Unifying Diffusion Identities via Tweedie's Formula

Jiayi Lin[1,†] **Thang D. Bui**[1,†]

[1]Australian National University, Acton, ACT, Australia
[†]Correspondence to jiayi.lin@anu.edu.au and thang.bui@anu.edu.au

Abstract

This paper consolidates several standard diffusion-model identities under a single Tweedie-style normaliser-differentiation viewpoint, making existing formulas easier to compare, reuse, and extend. Under Gaussian forward corruption, posterior moments such as means and covariances can be expressed in terms of derivatives of the log-marginal density, yielding unified expressions for intermediate-time conditionals, Gaussian bridge moments, and inverse-problem guidance terms. The resulting presentation is intended as a concise reference for theoretical derivations and implementation checks in diffusion modelling.

1 Introduction

Diffusion-based generative models have achieved strong empirical performance in many domains (Ho et al., 2020; Song et al., 2021b,a). These models are characterised by a judiciously chosen forward process $\{X_t\}_{t\in[0,1]}$ that progressively corrupts data samples $X_0 \sim p_0$ into a simple reference distribution p_1, together with a reverse-time generative process that transforms noise back into data. When the forward process is Markov, the initial law p_0 and transition kernels $\{p_{t|s}\}_{0\le s<t\le 1}$ determine a joint path law and hence a marginal density path $\{p_t\}_{t\in[0,1]}$ connecting p_0 and p_1. Many diffusion methods construct a reverse process that approximates the time-reversal of this forward Markov law.[1]

Although many objectives and sampling identities in diffusion models are closely related, they are often derived through different lenses such as score matching (Song et al., 2021b; Song and Ermon, 2019), time-reversal arguments (Song et al., 2021b; Albergo et al., 2025), the variational lower bound perspective (Ho et al., 2020; Nichol and Dhariwal, 2021; Kingma et al., 2021; Kingma and Gao, 2023), or forward-reverse joint KL minimisation (Bao et al., 2022b,a; Ou et al., 2025). These different perspectives are often expressed in different notations and, as a result, can obscure which quantities are genuinely distinct and which are merely reparameterisations of the same underlying identity.

To this end, this note aims to unify several diffusion identities via a Tweedie-style normaliser-differentiation viewpoint. Concretely, we show how posterior moments under a Gaussian forward corruption can be expressed in terms of derivatives of the log-marginal density, and for $0 \le s < t \le 1$ how this yields (i) the usual Tweedie posterior mean and covariance identities, (ii) intermediate-time identities such as $X_s \mid X_t$, (iii) Gaussian bridge moments $X_s \mid X_t, X_0$, and (iv) conditional and inverse-problem guidance identities. The exposition is organised using (1) a single exponential-family derivation template for Tweedie's identities, (2) a consistent mapping between identities at times 0,

1. For generation it is often sufficient to match the marginal path (or at least the endpoint p_0) rather than the full trajectory law (Holderrieth et al., 2025; Albergo et al., 2025).

s, and t, and (3) a bridge to conditional/inverse-problem guidance formulae. We first describe the notations and set-up in section 2, and then develop the aforementioned identities in section 3.

2 Set-up

Gaussian forward corruption. Let $X_0 \in \mathbb{R}^d$ denote a sample from a data distribution $p_0(x)$. We consider a family of Gaussian forward corruption kernels $\{p_{t|0}\}_{0<t\leq 1}$ indexed by $t \in (0, 1]$:

$$p_{t|0}(x_t \mid x_0) = \mathcal{N}\big(x_t; \alpha_t x_0, \beta_t^2 I_d\big), \qquad \alpha_t > 0,\ \beta_t \geq 0,\ \alpha_0 = 1,\ \beta_0 = 0, \tag{1}$$

or equivalently, $X_t = \alpha_t X_0 + \beta_t \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, I_d)$. For schedules with $\alpha_t \to 0$ and $\beta_t \to 1$ as $t \to 1$, the terminal marginal p_1 is approximately standard Gaussian, $p_1(x_1) \approx \mathcal{N}(x_1; 0, I_d)$. We index noise levels by $t \in [0, 1]$ and evaluate all quantities on a discrete grid $0 = t_0 < t_1 < \cdots < t_N = 1$. The identities in this note are discrete-time (applied at grid points), while the notation follows the continuous-time diffusion literature.

Forward Markov transitions and marginals. We assume the family (1) is Markov-consistent in forward time, i.e., for any $0 \leq s < t \leq 1$ there exists a kernel $p_{t|s}(x_t \mid x_s)$ such that $p_{t|0}(x_t \mid x_0) = \int p_{t|s}(x_t \mid x_s)\, p_{s|0}(x_s \mid x_0)\, dx_s$. For the linear Gaussian family (1), this holds with Gaussian transitions

$$p_{t|s}(x_t \mid x_s) = \mathcal{N}\big(x_t; \alpha_{t|s} x_s, \beta_{t|s}^2 I_d\big), \qquad \alpha_{t|s} := \frac{\alpha_t}{\alpha_s}, \quad \beta_{t|s}^2 := \beta_t^2 - \alpha_{t|s}^2 \beta_s^2, \tag{2}$$

and we assume the schedule of α_t and β_t satisfies $\beta_{t|s}^2 \geq 0$ for all $s < t$. Equivalently, one may realise the process via independent Gaussian increments $X_t = \alpha_{t|s} X_s + \beta_{t|s} \varepsilon_{t|s}$ with $\varepsilon_{t|s} \sim \mathcal{N}(0, I_d)$.

The marginal at time t is $p_t(x_t) = \int p_{t|0}(x_t \mid x_0)\, p_0(x_0)\, dx_0$. We use ∇_x and ∇_x^2 to denote the gradient and Hessian with respect to x.

3 Tweedie's identities

3.1 Tweedie's identities in the linear-Gaussian model

Let $p(y \mid x) = \mathcal{N}(y; Ax, C)$ be a linear Gaussian likelihood for $x \in \mathbb{R}^d, y \in \mathbb{R}^m, A \in \mathbb{R}^{m\times d}, C \in \mathbb{R}^{m\times m}$, and $p(x)$ be an arbitrary prior. By directly differentiating the log-marginal likelihood $p(y) = \int p(y \mid x) p(x)\, dx$, we can show that the marginal score and Hessian satisfy

$$\nabla_y \log p(y) = C^{-1}\big(A\,\mathbb{E}[X \mid Y = y] - y\big), \quad \nabla_y^2 \log p(y) = C^{-1} A \operatorname{Cov}[X \mid Y = y]\, A^\top C^{-1} - C^{-1}. \tag{3}$$

Equations (3) are the prototype of the identities we use throughout the paper: posterior moments appear as derivatives of the log marginal density. We will now re-express the same idea in a form that applies beyond the linear-Gaussian model using a single role-reversal or normaliser-differentiation viewpoint and then tailor the result to diffusion forward kernels.

3.2 Role reversal and normaliser differentiation

A key idea in Tweedie-style arguments is that, in many latent-variable models, the likelihood can be written so that the observation y enters only through a finite-dimensional feature map, and this feature appears linearly in the exponent (Stein, 1981; Efron, 2011). This enables a "role reversal": rather than treating y as the random variable and x as a parameter, we fix y and view the posterior over x as a member of a family indexed by y. Posterior moments then follow by differentiating the corresponding normaliser, which can be expressed via the marginal density.

Assume the likelihood can be written as

$$p(y \mid x) \;=\; \rho(\phi(y))\, g(\psi(x))\, \exp\{\phi(y)^\top \psi(x)\}, \tag{4}$$

where $\phi : \mathcal{Y} \to \mathbb{R}^k$ and $\psi : \mathcal{X} \to \mathbb{R}^k$ are vector-valued maps and $\rho, g > 0$. Let $m(x) \geq 0$ be any base measure (e.g., a prior $p(x)$ or a joint factor in a larger model). The marginal density of y is $p(y) = \int m(x)\, p(y \mid x)\, dx \;=\; Z(\phi(y))$ where we define the normaliser $Z(\eta) := \rho(\eta)\, \tilde{Z}(\eta)$ and $\tilde{Z}(\eta) := \int m(x)\, g(\psi(x))\, \exp\{\eta^\top \psi(x)\}\, dx$.

Moments from differentiating the normaliser. Assuming differentiation under the integral sign is valid, we have

$$\nabla_\eta \log \tilde{Z}(\eta) = \mathbb{E}[\psi(X) \mid \eta], \quad \nabla_\eta^2 \log \tilde{Z}(\eta) = \mathrm{Cov}[\psi(X) \mid \eta], \tag{5}$$

where $\mathbb{E}[\cdot \mid \eta]$ and $\mathrm{Cov}[\cdot \mid \eta]$ denote conditional expectation and covariance under the density $p(x \mid \eta) \propto m(x) g(\psi(x)) \exp\{\eta^\top \psi(x)\}$. In particular, evaluating at $\eta = \phi(y)$ yields

$$\mathbb{E}[\psi(X) \mid Y = y] = \nabla_\eta \log \tilde{Z}(\eta)\big|_{\eta=\phi(y)}, \quad \mathrm{Cov}[\psi(X) \mid Y = y] = \nabla_\eta^2 \log \tilde{Z}(\eta)\big|_{\eta=\phi(y)}. \tag{6}$$

From η-derivatives to y-derivatives. Since $\log p(y) = \log Z(\phi(y)) = \log \rho(\phi(y)) + \log \tilde{Z}(\phi(y))$, the chain rule gives

$$\nabla_y \log p(y) = J_\phi(y)^\top \nabla_\eta \log \tilde{Z}(\eta)\big|_{\eta=\phi(y)} + J_\phi(y)^\top \nabla_\eta \log \rho(\eta)\big|_{\eta=\phi(y)}, \tag{7}$$

and similarly for the Hessian. In the linear-Gaussian case (and in diffusion forward kernels), $\phi(y)$ is affine in y, thus reducing the right-hand side of (7) to simple linear transforms of posterior moments.

The identities (3) for the linear Gaussian case in Section 3.1 can also be obtained by expressing the Gaussian likelihood in the pairing form (4) and differentiating the associated normaliser $\tilde{Z}(\phi(y))$. We present this alternative derivation in Appendix B.2. The same role-reversal/normaliser-differentiation viewpoint extends directly to diffusion kernels by treating the noisy state as the "observation" and X_0 (or X_s) as the latent variable.

3.3 Deriving Tweedie's identities for diffusion kernels

We use vector-form first- and second-order sufficient statistics. For a scalar $a > 0$, let $\rho_a(x) := \mathcal{N}(x; 0, a^2 I_d)$ and define vector-valued maps $\phi_a : \mathbb{R}^d \to \mathbb{R}^{d+d^2}$ and $\psi_a : \mathbb{R}^d \to \mathbb{R}^{d+d^2}$ by

$$\phi_a(x)^\top = \begin{bmatrix} a^{-2} x^\top & -\tfrac{1}{2} a^{-2} \mathrm{vec}(I_d)^\top \end{bmatrix}, \qquad \psi_a(x)^\top = \begin{bmatrix} (ax)^\top & \big(a^2\, \mathrm{vec}(xx^\top)\big)^\top \end{bmatrix}. \tag{8}$$

Gaussian forward kernels in $\exp\{\phi^\top \psi\}$ form. Recall the forward kernel $p_{t|0}(x_t \mid x_0) = \mathcal{N}(x_t; \alpha_t x_0, \beta_t^2 I_d)$. A direct expansion of the Gaussian quadratic shows it can be written as

$$p_{t|0}(x_t \mid x_0) = \rho_{\beta_t}(x_t)\, \exp\Big\{\phi_{\beta_t}(x_t)^\top \psi_{\alpha_t}(x_0)\Big\}, \tag{9}$$

where we have used ρ_{β_t} to collect all terms depending only on x_t, and the pairing $\phi_{\beta_t}(x_t)^\top \psi_{\alpha_t}(x_0)$ to collect the cross- and x_0-quadratic terms. With prior $p_0(x_0)$, the induced marginal is $p_t(x_t) = \rho_{\beta_t}(x_t) \int p_0(x_0) \exp\Big\{\phi_{\beta_t}(x_t)^\top \psi_{\alpha_t}(x_0)\Big\}\, dx_0 \;\; =: \;\; Z_t\big(\phi_{\beta_t}(x_t)\big)$. Define the associated log-normaliser $A_t(x_t) := \log \frac{p_t(x_t)}{\rho_{\beta_t}(x_t)} = \log Z_t\big(\phi_{\beta_t}(x_t)\big) - \log \rho_{\beta_t}(x_t)$.

Posterior as a role-reversal exponential-family form. By Bayes' rule,

$$p_{0|t}(x_0 \mid x_t) = \frac{p_0(x_0) p_{t|0}(x_t \mid x_0)}{p_t(x_t)} = p_0(x_0) \exp\Big\{\phi_{\beta_t}(x_t)^\top \psi_{\alpha_t}(x_0) - A_t(x_t)\Big\}, \tag{10}$$

so the base measure is $p_0(x_0)$, the canonical parameter is $\phi_{\beta_t}(x_t)$, and the sufficient statistic is $\psi_{\alpha_t}(x_0)$.

Table 1. Instances of the normaliser-differentiation template under Gaussian diffusion kernels. Each row identifies the base measure m, the pairing $\exp\{\phi^\top \psi\}$ induced by a Gaussian forward kernel, and the associated log-normaliser A, which can be expressed via a marginal density. Posterior moments follow from derivatives of A with respect to x_t (via $\phi(x_t)$).

Target posterior	Forward kernel	Base measure $m(\cdot)$	$\phi(x_t)$	$\psi(\cdot)$	Log-norm. $A_t(\phi)$	Eq.
$p_{0\|t}(x_0 \mid x_t)$	$p_{t\|0}(x_t \mid x_0)$	$p_0(x_0)$	$\phi_{\beta_t}(x_t)$	$\psi_{\alpha_t}(x_0)$	$\log \frac{p_t(x_t)}{\rho_{\beta_t}(x_t)}$	(10), (11)
$p_{s\|t}(x_s \mid x_t)$	$p_{t\|s}(x_t \mid x_s)$	$p_s(x_s)$	$\phi_{\beta_{t\|s}}(x_t)$	$\psi_{\alpha_{t\|s}}(x_s)$	$\log \frac{p_t(x_t)}{\rho_{\beta_{t\|s}}(x_t)}$	(12), (13)
$p_{s\|0,t}(x_s \mid x_t, x_0)$	$p_{t\|s}(x_t \mid x_s)$	$p_{s\|0}(x_s \mid x_0)$	$\phi_{\beta_{t\|s}}(x_t)$	$\psi_{\alpha_{t\|s}}(x_s)$	$\log \frac{p_{t\|0}(x_t\|x_0)}{\rho_{\beta_{t\|s}}(x_t)}$	(15), (16)
$p_{0\|t}(x_0 \mid x_t, y)$	$p_{t\|0}(x_t \mid x_0)$	$p_0(x_0)\, p(y \mid x_0)$	$\phi_{\beta_t}(x_t)$	$\psi_{\alpha_t}(x_0)$	$\log \frac{p_t(x_t,y)}{\rho_{\beta_t}(x_t)}$	(19), (20)
$q_{0\|t}(x_0 \mid x_t)$	$p_{t\|0}(x_t \mid x_0)$	$R(x_0)p_0(x_0)$	$\phi_{\beta_t}(x_t)$	$\psi_{\alpha_t}(x_0)$	$\log \frac{p_t(x_t)\, r_t(x_t)}{\rho_{\beta_t}(x_t)}$	(23), (24)

Tweedie's identities for $X_0 \mid x_t$. Differentiating the log-normaliser $A_t(x_t)$ (see Appendix B.2) yields the posterior mean and covariance:

$$\mathbb{E}[X_0 \mid x_t] = \frac{1}{\alpha_t}\Big(x_t + \beta_t^2 \nabla_{x_t} \log p_t(x_t)\Big), \quad \mathrm{Cov}[X_0 \mid x_t] = \frac{\beta_t^2}{\alpha_t^2}\Big(I_d + \beta_t^2 \nabla_{x_t}^2 \log p_t(x_t)\Big). \tag{11}$$

This normaliser-differentiation viewpoint makes it explicit that the prior enters only through the base measure p_0, while posterior moments depend only on derivatives of the induced log-marginal density.

Related ideas appear in work on estimating higher-order posterior moments (Meng et al., 2021; Boys et al., 2024) and in extensions beyond Gaussian corruption (Kim and Ye, 2021). Our focus here is instead organisational: once a family of Gaussian Markov corruption kernels $\{p_{t|s}\}_{0 \le s < t \le 1}$ is fixed, the same reasoning can be reused to (i) connect identities across intermediate times and (ii) track how changing the base measure (e.g. conditioning/guidance or prior tilting) modifies the corresponding posterior moments.

Identity (11) characterises denoising at a single noise level t, but diffusion sampling is inherently multi-step: to propagate a state from time t to an earlier time $s < t$, a sampler needs the reverse-time conditional $p(x_s \mid x_t)$, typically approximated by a Gaussian transition whose mean and covariance depend on moments of $X_s \mid X_t$. The next subsection derives these intermediate-time moments and shows how they relate back to $X_0 \mid X_t$. Table 1 summarises the main instances of the template.

3.4 Tweedie's Formula in Diffusion Models

3.4.1 Connection between $X_0 \mid x_t$ and $X_s \mid x_t$

For $p_{t|s}(x_t \mid x_s) = \mathcal{N}(x_t; \alpha_{t|s} x_s, \beta_{t|s}^2 I_d)$ with base measure $p_s(x_s)$, the same normaliser-differentiation argument gives

$$p_{s|t}(x_s \mid x_t) = \frac{p_s(x_s) p_{t|s}(x_t \mid x_s)}{p_t(x_t)} = p_s(x_s) \exp\left\{\phi_{\beta_{t|s}}(x_t)^\top \psi_{\alpha_{t|s}}(x_s) - \log \frac{p_t(x_t)}{\rho_{\beta_{t|s}}(x_t)}\right\}, \tag{12}$$

and hence

$$\mathbb{E}[X_s \mid x_t] = \frac{1}{\alpha_{t|s}}\Big(x_t + \beta_{t|s}^2 \nabla_{x_t} \log p_t(x_t)\Big), \quad \mathrm{Cov}[X_s \mid x_t] = \frac{\beta_{t|s}^2}{\alpha_{t|s}^2}\Big(I_d + \beta_{t|s}^2 \nabla_{x_t}^2 \log p_t(x_t)\Big). \tag{13}$$

Because the forward kernels are Markov-consistent $p_t(x_t) = \int p_{t|s}(x_t \mid x_s) p_s(x_s)\, dx_s = \int p_{t|0}(x_t \mid x_0) p_0(x_0)\, dx_0$. Therefore the same marginal p_t underlies both $X_0 \mid X_t$ and $X_s \mid X_t$, yielding

$$\mathbb{E}[X_s \mid x_t] = \frac{\alpha_s \beta_{t|s}^2}{\beta_t^2}\mathbb{E}[X_0 \mid x_t] + \frac{\alpha_t \beta_s^2}{\alpha_s \beta_t^2} x_t, \quad \mathrm{Cov}[X_s \mid x_t] = \frac{\alpha_s^2 \beta_{t|s}^4}{\beta_t^4}\mathrm{Cov}[X_0 \mid x_t] + \frac{\beta_{t|s}^2 \beta_s^2}{\beta_t^2} I_d. \tag{14}$$

Thus, the linear relationships between the moments of $X_s \mid x_t$ and $X_0 \mid x_t$ show that a model trained to estimate the denoising quantities at time t (e.g., the posterior mean $\mathbb{E}[X_0 \mid x_t]$) implicitly determines the mean of intermediate-time conditionals $\mathbb{E}[X_s \mid x_t]$ for any $0 \le s < t \le 1$ under a fixed Markov-consistent Gaussian forward schedule. When the reverse transition $q_\theta(x_s \mid x_t)$ is modelled as Gaussian, these identities provide the corresponding moment targets used by samplers and by discrete-time KL objectives (Bao et al., 2022b,a; Ou et al., 2025).

3.4.2 Connection between $X_s \mid x_t$ and $X_s \mid x_t, x_0$

The same argument gives the moments of the Gaussian bridge distribution:

$$p_{s|0,t}(x_s \mid x_t, x_0) = \frac{p_{t|s}(x_t \mid x_s) p_{s|0}(x_s \mid x_0)}{p_{t|0}(x_t \mid x_0)} = p_{s|0}(x_s \mid x_0) \exp\left\{ \phi_{\beta_{t|s}}(x_t)^\top \psi_{\alpha_{t|s}}(x_s) - \log \frac{p_{t|0}(x_t \mid x_0)}{\rho_{\beta_{t|s}}(x_t)} \right\}, \tag{15}$$

$$\mathbb{E}[X_s \mid x_t, x_0] = \alpha_s \frac{\beta_{t|s}^2}{\beta_t^2} x_0 + \frac{\alpha_t \beta_s^2}{\alpha_s \beta_t^2} x_t, \quad \mathrm{Cov}[X_s \mid x_t, x_0] = \frac{\beta_{t|s}^2 \beta_s^2}{\beta_t^2} I_d. \tag{16}$$

Comparing (13) and (16), the variance decomposes as

$$\mathrm{Cov}[X_s \mid x_t] = \frac{\alpha_s^2 \beta_{t|s}^4}{\beta_t^4} \mathrm{Cov}[X_0 \mid x_t] + \mathrm{Cov}[X_s \mid x_t, x_0]. \tag{17}$$

In the original DDPM approximation (Ho et al., 2020), the reverse covariance commonly uses $\mathrm{Cov}[X_s \mid x_t, x_0]$. The decomposition above shows this approach ignores the additional uncertainty term tied to $\mathrm{Cov}[X_0 \mid x_t]$ (equivalently, it assumes that this term is negligible). Recent work considers data-dependent reverse variances estimated from scores (Bao et al., 2022b,a; Ou et al., 2025).

3.5 Tweedie's Formula in Conditional Generation

In many applications, we do not only want unconditional samples $X_0 \sim p_0$, but also samples from a conditional distribution $p(x_0 \mid y)$, where y encodes side information or measurements. This includes classical inverse problems (Daras et al., 2024; Chung et al., 2023; Boys et al., 2024) as well as general conditional generation (Dhariwal and Nichol, 2021; Ho and Salimans, 2021; Ho et al., 2022). A common probabilistic formulation treats $p_0(x_0)$ as a diffusion prior and introduces an observation model $p(y \mid x_0)$, so that $p(x_0 \mid y) \propto p_0(x_0)\, p(y \mid x_0)$.

Suppose we want to construct a reverse-time process whose endpoint distribution matches $X_0 \mid y \sim p(x_0 \mid y)$. Assume the standard conditional independence $Y \perp X_t \mid X_0$, for all $t \in (0, 1]$, i.e., observations depend on the clean variable but not directly on the intermediate noisy states given X_0. Then the posterior is $p(x_0 \mid y) = \frac{p(y|x_0)p_0(x_0)}{p(y)}$. Define the joint marginal $p_t(x_t, y) = \int p_0(x_0)\, p_{t|0}(x_t \mid x_0)\, p(y \mid x_0)\, dx_0$, and note that $p_t(x_t, y) = p_t(x_t \mid y)\, p(y)$, hence $\nabla_{x_t} \log p_t(x_t, y) = \nabla_{x_t} \log p_t(x_t \mid y)$. Also, define $p_t(y \mid x_t) := p_t(x_t, y)/p_t(x_t)$. Then the conditional score admits the standard decomposition

$$\nabla_{x_t} \log p_t(x_t \mid y) = \nabla_{x_t} \log p_t(x_t) + \nabla_{x_t} \log p_t(y \mid x_t), \tag{18}$$

where $\nabla_{x_t} \log p_t(y \mid x_t)$ is the guidance score (typically intractable without further structure).

Many diffusion-based conditioning methods can be viewed as approximating the conditional score $\nabla_{x_t} \log p_t(x_t \mid y)$: either by learning it directly (Batzolis et al., 2021; Ho and Salimans, 2021), or by combining an unconditional score model $\nabla_{x_t} \log p_t(x_t)$ with an additional estimate of the guidance score $\nabla_{x_t} \log p_t(y \mid x_t)$ via an explicit or approximate likelihood gradient (Boys et al., 2024), a classifier (Dhariwal and Nichol, 2021), or another surrogate (Denker et al., 2024). Using Bayes' rule and $Y \perp X_t \mid X_0$, we apply the same normaliser-differentiation argument as in the unconditional

case to obtain,

$$p_{0|t}(x_0 \mid x_t, y) = \frac{p_{t|0}(x_t \mid x_0)\, p(y \mid x_0)\, p_0(x_0)}{p_t(x_t, y)} = p_0(x_0) p(y \mid x_0) \exp\left\{ \phi_{\beta_t}(x_t)^\top \psi_{\alpha_t}(x_0) - \log \frac{p_t(x_t, y)}{\rho_{\beta_t}(x_t)} \right\}, \tag{19}$$

$$\mathbb{E}[X_0 \mid x_t, y] = \frac{1}{\alpha_t}\Big(x_t + \beta_t^2 \nabla_{x_t} \log p_t(x_t, y)\Big), \quad \mathrm{Cov}[X_0 \mid x_t, y] = \frac{\beta_t^2}{\alpha_t^2}\Big(I_d + \beta_t^2 \nabla_{x_t}^2 \log p_t(x_t, y)\Big). \tag{20}$$

3.6 Gluing other factors and changing priors

For a Gaussian forward kernel $p_{t|0}(x_t \mid x_0)$, the role-reversal viewpoint implies that we may multiply the base measure on x_0 by any nonnegative factor and still obtain posterior moments from derivatives of the corresponding marginal normaliser. Concretely, if we replace the prior $p_0(x_0)$ by an unnormalised target density $f(x_0)$ (known up to a constant), then the posterior is proportional to $p_{t|0}(x_t \mid x_0) f(x_0)$.

Let p_0 be a "reference" prior from which we can sample, and let f be an unnormalised target density with normalised version $q_0(x_0) = f(x_0)/Z$ where $Z = \int f(x)\, dx$. Assume q_0 is absolutely continuous w.r.t. p_0 and define the density ratio $R(x_0) := \frac{f(x_0)}{p_0(x_0)}$. Then $q_0(x_0) \propto R(x_0) p_0(x_0)$ and

$$q_{0|t}(x_0 \mid x_t) = \frac{p_{t|0}(x_t \mid x_0)\, q_0(x_0)}{q_t(x_t)} = \frac{p_{t|0}(x_t \mid x_0) R(x_0) p_0(x_0)}{p_t(x_t) \int p_{0|t}(x_0 \mid x_t) R(x_0)\, dx_0}. \tag{21}$$

Let $r_t(x_t) := \int p_{0|t}(x_0 \mid x_t) R(x_0)\, dx_0 = \mathbb{E}_{p_0}[R(X_0) \mid x_t]$. Then the induced marginal satisfies

$$q_t(x_t) = p_t(x_t)\, r_t(x_t), \qquad \text{and hence} \qquad \nabla_{x_t} \log q_t(x_t) - \nabla_{x_t} \log p_t(x_t) = \nabla_{x_t} \log r_t(x_t). \tag{22}$$

The quantity $r_t(x_t)$ is the same type of conditional normaliser that appears in Doob h-transform / tilted diffusion formulations; for example, Denker et al. (2025) use $R(x_0) = \exp(r(x_0)/\lambda)$ and obtain an analogous score decomposition for tilted target distributions.

Moreover, the posterior induced by q_0 admits the same $\exp\{\phi^\top \psi\}$ form:

$$q_{0|t}(x_0 \mid x_t) = R(x_0) p_0(x_0) \exp\left\{ \phi_{\beta_t}(x_t)^\top \psi_{\alpha_t}(x_0) - \log \frac{p_t(x_t) r_t(x_t)}{\rho_{\beta_t}(x_t)} \right\}. \tag{23}$$

Applying Tweedie's identities to p_0 and q_0 and subtracting yields

$$\mathbb{E}_{q_0}[X_0 \mid x_t] = \mathbb{E}_{p_0}[X_0 \mid x_t] + \frac{\beta_t^2}{\alpha_t} \nabla_{x_t} \log r_t(x_t), \quad \mathrm{Cov}_{q_0}[X_0 \mid x_t] = \mathrm{Cov}_{p_0}[X_0 \mid x_t] + \frac{\beta_t^4}{\alpha_t^2} \nabla_{x_t}^2 \log r_t(x_t). \tag{24}$$

Equations (24) show that changing the prior by a density ratio R shifts posterior moments through derivatives of $r_t(x_t) = \mathbb{E}_{p_0}[R(X_0) \mid x_t]$. In principle, the conditional mean $\mathbb{E}_{q_0}[X_0 \mid x_t]$ (equivalently the marginal score $\nabla_{x_t} \log q_t(x_t)$) defines a reverse process targeting q_0. When direct sampling from q_0 is unavailable, the Tweedie viewpoint reduces this to estimating $r_t(x_t)$, a reweighted conditional expectation under the reference model p_0.

4 Summary

This paper consolidates a collection of diffusion-model identities around a single Tweedie-style normaliser-differentiation viewpoint. Under Gaussian forward kernels, posterior moments of latent states (e.g. $X_0 \mid X_t$, $X_s \mid X_t$, and bridge distributions $X_s \mid X_t, X_0$) can be written as derivatives of a marginal normaliser, equivalently as derivatives of $\log p_t(x)$. The resulting "dictionary" makes explicit how several formulas that are often derived separately are different instances of the same underlying identities.

We also applied the same viewpoint to conditional generation and prior tilting. Under a fixed Markov Gaussian forward process, "gluing" additional factors into the base measure (e.g. likelihood terms for conditioning, or density ratios for tilting) induces systematic shifts in posterior moments through derivatives of the corresponding conditional normalisers. In turn, this reveals how reverse-transition moment approximations (mean and covariance) should be adjusted when targeting a modified prior or a conditional distribution.

Limitations. In this work, we focus on isotropic Gaussian corruption kernels and Markov-consistent Gaussian transitions. While the normaliser-differentiation template extends to broader exponential-family corruptions, carrying through the same moment identities beyond the Gaussian case typically requires additional modelling choices and is left to future work.

References

Michael Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

Fan Bao, Chongxuan Li, Jiacheng Sun, Jun Zhu, and Bo Zhang. Estimating the optimal covariance with imperfect mean in diffusion probabilistic models. In *International Conference on Machine Learning*, pages 2427–2441, 2022a.

Fan Bao, Chongxuan Li, Jun Zhu, and Bo Zhang. Analytic-DPM: an analytic estimate of the optimal reverse variance in diffusion probabilistic models. In *International Conference on Learning Representations*, 2022b.

Georgios Batzolis, Jan Stanczuk, Carola-Bibiane Schönlieb, and Christian Etmann. Conditional image generation with score-based diffusion models. *arXiv preprint arXiv:2111.13606*, 2021.

Benjamin Boys, Mark Girolami, Jakiw Pidstrigach, Sebastian Reich, Alan Mosca, and O. Deniz Akyildiz. Tweedie moment projected diffusions for inverse problems. *Transactions on Machine Learning Research*, 2024.

Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2):671–691, 2016.

Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023.

Marc Corstanje, Frank van der Meulen, and Moritz Schauer. Conditioning continuous-time Markov processes by guiding. *Stochastics*, 95(6):963–996, 2023.

Giannis Daras, Hyungjin Chung, Chieh-Hsin Lai, Yuki Mitsufuji, Jong Chul Ye, Peyman Milanfar, Alexandros G. Dimakis, and Mauricio Delbracio. A survey on diffusion models for inverse problems. *arXiv preprint arXiv:2410.00083*, 2024.

Alexander Denker, Francisco Vargas, Shreyas Padhy, Kieran Didi, Simon Mathis, Riccardo Barbano, Vincent Dutordoir, Emile Mathieu, Urszula Julia Komorowska, and Pietro Lio. DEFT: Efficient fine-tuning of diffusion models by learning the generalised h-transform. In *Advances in Neural Information Processing Systems*, volume 37, pages 19636–19682, 2024.

Alexander Denker, Shreyas Padhy, Francisco Vargas, and Johannes Hertrich. Iterative importance fine-tuning of diffusion models. *arXiv preprint arXiv:2502.04468*, 2025.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pages 8780–8794, 2021.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Ulrich G Haussmann and Etienne Pardoux. Time reversal of diffusions. *The Annals of Probability*, pages 1188–1205, 1986.

J Heng, V De Bortoli, A Doucet, and J Thornton. Simulating diffusion bridges with score matching. *Biometrika*, 112(4), 07 2025.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.

Jonathan Ho, Chitwan Saharia, William Chan, David J. Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *Journal of Machine Learning Research*, 23(47):1–33, 2022.

Peter Holderrieth, Marton Havasi, Jason Yim, Neta Shaul, Itai Gat, Tommi Jaakkola, Brian Karrer, Ricky T. Q. Chen, and Yaron Lipman. Generator matching: Generative modeling with arbitrary Markov processes. In *International Conference on Learning Representations*, pages 52153–52219, 2025.

Kwanyoung Kim and Jong Chul Ye. Noise2Score: Tweedie's approach to self-supervised image denoising without clean images. In *Advances in Neural Information Processing Systems*, volume 34, pages 864–874, 2021.

Diederik Kingma and Ruiqi Gao. Understanding diffusion objectives as the ELBO with simple data augmentation. In *Advances in Neural Information Processing Systems*, volume 36, pages 65484–65516, 2023.

Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. In *Advances in Neural Information Processing Systems*, volume 34, pages 21696–21707, 2021.

Siwei Lyu. Interpretation and generalization of score matching. *arXiv preprint arXiv:1205.2629*, 2012.

Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *arXiv preprint arXiv:1308.0215*, 2013.

Chenlin Meng, Yang Song, Wenzhe Li, and Stefano Ermon. Estimating high order gradients of the data distribution by denoising. In *Advances in Neural Information Processing Systems*, volume 34, pages 25359–25369, 2021.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171, 2021.

Zijing Ou, Mingtian Zhang, Andi Zhang, Tim Xiao, Yingzhen Li, and David Barber. Improving probabilistic diffusion models with optimal diagonal covariance matching. In *International Conference on Learning Representations*, pages 67782–67807, 2025.

Herbert E Robbins. An empirical Bayes approach to statistics. In *Breakthroughs in Statistics: Foundations and basic theory*, pages 388–394. Springer, 1992.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b.

Charles M Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, pages 1135–1151, 1981.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.

Appendix A. Summary of diffusion models

A.1 Noise Schedules and Markov families

Under the settings in Section 2, the deterministic functions $\alpha : [0,1] \to (0,\infty)$ and $\beta : [0,1] \to [0,\infty)$ with $\alpha_0 = 1$, $\beta_0 = 0$, define a time-indexed family of Gaussian corruption kernels

$$p_{t|0}(x_t \mid x_0) = \mathcal{N}(x_t; \alpha_t x_0, \beta_t^2 I_d), \qquad t \in (0,1]. \tag{25}$$

Given an initial distribution p_0 on X_0, this induces a marginal (probability) path $\{p_t\}_{t\in[0,1]}$ via

$$p_t(x_t) = \int p_{t|0}(x_t \mid x_0)\, p_0(x_0)\, dx_0. \tag{26}$$

When we additionally assume Markov consistency across times (Chapman–Kolmogorov), the schedule (α_t, β_t) determines a (time-inhomogeneous) Markov transition family $\{p_{t|s}\}_{0\le s<t\le 1}$ satisfying

$$p_{t|0}(\cdot \mid x_0) = \int p_{t|s}(\cdot \mid x_s)\, p_{s|0}(x_s \mid x_0)\, dx_s. \tag{27}$$

For the linear-Gaussian family above, the transitions are

$$p_{t|s}(x_t \mid x_s) = \mathcal{N}(x_t; \alpha_{t|s} x_s, \beta_{t|s}^2 I_d), \qquad \alpha_{t|s} := \frac{\alpha_t}{\alpha_s}, \quad \beta_{t|s}^2 := \beta_t^2 - \alpha_{t|s}^2 \beta_s^2, \tag{28}$$

and we assume $\beta_{t|s}^2 \ge 0$ for all $0 \le s < t \le 1$.

A.1.1 Continuous-time stochastic differential equation.

Under mild regularity conditions on α_t, β_t (ensuring $g_t^2 \ge 0$ and a continuous-path realisation), the above Markov family admits a continuous-time representation as the linear SDE

$$dX_t = f_t X_t\, dt + g_t\, dW_t, \qquad X_0 \sim p_0, \qquad t \in [0,1], \tag{29}$$

where W_t is standard Brownian motion and

$$f_t := \frac{\dot\alpha_t}{\alpha_t}, \qquad g_t^2 := \dot\beta_t^2 - 2\frac{\dot\alpha_t}{\alpha_t}\beta_t^2. \tag{30}$$

This choice ensures the conditional law $X_t \mid X_0$ matches $\mathcal{N}(\alpha_t X_0, \beta_t^2 I_d)$.

For the forward SDE (29), under standard regularity conditions the exact time-reversal is again an SDE whose drift involves the marginal score $\nabla_{x_t} \log p_t(x_t)$ (Anderson, 1982; Haussmann and Pardoux, 1986; Song et al., 2021b):

$$dX_t = \big(f_t X_t - g_t^2 \nabla_x \log p_t(x)\big)\Big|_{x=X_t} dt + g_t\, d\bar W_t, \qquad X_1 \sim p_1, \qquad t : 1 \to 0, \tag{31}$$

where $\bar W_t$ is a standard Brownian motion backward in time. Thus, modelling the reverse process reduces to learning or approximating $\nabla_{x_t} \log p_t(x_t)$, which is sufficient to specify the reverse-time drift.

In the Gaussian corruption model, Tweedie's identity (11) implies the equivalent score representation

$$\nabla_{x_t} \log p_t(x_t) = \frac{1}{\beta_t^2}\Big(\alpha_t\, \mathbb{E}[X_0 \mid x_t] - x_t\Big). \tag{32}$$

Hence, any estimator of the posterior mean $\mathbb{E}[X_0 \mid x_t]$ immediately yields an estimator of the marginal score $\nabla_{x_t} \log p_t(x_t)$, and vice versa, making the Tweedie viewpoint consistent with the reverse-SDE formulation.

A.1.2 Discrete-time Markov chain.

In discrete time, fix a grid $0 = t_0 < t_1 < \cdots < t_N = 1$ and define the forward Markov chain by the one-step kernels

$$p_{t_n|t_{n-1}}(x_{t_n} \mid x_{t_{n-1}}) = \mathcal{N}\big(x_{t_n}; \alpha_{t_n|t_{n-1}} x_{t_{n-1}},\ \beta^2_{t_n|t_{n-1}} I_d\big), \qquad n = 1, \ldots, N. \tag{33}$$

The joint distribution on the discrete trajectory $x_{0:N} := (x_{t_0}, \ldots, x_{t_N})$ factorises as

$$p(x_{0:N}) = p_0(x_{t_0}) \prod_{n=1}^{N} p_{t_n|t_{n-1}}(x_{t_n} \mid x_{t_{n-1}}). \tag{34}$$

Many discrete-time diffusion objectives, including ELBO/KL formulations (Ho et al., 2020; Song et al., 2021a; Nichol and Dhariwal, 2021; Kingma et al., 2021; Kingma and Gao, 2023) and related KL-based objectives (Bao et al., 2022b,a; Ou et al., 2025), can be interpreted as fitting a parameterised reverse-time chain whose transitions approximate the reverse-time conditionals induced by the forward joint. Concretely, let

$$q_\theta(x_{t_{n-1}} \mid x_{t_n}) = \mathcal{N}\big(x_{t_{n-1}}; \mu_\theta(t_n, x_{t_n}),\ \Sigma_\theta(t_n, x_{t_n})\big), \qquad n = 1, \ldots, N, \tag{35}$$

and define the reverse-chain joint

$$q_\theta(x_{t_{0:N}}) = q_{t_N}(x_{t_N}) \prod_{n=1}^{N} q_\theta(x_{t_{n-1}} \mid x_{t_n}), \qquad q_{t_N}(x_{t_N}) = p_{t_N}(x_{t_N}) \approx \mathcal{N}(0, I_d). \tag{36}$$

A natural objective is to minimise a path-space divergence between the forward and reverse joints. For example, the forward KL decomposes as

$$\mathrm{KL}\big(p(x_{t_{0:N}}) \,\|\, q_\theta(x_{t_{0:N}})\big) = \mathrm{KL}\big(p_{t_N} \,\|\, q_{t_N}\big) + \sum_{n=1}^{N} \mathbb{E}_{p_{t_n}} \left[\mathrm{KL}\big(p(x_{t_{n-1}} \mid x_{t_n}) \,\|\, q_\theta(x_{t_{n-1}} \mid x_{t_n})\big)\right]. \tag{37}$$

Thus, fitting a Gaussian reverse kernel reduces to approximating (directly or indirectly) the conditional moments of the reverse-time conditionals $p(x_{t_{n-1}} \mid x_{t_n})$.

In our setting, Tweedie identities (Section 3.4) provide closed-form expressions for these conditional means and covariances in terms of derivatives of the log-marginal density $\log p_{t_n}(x_{t_n})$ (i.e., the marginal score and, for covariances, the Hessian). This connects the discrete-time Gaussian reverse-kernel viewpoint to the score-based formulation, and clarifies which moments are being approximated by common samplers.

Appendix B. Tweedie's formula for linear Gaussian likelihood

B.1 A direct derivation

Let $p(y \mid x) = \mathcal{N}(y \mid Ax, C)$ be a linear Gaussian likelihood for $x \in \mathbb{R}^d, y \in \mathbb{R}^m, A \in \mathbb{R}^{m\times d}, C \in \mathbb{R}^{m\times m}$. Given $p(y|x)$ and a distribution $p(x)$, we can link the score of the marginal distribution, $\nabla_y \log p(y)$ to the expectation of x with respect to the conditional distribution $p(x|y)$, $\mathbb{E}[x \mid y]$, as

follows,

$$\nabla_y \log p(y) = \frac{\nabla_y p(y)}{p(y)} \quad \text{derivative of natural logarithm} \tag{38}$$

$$= \frac{\nabla_y \int p(y|x)p(x)dx}{p(y)} \quad \text{marginalisation property} \tag{39}$$

$$= \frac{\int \nabla_y p(y|x)p(x)dx}{p(y)} \quad \text{differentiation under integral sign} \tag{40}$$

$$= \frac{\int p(y|x)C^{-1}(Ax-y)p(x)dx}{p(y)} \quad \text{since } p(y|x) = \mathcal{N}(y; Ax, C) \tag{41}$$

$$= \int p(x|y)C^{-1}(Ax-y)dx \quad \text{since } p(y) \text{ does not depend on } x\text{, apply Bayes' rule} \tag{42}$$

$$\boxed{\nabla_y \log p(y) = C^{-1}(A\mathbb{E}[x|y] - y).} \tag{43}$$

This is often known as Tweedie's formula (Efron, 2011). We can extend this to relate the score to the conditional covariance. Consider the derivative of the score, or the second derivative of the log marginal distribution,

$$\nabla_y^2 \log p(y) = \nabla_y \left[C^{-1}(A\mathbb{E}[x|y] - y)\right] = C^{-1}(A\nabla_y \mathbb{E}[x|y] - I_m). \tag{44}$$

We note that

$$\nabla_y \mathbb{E}[x|y] = \nabla_y \int x \frac{p(y|x)p(x)}{p(y)} dx \quad \text{Bayes' rule} \tag{45}$$

$$= \int x \left[\frac{\nabla_y p(y|x)p(x)}{p(y)} - \frac{p(y|x)p(x)\nabla_y p(y)}{p(y)^2}\right] dx \tag{46}$$

$$= \int xp(x|y)[\nabla_y \log p(y|x) - \nabla_y \log p(y)]dx \quad \text{derivative of natural log} \tag{47}$$

$$= \int xp(x|y)[(x^\intercal A - y^\intercal)C^{-1}]dx - \nabla_y \log p(y) \int xp(x|y)dx \quad \log p(y) \text{ does not depend on } x \tag{48}$$

$$= \left(\mathbb{E}[xx^\intercal|y]A^\intercal - \mathbb{E}[x|y]y^\intercal\right)C^{-1} - C^{-1}(A\mathbb{E}[x|y] - y)\mathbb{E}[x|y]^\intercal \tag{49}$$

$$= \mathrm{Cov}[x|y]A^\intercal C^{-1}. \tag{50}$$

Therefore,

$$\boxed{\nabla_y^2 \log p(y) = C^{-1}A\mathrm{Cov}[x|y]A^\intercal C^{-1} - C^{-1}I_m.} \tag{51}$$

We also have:

$$\nabla_y^2 \log p(y) = \nabla_y \nabla_y \log p(y) \tag{52}$$

$$= \nabla_y \frac{\nabla_y p(y)}{p(y)} \tag{53}$$

$$= \frac{\nabla_y^2 p(y)}{p(y)} - \left(\frac{\nabla_y p(y)}{p(y)}\right)\left(\frac{\nabla_y p(y)}{p(y)}\right)^\intercal \tag{54}$$

$$= \frac{\nabla_y^2 p(y)}{p(y)} - \nabla_y \log p(y)\left(\nabla_y \log p(y)\right)^\intercal \tag{55}$$

Integrating both sides with respect to $p(y)$ leads to

$$\int \left[\nabla_y^2 \log p(y)\right] p(y)dy = \int \frac{\nabla_y^2 p(y)}{p(y)} p(y)dy - \int \left[\nabla_y \log p(y) \left(\nabla_y \log p(y)\right)^\intercal\right] p(y) \tag{56}$$

$$= \underbrace{\nabla_y^2 \int p(y)dy}_{=0} - \int \left[\nabla_y \log p(y) \left(\nabla_y \log p(y)\right)^\intercal\right] p(y)dy \tag{57}$$

$$\boxed{\int \left[\nabla_y^2 \log p(y)\right] p(y)dy = -\int \left[\nabla_y \log p(y) \left(\nabla_y \log p(y)\right)^\intercal\right] p(y)dy} \tag{58}$$

Combining the two boxed results above gives

$$C^{-1}A\left[\int \mathrm{Cov}[x|y]p(y)dy\right]A^\intercal C^{-1} - C^{-1}I_m = -\int \left[\nabla_y \log p(y) \left(\nabla_y \log p(y)\right)^\intercal\right] p(y)dy \tag{59}$$

When $A = a$ and $C = c^2 I_m$, we obtain

$$\nabla_y^2 \log p(y) = \frac{a^2}{c^4}\mathrm{Cov}[x|y] - \frac{1}{c^2}I_m, \tag{60}$$

$$\frac{a^2}{c^4}\left[\int \mathrm{Cov}[x|y]p(y)dy\right] - \frac{1}{c^2}I_m = -\int \left[\nabla_y \log p(y) \left(\nabla_y \log p(y)\right)^\intercal\right] p(y)dy, \tag{61}$$

$$\text{or,} \boxed{\int \mathrm{Cov}[x|y]p(y)dy = \frac{c^2}{a^2}I_m - \frac{c^4}{a^2}\int \left[\nabla_y \log p(y) \left(\nabla_y \log p(y)\right)^\intercal\right] p(y)dy} \tag{62}$$

Furthermore,

$$\boxed{\int \mathrm{trace}(\mathrm{Cov}[x|y])p(y)dy = \frac{dc^2}{a^2} - \frac{c^4}{a^2}\int ||\nabla_y \log p(y)||_2^2 p(y)dy} \tag{63}$$

B.2 A derivation using the exponential family

Consider the linear Gaussian case in Section 3.1 where $p(y \mid x) = \mathcal{N}(y; Ax, C)$ with $x \in \mathbb{R}^d$, $y \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times d}$, and $C \in \mathbb{R}^{m\times m}$ symmetric positive definite. Define the mapping

$$\phi_C(y)^\top = \begin{bmatrix} y^\top C^{-1} & -\frac{1}{2}\,\mathrm{vec}(C^{-1})^\top \end{bmatrix}, \qquad \psi_A(x)^\top = \begin{bmatrix} (Ax)^\top & \mathrm{vec}(Axx^\top A^\top)^\top \end{bmatrix}. \tag{64}$$

We also define the base Gaussian density $\rho_C(y) := \mathcal{N}(y; 0, C)$.

Gaussian pairing form. Expanding the quadratic in the likelihood gives

$$\begin{aligned} p(y \mid x) &= (2\pi)^{-m/2}|C|^{-1/2}\exp\Big\{-\tfrac{1}{2}(y-Ax)^\top C^{-1}(y-Ax)\Big\} \\ &= \rho_C(y)\exp\Big\{y^\top C^{-1}Ax - \tfrac{1}{2}(Ax)^\top C^{-1}(Ax)\Big\} \\ &= \rho_C(y)\exp\Big\{\phi_C(y)^\top \psi_A(x)\Big\}. \end{aligned} \tag{65}$$

Marginal, normaliser, and posterior. Define $\eta := \phi_C(y) \in \mathbb{R}^{m+m^2}$ and split $\eta = (\eta_1, \eta_2)$ with

$$\eta_1 = C^{-1}y \in \mathbb{R}^m, \qquad \eta_2 = -\tfrac{1}{2}\,\mathrm{vec}(C^{-1}) \in \mathbb{R}^{m^2}. \tag{66}$$

We view $\rho_C(y)$ as a function of η by introducing

$$U(\eta_2) = -2\,\mathrm{unvec}(\eta_2) \in \mathbb{R}^{m\times m}, \qquad \text{so that } U(\eta_2) = C^{-1}. \tag{67}$$

Then $y = U(\eta_2)^{-1}\eta_1$ and

$$\rho(\eta) = \rho_C(y)\Big|_{y=U(\eta_2)^{-1}\eta_1} = (2\pi)^{-m/2}\,|U(\eta_2)|^{1/2}\exp\Big\{-\tfrac{1}{2}\,\eta_1^\top U(\eta_2)^{-1}\eta_1\Big\}. \tag{68}$$

Let $p(x)$ be any prior/base measure on x. The marginal is

$$p(y) = \int p(x)p(y \mid x)\,dx = \rho(\phi_C(y))\int p(x)\exp\{\phi_C(y)^\top\psi_A(x)\}\,dx =: Z(\phi_C(y)). \tag{69}$$

Using $\eta = \phi_C(y)$,

$$Z(\eta) = \rho(\eta)\int p(x)\exp\{\eta^\top\psi_A(x)\}\,dx, \qquad A(\eta) := \log\frac{Z(\eta)}{\rho(\eta)} = \log\int p(x)\exp\{\eta^\top\psi_A(x)\}\,dx. \tag{70}$$

By Bayes' rule, the posterior can be written as

$$p(x \mid y) = \frac{p(x)p(y \mid x)}{p(y)} = p(x)\exp\{\eta^\top\psi_A(x) - A(\eta)\}, \qquad \eta = \phi_C(y). \tag{71}$$

Derivatives in η_1-space Let $\eta_1 = \phi_{C,1}(y)$. Differentiating under the integral sign,

$$\begin{aligned}
\nabla_{\eta_1}A(\eta) &= \nabla_{\eta_1}\log\int p(x)\exp\{\eta^\top\psi_A(x)\}\,dx \\
&= \frac{\int p(x)\,\nabla_{\eta_1}\exp\{\eta^\top\psi_A(x)\}\,dx}{\int p(x)\exp\{\eta^\top\psi_A(x)\}\,dx} \\
&= \frac{\int p(x)\,\psi_{A,1}(x)\exp\{\eta^\top\psi_A(x)\}\,dx}{\int p(x)\exp\{\eta^\top\psi_A(x)\}\,dx} \\
&= \int \psi_{A,1}(x)\,p(x)\exp\{\eta^\top\psi_A(x) - A(\eta)\}\,dx \\
&= \mathbb{E}[\psi_{A,1}(X) \mid \eta] = \mathbb{E}[AX \mid \eta] = A\,\mathbb{E}[X \mid \eta].
\end{aligned} \tag{72}$$

Similarly,

$$\begin{aligned}
\nabla^2_{\eta_1}A(\eta) &= \nabla_{\eta_1}\Big(\mathbb{E}[AX \mid \eta]\Big) \\
&= \int (Ax)\big(Ax - \nabla_{\eta_1}A(\eta)\big)^\top p(x)\exp\{\eta^\top\psi_A(x) - A(\eta)\}\,dx \\
&= \mathbb{E}[AXX^\top A^\top \mid \eta] - \mathbb{E}[AX \mid \eta]\mathbb{E}[AX \mid \eta]^\top \\
&= A\,\mathrm{Cov}[X \mid \eta]\,A^\top.
\end{aligned} \tag{73}$$

Chain rule from η_1 to y. For $\eta = \phi_C(y)$, the Jacobian of $\eta_1 = C^{-1}y$ is $J_{\eta_1}(y) = C^{-1}$. Since C is fixed, η_2 is constant in y, so only η_1 contributes in y-derivatives. Using $\log p(y) = \log\rho_C(y) + A(\phi_C(y))$,

$$\begin{aligned}
\nabla_y A(\eta)\big|_{\eta=\phi_C(y)} &= \nabla_y\log p(y) - \nabla_y\log\rho_C(y) = J_{\eta_1}(y)^\top\nabla_{\eta_1}A(\eta)\big|_{\eta=\phi_C(y)} \\
&= C^{-1}\,A\,\mathbb{E}[X \mid y],
\end{aligned} \tag{74}$$

$$\begin{aligned}
\nabla^2_y A(\eta)\big|_{\eta=\phi_C(y)} &= \nabla^2_y\log p(y) - \nabla^2_y\log\rho_C(y) = C^{-1}\nabla^2_{\eta_1}A(\eta)\big|_{\eta=\phi_C(y)}C^{-1} \\
&= C^{-1}A\,\mathrm{Cov}[X \mid y]\,A^\top C^{-1}.
\end{aligned} \tag{75}$$

Since $\nabla_y\log\rho_C(y) = -C^{-1}y$ and $\nabla^2_y\log\rho_C(y) = -C^{-1}$, we obtain

$$\nabla_y\log p(y) = C^{-1}\big(A\mathbb{E}[X \mid y] - y\big), \tag{76}$$

$$\nabla^2_y\log p(y) = C^{-1}A\,\mathrm{Cov}[X \mid y]\,A^\top C^{-1} - C^{-1}. \tag{77}$$

Thus, for $A = aI$ and $C = c^2 I_m$,

$$\mathbb{E}[X \mid y] = \frac{1}{a}\big(y + c^2 \nabla_y \log p(y)\big), \tag{78}$$

$$\mathrm{Cov}[X \mid y] = \frac{c^2}{a^2}\big(c^2 \nabla_y^2 \log p(y) + I_m\big). \tag{79}$$

Appendix C. Related Work

Tweedie's identities, empirical Bayes, and exponential-family derivations. Tweedie-style identities relate posterior moments under a corruption model to derivatives of the corresponding log-marginal density, with classical roots in empirical Bayes (Robbins, 1992; Efron, 2011). In the Gaussian corruption setting, Meng et al. (2021) use such identities to estimate higher-order derivatives of the log-likelihood, providing an alternative to differentiating a learned score model using the denoising score matching objectives (Vincent, 2011; Lyu, 2012). Kim and Ye (2021) extend Tweedie-style reasoning beyond Gaussian noise to broader exponential-family corruption models for self-supervised denoising. Here, we present a compact normaliser-differentiation template that (i) unifies these derivations under a common notation and (ii) makes explicit how changing the base measure (e.g. conditioning, guidance, or prior tilting) systematically shifts posterior moments and how these posterior moments can be connected when a fixed Markov Gaussian corruption family is used.

Reverse-process covariance and intermediate-time conditionals in diffusion models. In discrete-time diffusion sampling, constructing Gaussian reverse transitions requires moments of intermediate-time conditionals such as $X_s \mid X_t$, not only $X_0 \mid X_t$. Several works derive score-based expressions for these conditional covariances and propose practical estimators or parameterisations for reverse-time variances (Bao et al., 2022b,a; Ou et al., 2025). These works commonly rely on identities such as the law of total variance and relate conditional variances to the marginal score. In contrast, we derive the same intermediate-time moment identities (including bridge decompositions) from a single Tweedie/normaliser-differentiation viewpoint, and we use this to organise how different "gluing" operations (conditioning and prior tilting) modify the corresponding reverse-time moments.

Conditional generation and changing the prior. A standard route to conditional generation with diffusion priors starts from Bayes' rule and yields the score decomposition in (18), which motivates guidance methods that combine an unconditional diffusion score with an additional "guidance" term (Dhariwal and Nichol, 2021; Chung et al., 2023). More generally, conditioning and target reweighting can be formalised as a change of path measure via Doob's h-transform and related bridge constructions, providing a principled view of conditional diffusion dynamics (Corstanje et al., 2023; Heng et al., 2025; Denker et al., 2024, 2025). These formulations are closely connected to Schrödinger bridge and half-bridge problems, which characterise conditional path measures under endpoint or observation constraints (Léonard, 2013; Chen et al., 2016; Heng et al., 2025; Denker et al., 2025). In contrast to deriving reverse-time conditioning dynamics primarily at the path/SDE level, we present a compact Tweedie-style normaliser-differentiation derivation that makes the induced moment shifts (mean and covariance) explicit under Gaussian forward kernels, which is directly compatible with the Gaussian reverse-transition approximations used in discrete-time samplers.