Assignment 3

Joshmitha Gunda

September 2025

1 Question 1

Consider the following definitions:

- $I(x, y)$: True noise-free intensity at pixel coordinate (x, y). For this exercise, it is constant across all $N = 10{,}000$ images.
- $J_i(x, y)$: Measured intensity at pixel (x, y) in the i-th image ($i = 1, \dots, N$).
- $W_i(x, y)$: Noise affecting pixel (x, y) in the i-th image.
- $N = 10{,}000$: Total number of images captured.

The noisy image can be modeled as:

$$J_i(x, y) = I(x, y) + W_i(x, y),$$

where $E[W_i(x, y)] = 0$ and the noise samples are independent across all pixels.

Method to Estimate Noise Distribution

1. **Estimate true intensity:**
   Since the noise has zero mean, the true intensity can be estimated by averaging over all N images:

   $$\hat{I}(x, y) = \frac{1}{N} \sum_{i=1}^{N} J_i(x, y).$$

   This works because, on average, the positive and negative noise values cancel each other out.

2. **Extract noise samples:**
   For each image i, the noise affecting pixel (x, y) can be estimated as:

   $$\hat{W}_i(x, y) = J_i(x, y) - \hat{I}(x, y).$$

   Collecting these values across all N images provides a large set of noise samples for each pixel. Since the noise in each pixel is independent, we can treat all these samples as separate examples of the noise.

3. **Estimate noise distribution:**
   To get an idea of the noise distribution, we can plot a histogram of all the $\hat{W}_i(x, y)$ values. Then, divide the count in each bin by the total number of samples to convert it into probabilities:

$$P(\hat{W} = w) \approx \frac{\text{count of samples in bin containing } w}{\text{total number of samples}}.$$

   This normalized histogram provides an estimate of the noise distribution.

2 Question 2

(a) Proof that $\{v_i\}_{i=1}^n$ follow distribution F

Let:

- F: Continuous cumulative distribution function (CDF), and is invertible
- $u_i \sim \text{Uniform}[0, 1]$, for $i = 1, \ldots, n$
- $v_i = F^{-1}(u_i)$, for $i = 1, \ldots, n$

We want to prove that each v_i follows distribution F, i.e., $P(v_i \leq t) = F(t)$ for all t.

For each $i = 1, \ldots, n$, let $G_i(t)$ denote the CDF of v_i. Then:

$$G_i(t) = P(v_i \leq t) = P(F^{-1}(u_i) \leq t).$$

Since F is invertible and increasing, we can apply F to both sides of the inequality:

$$P(F^{-1}(u_i) \leq t) = P(u_i \leq F(t)).$$

Because $u_i \sim \text{Uniform}[0, 1]$, we have:

$$P(u_i \leq F(t)) = F(t).$$

Therefore, for each $i = 1, \ldots, n$:

$$G_i(t) = F(t) \quad \text{for all } t.$$

Hence, the collection $\{v_i\}_{i=1}^n$ all follow the distribution F.

(b) Proof that $P(E \geq d) = P(D \geq d)$

Definitions:

- $Y_1, Y_2, \ldots, Y_n \sim F$ (continuous distribution)
- Empirical distribution:

$$F_e(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(Y_i \leq x),$$

- $D = \max_x |F_e(x) - F(x)|$.
- $U_1, U_2, \ldots, U_n \sim \text{Uniform}[0,1]$.
- $$E = \max_{0 \le y \le 1} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(U_i \le y) - y \right|.$$

From part (a), if $U_i \sim \text{Uniform}[0,1]$, then

$$Y_i = F^{-1}(U_i) \sim F.$$

Now consider the empirical distribution:

$$F_e(x) = \frac{1}{n} \sum_{i=1}^{n} 1(Y_i \le x) = \frac{1}{n} \sum_{i=1}^{n} 1(F^{-1}(U_i) \le x).$$

Since F is increasing and continuous:

$$1(F^{-1}(U_i) \le x) = 1(U_i \le F(x)).$$

Therefore, the empirical distribution can be rewritten as:

$$F_e(x) = \frac{1}{n} \sum_{i=1}^{n} 1(U_i \le F(x)).$$

Let $y = F(x)$. As x ranges over all real numbers, y ranges over $[0,1]$. Thus:

$$F_e(x) - F(x) = \frac{1}{n} \sum_{i=1}^{n} 1(U_i \le F(x)) - F(x) = \frac{1}{n} \sum_{i=1}^{n} 1(U_i \le y) - y.$$

Taking the maximum over all x gives:

$$D = \max_x |F_e(x) - F(x)| = \max_{0 \le y \le 1} \left| \frac{1}{n} \sum_{i=1}^{n} 1(U_i \le y) - y \right| = E.$$

Since $D = E$, their probabilities are identical:

$$P(D \ge d) = P(E \ge d) \quad \text{for all } d \ge 0.$$

(c) Practical significance

This result basically shows that we can make random numbers follow any distribution we want, just by starting with uniform random numbers and applying the inverse CDF
This lets us simulate data from more complicated distributions So, it's a simple way to get random samples from any distribution, even if we can't generate them directly

3 Question 3

(a)

Given equation of the plane is $z = a$x+by+c
We have access to accurate X and Y coordinates of some N points lying on the plane. We also have access to the Z coordinates of these points, but those have been corrupted independently by noise from $N(0, \sigma^2)$
so z_i=ax_i+by_i+c+θ_i where $\theta_i \sim N(0, \sigma^2)$
so

Log-Likelihood Function

The probability density function for each observation:

$$p(z_i|x_i, y_i, a, b, c) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z_i - (ax_i + by_i + c))^2}{2\sigma^2}\right]$$

The log-likelihood function:

$$\begin{aligned} L(a, b, c) &= \log \prod_{i=1}^{N} p(z_i|x_i, y_i, a, b, c) \\ &= \sum_{i=1}^{N} \left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(z_i - ax_i - by_i - c)^2}{2\sigma^2}\right] \\ &= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(z_i - ax_i - by_i - c)^2 \end{aligned}$$

To maximize log-likelihood, we minimize:

$$P = \sum_{i=1}^{N}(z_i - ax_i - by_i - c)^2$$

Minimizing P

Setting partial derivatives to zero:

$$\begin{aligned} \frac{\partial P}{\partial a} &= \sum_{i=1}^{N} 2(z_i - ax_i - by_i - c)(-x_i) = 0 \\ &\Rightarrow \sum_{i=1}^{N} x_i z_i = a\sum_{i=1}^{N} x_i^2 + b\sum_{i=1}^{N} x_i y_i + c\sum_{i=1}^{N} x_i \end{aligned}$$

$$\frac{\partial P}{\partial b} = \sum_{i=1}^{N} 2(z_i - ax_i - by_i - c)(-y_i) = 0$$

$$\Rightarrow \sum_{i=1}^{N} y_i z_i = a \sum_{i=1}^{N} x_i y_i + b \sum_{i=1}^{N} y_i^2 + c \sum_{i=1}^{N} y_i$$

$$\frac{\partial P}{\partial c} = \sum_{i=1}^{N} 2(z_i - ax_i - by_i - c)(-1) = 0$$

$$\Rightarrow \sum_{i=1}^{N} z_i = a \sum_{i=1}^{N} x_i + b \sum_{i=1}^{N} y_i + cN$$

Matrix Form

The system of equations can be expressed in matrix form as:

$$\begin{bmatrix} \sum x_i^2 & \sum x_i y_i & \sum x_i \\ \sum x_i y_i & \sum y_i^2 & \sum y_i \\ \sum x_i & \sum y_i & N \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum x_i z_i \\ \sum y_i z_i \\ \sum z_i \end{bmatrix}$$

More compactly:

$$\mathbf{P}\mathbf{t} = \mathbf{k}$$

where

$$\mathbf{P} = \begin{bmatrix} \sum x_i^2 & \sum x_i y_i & \sum x_i \\ \sum x_i y_i & \sum y_i^2 & \sum y_i \\ \sum x_i & \sum y_i & N \end{bmatrix}, \quad \mathbf{t} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} \sum x_i z_i \\ \sum y_i z_i \\ \sum z_i \end{bmatrix}$$

(b)

Given equation of the plane is now

$$z = a_1 x^2 + a_2 y^2 + a_3 xy + a_4 x + a_5 y + a_6$$

We have access to accurate X and Y coordinates of some N points lying on the plane, we also have access to the Z coordinate of these points but those have been corrupted independently by noise from $N(0, \sigma^2)$

$$z_i = a_1 x_i^2 + a_2 y_i^2 + a_3 x_i y_i + a_4 x_i + a_5 y_i + a_6 + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

Log-Likelihood Function

The probability density function:

$$p(z_i|x_i, y_i, a_1, \ldots, a_6) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z_i - (a_1x_i^2 + a_2y_i^2 + a_3x_iy_i + a_4x_i + a_5y_i + a_6))^2}{2\sigma^2}\right]$$

The log-likelihood function:

$$\begin{aligned}
L(a_1, \ldots, a_6) &= \log \prod_{i=1}^{N} p(z_i|x_i, y_i, a_1, \ldots, a_6) \\
&= \sum_{i=1}^{N} \left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(z_i - (a_1x_i^2 + a_2y_i^2 + a_3x_iy_i + a_4x_i + a_5y_i + a_6))^2}{2\sigma^2}\right] \\
&= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6)^2
\end{aligned}$$

To maximize log-likelihood, we minimize:

$$P = \sum_{i=1}^{N}(z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6)^2$$

Partial Derivatives

Setting partial derivatives to zero:

$$\begin{aligned}
\frac{\partial P}{\partial a_1} &= \sum_{i=1}^{N} 2(z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6)(-x_i^2) = 0 \\
&\Rightarrow \sum x_i^2 z_i = a_1 \sum x_i^4 + a_2 \sum x_i^2y_i^2 + a_3 \sum x_i^3y_i + a_4 \sum x_i^3 + a_5 \sum x_i^2y_i + a_6 \sum x_i^2
\end{aligned}$$

$$\begin{aligned}
\frac{\partial P}{\partial a_2} &= \sum_{i=1}^{N} 2(z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6)(-y_i^2) = 0 \\
&\Rightarrow \sum y_i^2 z_i = a_1 \sum x_i^2y_i^2 + a_2 \sum y_i^4 + a_3 \sum x_iy_i^3 + a_4 \sum x_iy_i^2 + a_5 \sum y_i^3 + a_6 \sum y_i^2
\end{aligned}$$

$$\begin{aligned}
\frac{\partial P}{\partial a_3} &= \sum_{i=1}^{N} 2(z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6)(-x_iy_i) = 0 \\
&\Rightarrow \sum x_iy_iz_i = a_1 \sum x_i^3y_i + a_2 \sum x_iy_i^3 + a_3 \sum x_i^2y_i^2 + a_4 \sum x_i^2y_i + a_5 \sum x_iy_i^2 + a_6 \sum x_iy_i
\end{aligned}$$

$$\frac{\partial P}{\partial a_4} = \sum_{i=1}^{N} 2(z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)(-x_i) = 0$$
$$\Rightarrow \sum x_i z_i = a_1 \sum x_i^3 + a_2 \sum x_i y_i^2 + a_3 \sum x_i^2 y_i + a_4 \sum x_i^2 + a_5 \sum x_i y_i + a_6 \sum x_i$$

$$\frac{\partial P}{\partial a_5} = \sum_{i=1}^{N} 2(z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)(-y_i) = 0$$
$$\Rightarrow \sum y_i z_i = a_1 \sum x_i^2 y_i + a_2 \sum y_i^3 + a_3 \sum x_i y_i^2 + a_4 \sum x_i y_i + a_5 \sum y_i^2 + a_6 \sum y_i$$

$$\frac{\partial P}{\partial a_6} = \sum_{i=1}^{N} 2(z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)(-1) = 0$$
$$\Rightarrow \sum z_i = a_1 \sum x_i^2 + a_2 \sum y_i^2 + a_3 \sum x_i y_i + a_4 \sum x_i + a_5 \sum y_i + a_6 N$$

Matrix Form

The system of equations can be expressed in matrix form as:

$$\begin{bmatrix} \sum x_i^4 & \sum x_i^2 y_i^2 & \sum x_i^3 y_i & \sum x_i^3 & \sum x_i^2 y_i & \sum x_i^2 \\ \sum x_i^2 y_i^2 & \sum y_i^4 & \sum x_i y_i^3 & \sum x_i y_i^2 & \sum y_i^3 & \sum y_i^2 \\ \sum x_i^3 y_i & \sum x_i y_i^3 & \sum x_i^2 y_i^2 & \sum x_i^2 y_i & \sum x_i y_i^2 & \sum x_i y_i \\ \sum x_i^3 & \sum x_i y_i^2 & \sum x_i^2 y_i & \sum x_i^2 & \sum x_i y_i & \sum x_i \\ \sum x_i^2 y_i & \sum y_i^3 & \sum x_i y_i^2 & \sum x_i y_i & \sum y_i^2 & \sum y_i \\ \sum x_i^2 & \sum y_i^2 & \sum x_i y_i & \sum x_i & \sum y_i & N \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} = \begin{bmatrix} \sum x_i^2 z_i \\ \sum y_i^2 z_i \\ \sum x_i y_i z_i \\ \sum x_i z_i \\ \sum y_i z_i \\ \sum z_i \end{bmatrix}$$

More compactly:

$$\mathbf{P}\mathbf{t} = \mathbf{k}$$

where

$$\mathbf{t} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} \sum x_i^2 z_i \\ \sum y_i^2 z_i \\ \sum x_i y_i z_i \\ \sum x_i z_i \\ \sum y_i z_i \\ \sum z_i \end{bmatrix}$$

and the matrix **P** is as defined above

(c)

Is knowledge of σ^2 required for model fitting?

No, knowledge of the noise variance σ^2 is not required for estimating the parameters

Reason:

From the log-likelihood function:

$$L(a,b,c) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(z_i - ax_i - by_i - c)^2$$

- The first term $-\frac{N}{2}\log(2\pi\sigma^2)$ is constant with respect to a, b, c
- The second term has σ^2 as a constant multiplier
- **Maximizing $L(a,b,c)$ is equivalent to minimizing $\sum_{i=1}^{N}(z_i - ax_i - by_i - c)^2$**

The same applies to the polynomial surface fitting case.

Estimating the noise variance σ^2

Maximum Likelihood Estimator:

After obtaining parameter estimates $\hat{a}, \hat{b}, \hat{c}$, maximize likelihood with respect to σ^2:

$$\frac{\partial L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{N}(z_i - \hat{a}x_i - \hat{b}y_i - \hat{c})^2 = 0$$

Solving gives:

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{N}\sum_{i=1}^{N}(z_i - \hat{a}x_i - \hat{b}y_i - \hat{c})^2$$

4 Question 4

(b)

$$(b) \quad V = \{v_1, v_2, \ldots, v_{250}\}$$

$$\hat{p}_n(x;\sigma) = \frac{1}{n\sqrt{\pi}}\sum_{i=1}^{n}\exp\left(-\frac{(x-x_i)^2}{2\sigma^2}\right)$$

$$L(V \mid T;\sigma) = \prod_{i=1}^{250}\hat{p}_n(v_i;\sigma)$$

$$= \prod_{i=1}^{250}\frac{1}{750\cdot\sqrt{\pi}}\sum_{j=1}^{250}\exp\left(-\frac{(v_i-x_j)^2}{2\sigma^2}\right)$$

Here $n = 750$, since it is the number of samples used to build the pdf.