Here are the comprehensive lecture notes covering the mathematical derivations and concepts from the provided PDF regarding Diffusion Models, Tweedie's Formula, and DDPM.

Lecture Notes: Diffusion Models & Mathematical Preliminaries

1. Tweedie's Formula (Preliminary)

Goal: To find the posterior mean $E[X|Y=y]$ and variance $Var[X|Y=y]$ given a noisy observation, using the score function (gradient of the log-likelihood).

A. Problem Setup

Consider a clean signal X and a noisy observation Y:

$$Y = X + \sigma Z$$

- $X \sim P_X$ (Unknown data distribution)
- $Z \sim \mathcal{N}(0, I)$ (Gaussian noise)
- $Y|X \sim \mathcal{N}(X, \sigma^2 I)$

We define $\phi_\sigma(u)$ as the probability density function (PDF) of the noise $\mathcal{N}(0, \sigma^2 I_d)$:

$$\phi_\sigma(u) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{\|u\|^2}{2\sigma^2}\right)$$

The marginal density of Y is:

$$P_Y(y) = \int_{\mathbb{R}^d} P_X(x)\phi_\sigma(y-x)dx$$

B. Derivation of the Gradient

We compute the gradient of the marginal density $\nabla_y P_Y(y)$:

1. **Differentiation under the integral:**

$$\nabla_y P_Y(y) = \int P_X(x) \nabla_y \phi_\sigma(y-x) dx$$

2. **Kernel Derivative Property:**
   Note that for the Gaussian kernel:

$$\nabla_y \phi_\sigma(y-x) = -\frac{y-x}{\sigma^2} \phi_\sigma(y-x)$$

3. **Substitution:**
   Substituting this back into the integral:

$$\nabla_y P_Y(y) = -\frac{1}{\sigma^2} \int (y-x) P_X(x) \phi_\sigma(y-x) dx$$

$$\nabla_y P_Y(y) = -\frac{1}{\sigma^2} \left[y \underbrace{\int P_X(x) \phi_\sigma(y-x) dx}_{P_Y(y)} - \int x P_X(x) \phi_\sigma(y-x) dx \right]$$

4. **Bayes' Rule Identity:**
   Recall that $P_{X,Y}(x,y) = P_X(x)\phi_\sigma(y-x) = P_Y(y)P_{X|Y}(x|y)$.
   Therefore, the second integral becomes:

$$\int x P_{X,Y}(x,y) dx = P_Y(y) \int x P_{X|Y}(x|y) dx = P_Y(y) E[X|Y=y]$$

5. **Final Equation:**

$$\nabla_y P_Y(y) = -\frac{1}{\sigma^2} \left(y P_Y(y) - P_Y(y) E[X|Y=y] \right)$$

C. Resulting Formula (Tweedie's Formula)

Rearranging the terms to solve for $E[X|Y]$:

$$\frac{\nabla_y P_Y(y)}{P_Y(y)} = \frac{1}{\sigma^2} (E[X|Y=y] - y)$$

$$\nabla_y \log P_Y(y) = \frac{1}{\sigma^2} (E[X|Y=y] - y)$$

Expectation:

$$E[X|Y=y] = y + \sigma^2 \nabla_y \log P_Y(y)$$

Variance:

$$Var[X|Y=y] = \sigma^2 (I + \sigma^2 \nabla_y^2 \log P_Y(y))$$

2. Gaussian Approximation of the Posterior

Goal: Prove that if σ is sufficiently small, the posterior $P_{X|Y}(x|y)$ approximates a Gaussian distribution.

A. Bayes' Expansion

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x) P_X(x)}{P_Y(y)}$$

Ignoring the normalization constant $P_Y(y)$, we focus on the numerator:

$$P_{X|Y}(x|y) \propto \exp\left(-\frac{\|y-x\|^2}{2\sigma^2}\right) P_X(x)$$

B. Taylor Expansion

We expand $P_X(x)$ around y (assuming $\sigma \to 0$, x is close to y):

$$P_X(x) \approx P_X(y) + \langle \nabla P_X(y), x-y \rangle + O(\|x-y\|^2)$$

Using the approximation $1 + a \approx e^a$, we rewrite the linear term in log-space:

$$P_X(x) \approx P_X(y) \left(1 + \langle \nabla \log P_X(y), x-y \rangle\right)$$

$$P_X(x) \approx P_X(y) \exp(\langle \nabla \log P_X(y), x-y \rangle)$$

C. Completing the Square

Combining the likelihood and the prior approximation:

$$P_{X|Y}(x|y) \propto \exp\left(-\frac{\|x-y\|^2}{2\sigma^2} + \langle \nabla \log P_X(y), x-y \rangle\right)$$

Let $l_1 = x - y$ and $l_2 = \nabla \log P_X(y)$. We seek to complete the square for the term in the exponent:

$$-\frac{1}{2\sigma^2}\|l_1\|^2 + \langle l_1, l_2 \rangle = -\frac{1}{2\sigma^2}\left(\|l_1\|^2 - 2\sigma^2 \langle l_1, l_2 \rangle\right)$$

By adding and subtracting the squared term $\sigma^4 \|l_2\|^2$:

$$= -\frac{1}{2\sigma^2}\|l_1 - \sigma^2 l_2\|^2 + \frac{\sigma^2}{2}\|l_2\|^2$$

Substituting l_1 and l_2 back:

$$\text{Exponent} \approx -\frac{1}{2\sigma^2}\|(x-y) - \sigma^2 \nabla \log P_X(y)\|^2 + \text{Constant}$$

D. Final Approximation

The posterior follows a Gaussian distribution:

$$P_{X|Y}(x|y) \approx \mathcal{N}\left(y + \sigma^2 \nabla \log P_X(y), \sigma^2 I\right)$$

Note: For small σ, $\nabla \log P_X(y) \approx \nabla \log P_Y(y)$, making this consistent with Tweedie's formula.

3. Denoising Diffusion Probabilistic Models (DDPM)

A. Forward Diffusion Process

We define a forward process that gradually adds noise to data $x_0 \sim P_{\text{data}}$.
For $t = 1, \dots, T$:

$$x_t | x_{t-1} \sim \mathcal{N}(\sqrt{1-\beta_t}\, x_{t-1}, \beta_t I)$$

where $0 < \beta_t < 1$ represents the variance schedule.

Let $\alpha_t = 1 - \beta_t$. The transition can be written using reparameterization:

$$x_t = \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t}\, z_t, \quad z_t \sim \mathcal{N}(0, I)$$

B. Deriving the Marginal $q(x_t|x_0)$

We want to sample x_t directly from x_0.
Define $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$.

Recursive Substitution:

1. Start with x_t:

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1-\alpha_t} z_t$$

2. Substitute $x_{t-1} = \sqrt{\alpha_{t-1}} x_{t-2} + \sqrt{1-\alpha_{t-1}} z_{t-1}$:

$$x_t = \sqrt{\alpha_t}(\sqrt{\alpha_{t-1}} x_{t-2} + \sqrt{1-\alpha_{t-1}} z_{t-1}) + \sqrt{1-\alpha_t} z_t$$

$$x_t = \sqrt{\alpha_t \alpha_{t-1}} x_{t-2} + \sqrt{\alpha_t(1-\alpha_{t-1})} z_{t-1} + \sqrt{1-\alpha_t} z_t$$

Gaussian Summation Tip:
If $X_1 \sim \mathcal{N}(0, \sigma_1^2 I)$ and $X_2 \sim \mathcal{N}(0, \sigma_2^2 I)$ are independent, then:

$$X_1 + X_2 \sim \mathcal{N}(0, (\sigma_1^2 + \sigma_2^2) I)$$

Applying this to the noise terms:

- Variance of first noise term: $\alpha_t(1-\alpha_{t-1})$
- Variance of second noise term: $(1-\alpha_t)$
- Total Variance: $\alpha_t - \alpha_t \alpha_{t-1} + 1 - \alpha_t = 1 - \alpha_t \alpha_{t-1}$

Thus, the merged noise term is $\sqrt{1-\alpha_t \alpha_{t-1}} \bar{z}$.

C. General Result

By induction, extending this to $t = 0$:

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t} \epsilon, \quad \epsilon \sim \mathcal{N}(0, I)$$

This yields the marginal distribution:

$$x_t | x_0 \sim \mathcal{N}(\sqrt{\bar{\alpha}_t} x_0, (1-\bar{\alpha}_t) I)$$