Image Denoising

Homework n°3 - Exercises

Exercise 3.11

1) For any real number t, we have

$$\begin{aligned}\mathbb{P}(y \leq t) &= \mathbb{P}(ux \leq t)\\ &= \mathbb{P}(ux \leq t, u = 1) + \mathbb{P}(ux \leq t, u = -1)\\ &= \mathbb{P}(x \leq t, u = 1) + \mathbb{P}(-x \leq t, u = -1)\\ &= \mathbb{P}(u = 1)\mathbb{P}(x \leq t) + \mathbb{P}(u = -1)\mathbb{P}(-x \leq t) \text{ assuming } x \text{ and } u \text{ are independent}\\ &= \frac{1}{2}(\mathbb{P}(x \leq t) + \mathbb{P}(-x \leq t))\\ &= \mathbb{P}(x \leq t) \text{ because the law } \mathcal{N}(0, \sigma^2) \text{ of } x \text{ is symmetric.}\end{aligned}$$

This proves that x and y have the same law, or equivalently:

$$\boxed{y \text{ is a Gaussian variable with same density as } x}.$$

2) $\mathbb{E}[xy] = \mathbb{E}[x(ux)] = \mathbb{E}[ux^2] = \mathbb{E}[u]\mathbb{E}[x^2]$ because x and u are independent, and $\mathbb{E}[u] = 0$ so $\mathbb{E}[xy] = 0$.
And $\mathbb{E}[x]\mathbb{E}[y] = 0$ because $\mathbb{E}[x] = 0$.
So $\boxed{x \text{ and } y \text{ are uncorrelated}}$.

3) If we assume $x \sim \mathcal{N}(0, \sigma^2)$, then $\boxed{\mathbb{E}[x^2] = \sigma^2}$.
Using question 1), we then also have $\boxed{\mathbb{E}[y^2] = \sigma^2}$.
And $\mathbb{E}[x^2y^2] = \mathbb{E}[x^2(ux)^2] = \mathbb{E}[x^4 \underbrace{u^2}_{=1}] = \mathbb{E}[x^4]$, and it is known that $\mathbb{E}[\mathcal{N}(0, 1)^4] = 3$, so $\boxed{\mathbb{E}[x^2y^2] = 3\sigma^4}$.
Therefore $\mathbb{E}[x^2y^2] \neq \mathbb{E}[x^2]\mathbb{E}[y^2]$, so $\boxed{x \text{ and } y \text{ are not independent}}$.

Exercise 3.12

Let A be an $n \times m$ matrix, and B an $m \times n$ matrix.
We have

$$\begin{aligned}\mathrm{Tr}(AB) &= \sum_{i=1}^{n} (AB)_{i,i}\\ &= \sum_{i=1}^{n} \sum_{k=1}^{m} A_{i,k}B_{k,i} \text{ by definition of } AB\\ &= \sum_{k=1}^{m} \sum_{i=1}^{n} B_{k,i}A_{i,k} \text{ by permuting the sums}\\ &= \sum_{k=1}^{m} (BA)_{k,k} \text{ by definition of } BA\\ &= \mathrm{Tr}(BA).\end{aligned}$$

so $\boxed{\mathrm{Tr}(AB) = \mathrm{Tr}(BA)}$.

Exercise 3.13

For a matrix H, we have $\mathrm{Tr}((A+H)B) = \mathrm{Tr}(AB + HB) = \mathrm{Tr}(AB) + \mathrm{Tr}(HB)$.
And $\mathrm{Tr}(HB) = \mathrm{Tr}(H(B^T)^T) = \langle H, B^T \rangle$ using the scalar product $\langle M, N \rangle = \mathrm{Tr}(MN^T)$.

So we have shown that $\mathrm{Tr}((A+H)B) - \mathrm{Tr}(AB) = \langle H, B^T \rangle$, which shows that $\dfrac{\partial \mathrm{Tr}(AB)}{\partial A} = B^T$.
And since $\mathrm{Tr}(BA) = \mathrm{Tr}(AB)$ (using the previous exercise), we have $\boxed{\dfrac{\partial \mathrm{Tr}(AB)}{\partial A} = \dfrac{\partial \mathrm{Tr}(BA)}{\partial A} = B^T}$.

Exercise 3.14

From Terence Tao's blog, we have $\dfrac{\partial \det(A)}{\partial A} = \det(A) A^{-T}$.
This is because $\det(A+H) = \det(A)\det(I + A^{-1}H) = \det(A)(1 + \mathrm{Tr}(A^{-1}H) + O(\|H\|^2)) = \det(A) + \det(A)\langle A^{-T}, H \rangle + O(\|H\|^2)$.

As a consequence, $\dfrac{\partial \log\det(A)}{\partial A} = \dfrac{1}{\det(A)} \dfrac{\partial \det(A)}{\partial A} = \dfrac{1}{\det(A)} \det(A) A^{-T} = A^{-T}$.
We then have $\boxed{\dfrac{\partial \log\det(A)}{\partial A} = A^{-T}}$.

Exercise 8.1

We have $\mathbb{P}\left(\tilde{P} \mid P\right) = \frac{1}{(2\pi\sigma^2)^{\kappa/2}} \exp\left(-\frac{\|P - \tilde{P}\|^2}{2\sigma^2}\right)$, and P and the noise are independent.

We have $\tilde{P} = P + N$ with $N \sim \mathcal{N}(0, \sigma^2 I_{\kappa^2})$ white noise.

Then for each $1 \leq i, j \leq \kappa^2$, we have

$$\begin{aligned}
\mathrm{Cov}(\tilde{P}_i, \tilde{P}_j) &= \mathrm{Cov}(P_i + N_i, P_j + N_j) \\
&= \mathrm{Cov}(P_i, P_j) + \mathrm{Cov}(P_i, N_j) + \mathrm{Cov}(N_i, P_j) + \mathrm{Cov}(N_i, N_j) \\
&= \mathrm{Cov}(P_i, P_j) + 0 + 0 + \sigma^2 \delta_{i,j},
\end{aligned}$$

which can be rewritten $\boxed{\mathbf{C}_{\tilde{P}} = \mathbf{C}_P + \sigma^2 \mathbf{I}_{\kappa^2}}$.

And since $\mathbb{E}[N] = 0$, we also have $\boxed{\mathbb{E}[\tilde{P}] = \mathbb{E}[P]}$.

Exercise 8.3

We first consider

$$\hat{P}_1 = \overline{P} + \left[\mathbf{C}_{\tilde{P}} - \sigma^2 \mathbf{I}\right] \mathbf{C}_{\tilde{P}}^{-1} \left(\tilde{P} - \overline{P}\right). \tag{1}$$

Let $(G_i)_i$ be an orthonormal basis constituted of eigenvectors of $\mathbf{C}_{\tilde{P}}$ with eigenvalues $\lambda_i > 0$ ($\mathbf{C}_{\tilde{P}}$ is a symmetric positive definite matrix).
Then $(G_i)_i$ is also a family of eigenvectors of $\mathbf{C}_{\tilde{P}}^{-1}$ with eigenvalues $1/\lambda_i > 0$.

We can now rewrite (1):

$$\begin{aligned}
\hat{P}_1 &= \overline{P} + \left[\mathbf{C}_{\tilde{P}} - \sigma^2 \mathbf{I}\right] \mathbf{C}_{\tilde{P}}^{-1} \left(\tilde{P} - \overline{P}\right) \\
&= \overline{P} + \left[\mathbf{I} - \sigma^2 \mathbf{C}_{\tilde{P}}^{-1}\right] \left(\tilde{P} - \overline{P}\right) \\
&= \sum_i \left\langle \overline{P} + \left[\mathbf{I} - \sigma^2 \mathbf{C}_{\tilde{P}}^{-1}\right] \left(\tilde{P} - \overline{P}\right), G_i \right\rangle G_i \text{ because } (G_i)_i \text{ is an orthonormal basis} \\
&= \sum_i \left(\langle \overline{P}, G_i \rangle + \left\langle \left[\mathbf{I} - \sigma^2 \mathbf{C}_{\tilde{P}}^{-1}\right] \left(\tilde{P} - \overline{P}\right), G_i \right\rangle \right) G_i \\
&= \sum_i \left(\langle \overline{P}, G_i \rangle + \left\langle \left(\tilde{P} - \overline{P}\right), \left[\mathbf{I} - \sigma^2 \mathbf{C}_{\tilde{P}}^{-1}\right]^T G_i \right\rangle \right) G_i \\
&= \sum_i \left(\langle \overline{P}, G_i \rangle + \left(1 - \frac{\sigma^2}{\lambda_i}\right) \left\langle \tilde{P} - \overline{P}, G_i \right\rangle \right) G_i \\
&= \sum_i \left(\frac{\sigma^2}{\lambda_i} \langle \overline{P}, G_i \rangle + \left(1 - \frac{\sigma^2}{\lambda_i}\right) \left\langle \tilde{P}, G_i \right\rangle \right) G_i,
\end{aligned}$$

so

$$\boxed{\hat{P}_1 = \sum_i \frac{\sigma^2}{\lambda_i} \langle \overline{P}, G_i \rangle G_i + \sum_i \left(1 - \frac{\sigma^2}{\lambda_i}\right) \left\langle \tilde{P}, G_i \right\rangle G_i.}$$

The second term in the above expression can be written $\sum_i a(i) \left\langle \tilde{P}, G_i \right\rangle G_i$, where $a(i) = 1 - \frac{\sigma^2}{\lambda_i}$. This is what is used in Wiener filtering.

As for the other formula:

$$\hat{P}_2 = \overline{P}^1 + \mathbf{C}_P^1 \left[\mathbf{C}_P^1 + \sigma^2 \mathbf{I}\right]^{-1} \left(\tilde{P} - \overline{P}^1\right), \tag{2}$$

let $(H_i)_i$ be an orthonormal basis constituted of eigenvectors of $\mathbf{C}_P^1$ with eigenvalues $\mu_i > 0$ ($\mathbf{C}_P^1$ is a symmetric positive definite matrix).

Then we can rewrite (2):

$$\begin{aligned}
\hat{P}_2 &= \overline{P}^1 + \mathbf{C}_P^1 \left[\mathbf{C}_P^1 + \sigma^2 \mathbf{I}\right]^{-1} \left(\tilde{P} - \overline{P}^1\right) \\
&= \overline{P}^1 + \left[\mathbf{I} + \sigma^2 (\mathbf{C}_P^1)^{-1}\right]^{-1} \left(\tilde{P} - \overline{P}^1\right) \\
&= \sum_i \langle \overline{P}^1 + \left[\mathbf{I} + \sigma^2 (\mathbf{C}_P^1)^{-1}\right]^{-1} \left(\tilde{P} - \overline{P}^1\right), H_i \rangle H_i \\
&= \sum_i \left(\langle \overline{P}^1, H_i \rangle + \langle \left[\mathbf{I} + \sigma^2 (\mathbf{C}_P^1)^{-1}\right]^{-1} \left(\tilde{P} - \overline{P}^1\right), H_i \rangle \right) H_i \\
&= \sum_i \left(\langle \overline{P}^1, H_i \rangle + \langle \tilde{P} - \overline{P}^1, \left[\mathbf{I} + \sigma^2 (\mathbf{C}_P^1)^{-1}\right]^{-1} H_i \rangle \right) H_i \\
&= \sum_i \left(\langle \overline{P}^1, H_i \rangle + \frac{1}{1 + \sigma^2/\mu_i} \langle \tilde{P} - \overline{P}^1, H_i \rangle \right) H_i \\
&= \sum_i \left(\left(1 - \frac{1}{1 + \sigma^2/\mu_i}\right) \langle \overline{P}^1, H_i \rangle + \frac{1}{1 + \sigma^2/\mu_i} \langle \tilde{P}, H_i \rangle \right) H_i
\end{aligned}$$

so

$$\boxed{\hat{P}_2 = \sum_i \frac{\sigma^2}{\sigma^2 + \mu_i} \langle \overline{P}^1, H_i \rangle H_i + \sum_i \frac{\mu_i}{\sigma^2 + \mu_i} \langle \tilde{P}, H_i \rangle H_i}$$

The second term in the above expression can be written $\sum_i b(i) \left\langle \tilde{P}, H_i \right\rangle H_i$, where $b(i) = \frac{\mu_i}{\sigma^2 + \mu_i}$. This is once again what is used in Wiener filtering.

So in both cases, we use a Wiener filtering. However, we also add either $\sum_i \frac{\sigma^2}{\lambda_i} \left\langle \overline{P}, G_i \right\rangle G_i$ or $\sum_i \frac{\sigma^2}{\sigma^2 + \mu_i} \langle \overline{P}^1, H_i \rangle H_i$. The idea is that we mostly use the components of $\overline{P}$ and $\overline{P}^1$ corresponding to small eigenvalues λ_i or μ_i.

Exercise 8.4

We rewrite:

$$\text{MSE} = \int \mathbb{P}(P) \int \mathbb{P}(\tilde{P} \mid P) \|P - \hat{P}\|^2 d\tilde{P} dP. \tag{8.16}$$

Using (8.16), we have:

$$\begin{aligned}
\text{MSE} &= \int \mathbb{P}(P) \int \mathbb{P}(\tilde{P} \mid P) \|P - \hat{P}\|^2 d\tilde{P} dP \\
&= \int \int \underbrace{\mathbb{P}(P)\mathbb{P}(\tilde{P} \mid P)}_{=\mathbb{P}(\tilde{P})\mathbb{P}(P|\tilde{P})} \|P - \hat{P}\|^2 d\tilde{P} dP \\
&= \int \int \mathbb{P}(\tilde{P})\mathbb{P}(P \mid \tilde{P}) \|P - \hat{P}\|^2 d\tilde{P} dP.
\end{aligned}$$

Then, we apply Fubini-Tonelli's theorem, by verifying that for each P, $\int \mathbb{P}(\tilde{P})\mathbb{P}(P \mid \tilde{P}) \|P - \hat{P}\|^2 d\tilde{P} < +\infty$ and for each $\tilde{P}$, $\int \mathbb{P}(\tilde{P})\mathbb{P}(P \mid \tilde{P}) \|P - \hat{P}\|^2 d\tilde{P} < +\infty$.

In fact, this theorem also works if the functions are not integrable, in the case where they are non-negative, which is the case here.

We then have:

$$\begin{aligned}
\text{MSE} &= \int \int \mathbb{P}(\tilde{P})\mathbb{P}(P \mid \tilde{P}) \|P - \hat{P}\|^2 dP d\tilde{P} \\
&= \int \mathbb{P}(\tilde{P}) \int \mathbb{P}(P \mid \tilde{P}) \|P - \hat{P}\|^2 dP d\tilde{P}.
\end{aligned}$$

We have shown:

$$\boxed{\text{MSE} = \int \mathbb{P}(\tilde{P}) \int \mathbb{P}(P \mid \tilde{P}) \|P - \hat{P}\|^2 dP d\tilde{P}}.$$

Exercise 8.5

$$\mathrm{MMSE}(\tilde{P}) = \int \mathbb{P}(P \mid \tilde{P})(P - \hat{P})^2 \, dP. \tag{8.18}$$

We can differentiate (8.18) with respect to $\hat{P}$:

$$\begin{aligned}
\frac{\partial \mathrm{MSSE}}{\partial \hat{P}}(\hat{P}) &= 2 \int \mathbb{P}(P \mid \tilde{P})(\hat{P} - P) \, dP \\
&= 2\hat{P} \int \mathbb{P}(P \mid \tilde{P}) \, dP - 2 \int \mathbb{P}(P \mid \tilde{P}) P \, dP.
\end{aligned}$$

Setting this derivative to 0 gives

$$\hat{P} = \frac{\int \mathbb{P}(P \mid \tilde{P}) P \, dP}{\int \mathbb{P}(P \mid \tilde{P}) \, dP} = \frac{\int \mathbb{P}(P \mid \tilde{P}) P \, dP}{1} = \int \mathbb{P}(P \mid \tilde{P}) P \, dP = \mathbb{E}\left[P \mid \tilde{P}\right].$$

We have shown that we get the MSSE estimator (8.15):

$$\boxed{\hat{P} = \mathbb{E}\left[P \mid \tilde{P}\right]}.$$