \sqrt{\alpha_t} x)}
\end{aligned}$$

is density for $\mathcal{N}(\sqrt{\alpha_t} x, (1 - \alpha_t) \mathbb{I}_D)$.

(\dagger) Change of variable with $g(u, v) := (\sqrt{\alpha_t} v + u, v)$

(\ddagger) $(\mathcal{E}_t)_t$ is assumed to be ‘independent noise’, so X independent of $(\mathcal{E}_\tau)_{1 \leq t}$ for all t . It follows that X also independent of measurable function $\tilde{\mathcal{E}}_t$ of $(\mathcal{E}_\tau)_{\tau \leq t}$.

20.4

By assumption, $0 < \beta_t < 1$ and thus $0 < (1 - \beta_t) < 1$ for all t . Consequently $\alpha_t = \prod_{\tau=1}^t (1 - \beta_\tau) \xrightarrow{t \rightarrow \infty} 0$ and it follows that

$$\begin{aligned} q_{Z_t|X=x}(z) &= \frac{1}{\sqrt{(2\pi)^D \det((1 - \alpha_t)\mathbb{I}_D)}} e^{-\frac{1}{2}(z_t - \sqrt{\alpha_t}x)^\top ((1 - \alpha_t)\mathbb{I}_D)^{-1} (z_t - \sqrt{\alpha_t}x)} \\ &= \frac{1}{\sqrt{(2\pi)^D \underbrace{(1 - \alpha_t)^D}_{\rightarrow 1}}} e^{-\frac{1}{2} \underbrace{\frac{1}{1 - \alpha_t}}_{\rightarrow 1} \underbrace{(z_t - \sqrt{\alpha_t}x)^\top}_{\rightarrow 0} \underbrace{(z_t - \sqrt{\alpha_t}x)}_{\rightarrow 0}} \\ &\xrightarrow{t \rightarrow \infty} \frac{1}{\sqrt{(2\pi)^D}} e^{-\frac{1}{2} z_t^\top z_t} \end{aligned}$$

which is density for $\mathcal{N}(0, \mathbb{I}_D)$.

20.5

$$\begin{aligned} \text{cov}[A + B] &= \mathbb{E}[(A + B)(A + B)^\top] - \mathbb{E}[A + B]\mathbb{E}[A + B]^\top \\ &= \mathbb{E}[AA^\top + AB^\top + BA^\top + BB^\top] \\ &\quad - \mathbb{E}[A + B]\mathbb{E}[A + B]^\top \\ &= \mathbb{E}[AA^\top] + \mathbb{E}[AB^\top] + \mathbb{E}[BA^\top] + \mathbb{E}[BB^\top] \\ &\quad - (\mathbb{E}[A] + \mathbb{E}[B])(\mathbb{E}[A] + \mathbb{E}[B])^\top \\ &= \mathbb{E}[AA^\top] + \mathbb{E}[AB^\top] + \mathbb{E}[BA^\top] + \mathbb{E}[BB^\top] \\ &\quad - \mathbb{E}[A]\mathbb{E}[A]^\top - \mathbb{E}[A]\mathbb{E}[B]^\top - \mathbb{E}[B]\mathbb{E}[A]^\top - \mathbb{E}[B]\mathbb{E}[B]^\top \\ &\stackrel{(*)}{=} \mathbb{E}[AA^\top] + \cancel{\mathbb{E}[A]\mathbb{E}[B^\top]} + \cancel{\mathbb{E}[B]\mathbb{E}[A^\top]} + \mathbb{E}[BB^\top] \\ &\quad - \mathbb{E}[A]\mathbb{E}[A]^\top - \cancel{\mathbb{E}[A]\mathbb{E}[B]^\top} - \cancel{\mathbb{E}[B]\mathbb{E}[A]^\top} - \mathbb{E}[B]\mathbb{E}[B]^\top \\ &= \mathbb{E}[AA^\top] - \mathbb{E}[A]\mathbb{E}[A]^\top + \mathbb{E}[BB^\top] - \mathbb{E}[B]\mathbb{E}[B]^\top \\ &= \text{cov}[A] + \text{cov}[B] \end{aligned}$$

(*) $A \perp B$

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

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

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

(\dagger) properties of cov as shown above