Inverse filtering for image restoration

- Inverse filtering is a deterministic and direct method for image restoration.

- The images involved must be lexicographically ordered. That means that an image is converted to a column vector by stacking the rows one by one after converting them to columns.

- Therefore, an image of size $M \times N = 256 \times 256$ is converted to a column vector of size $(256 \times 256) \times 1 = 65536 \times 1$.

- The degradation model is written in a matrix form as
$$\mathbf{y} = \mathbf{Hf}$$
where the images are vectors and the degradation process is a huge but sparse matrix of size $MN \times MN$.

- The above relationship is ideal. The true degradation model is $\mathbf{y} = \mathbf{Hf} + \mathbf{n}$ where $\mathbf{n}$ is a lexicographically ordered two dimensional noisy signal which corrupts the distorted image $y(x, y)$.

Inverse Filtering for image restoration

- We formulate an unconstrained optimisation problem as follows:

$$\underset{\mathbf{f}}{\text{minimise}}\, J(\mathbf{f}) = \|\mathbf{n}(\mathbf{f})\|^2 = \|\mathbf{y}-\mathbf{Hf}\|^2$$

$$\|\mathbf{y}-\mathbf{Hf}\|^2 = (\mathbf{y}-\mathbf{Hf})^T(\mathbf{y}-\mathbf{Hf}) = [\mathbf{y}^T-(\mathbf{Hf})^T](\mathbf{y}-\mathbf{Hf})$$
$$= (\mathbf{y}^T-\mathbf{f}^T\mathbf{H}^T)(\mathbf{y}-\mathbf{Hf}) = \mathbf{y}^T\mathbf{y}-\mathbf{y}^T\mathbf{Hf}-\mathbf{f}^T\mathbf{H}^T\mathbf{y}+\mathbf{f}^T\mathbf{H}^T\mathbf{Hf}$$

- We set the first derivative of $J(\mathbf{f})$ equal to 0.

$$\frac{\partial J(\mathbf{f})}{\partial \mathbf{f}} = \mathbf{0} \Rightarrow \frac{\partial \mathbf{y}^T\mathbf{y}}{\partial \mathbf{f}} - \frac{\partial \mathbf{y}^T\mathbf{Hf}}{\partial \mathbf{f}} - \frac{\partial \mathbf{f}^T\mathbf{H}^T\mathbf{y}}{\partial \mathbf{f}} + \frac{\partial \mathbf{f}^T\mathbf{H}^T\mathbf{Hf}}{\partial \mathbf{f}} = \mathbf{0}$$

Note that:

- $\frac{\partial(\cdot)}{\partial \mathbf{f}}$ indicates a vector of partial derivatives
- $\frac{\partial \mathbf{y}^T\mathbf{y}}{\partial \mathbf{f}} = 0$
- $\frac{\partial \mathbf{y}^T\mathbf{Hf}}{\partial \mathbf{f}} = \frac{\partial \mathbf{f}^T\mathbf{H}^T\mathbf{y}}{\partial \mathbf{f}}$

- Therefore,

$$\frac{\partial J(\mathbf{f})}{\partial \mathbf{f}} = \mathbf{0} \Rightarrow -2\frac{\partial \mathbf{f}^T\mathbf{H}^T\mathbf{y}}{\partial \mathbf{f}} + \frac{\partial \mathbf{f}^T\mathbf{H}^T\mathbf{Hf}}{\partial \mathbf{f}} = \mathbf{0} \Rightarrow -2\mathbf{H}^T\mathbf{y} + 2\mathbf{H}^T\mathbf{Hf} = 0 \Rightarrow$$

$$\mathbf{H}^T\mathbf{Hf} = \mathbf{H}^T y$$

- If the matrix $\mathbf{H}^T\mathbf{H}$ is invertible then $\mathbf{f} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T y$
- If $\mathbf{H}$ is square and invertible then $\mathbf{f} = \mathbf{H}^{-1}(\mathbf{H}^T)^{-1}\mathbf{H}^T y = \mathbf{H}^{-1} y$

Inverse Filtering for image restoration

- According to the previous analysis if $\mathbf{H}$ (and therefore $\mathbf{H}^{-1}$) is block circulant the above problem can be solved as a set of $M \times N$ scalar problems as follows.

$$F(u,v) = \frac{H^*(u,v)Y(u,v)}{|H(u,v)|^2} \Rightarrow f(i,j) = \mathfrak{I}^{-1}\left[\frac{H^*(u,v)Y(u,v)}{|H(u,v)|^2}\right] = \frac{Y(u,v)}{H(u,v)}$$

- Diagonalisation of all matrices involved is necessary.

Computational issues concerning inverse filtering
Noise free case

- Suppose first that the additive noise $n(i,j)$ is negligible. A problem arises if $H(u,v)$ becomes very small or zero for some point (u,v) or for a whole region in the (u,v) plane. In that region inverse filtering cannot be applied.
- Note that in most real applications $H(u,v)$ drops off rapidly as a function of distance from the origin !

  Solution: if these points are known they can be neglected in the computation of $F(u,v)$.

Imperial College London

Computational issues concerning inverse filtering
Noisy case

- In the presence of external noise we have that

$$\hat{F}(u,v) = \frac{H^*(u,v)(Y(u,v) - N(u,v))}{|H(u,v)|^2} =$$

$$\frac{H^*(u,v)Y(u,v)}{|H(u,v)|^2} - \frac{H^*(u,v)N(u,v)}{|H(u,v)|^2} \Rightarrow$$

$$\hat{F}(u,v) = F(u,v) - \frac{N(u,v)}{H(u,v)}$$

- If $H(u,v)$ becomes very small, the term $N(u,v)$ dominates the result.

Psedoinverse Filtering

- To cope with noise amplification we carry out the restoration process in a limited neighborhood about the origin where $H(u,v)$ is not very small.
- This procedure is called **pseudoinverse or generalized inverse filtering**.
- In that case we set

$$\hat{F}(u,v) = \begin{cases} \dfrac{H^*(u,v)Y(u,v)}{|H(u,v)|^2} & H(u,v) \neq 0 \\ 0 & H(u,v) = 0 \end{cases}$$

or

$$\hat{F}(u,v) = \begin{cases} \dfrac{H^*(u,v)Y(u,v)}{|H(u,v)|^2} = \dfrac{Y(u,v)}{H(u,v)} & |H(u,v)| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

Psedoinverse restoration examples

Figure 3: Degraded by a 7×7 pill-box blur, 20 dB BSNR

Figure 5: Degraded by a 5×5 Gaussian blur ($\sigma^2 = 1$), 20 dB BSNR

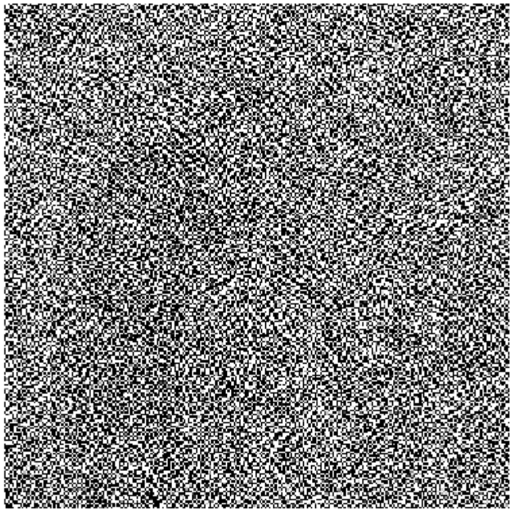


Figure 11: Result of Figure 3 restored by a generalized inverse filter with a threshold of 10^{-3}, ISNR $= -32.9$ dB

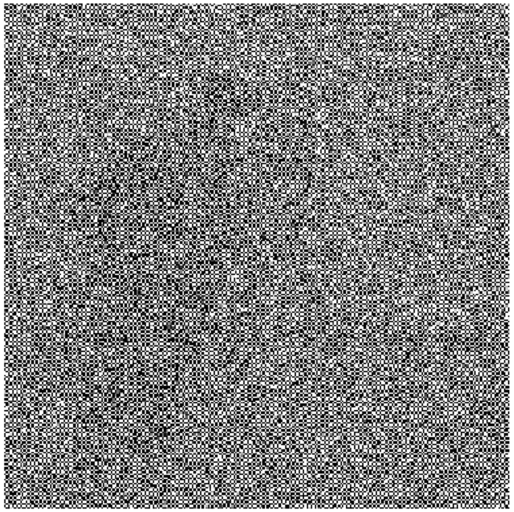


Figure 17: Result of Figure 5 restored by a generalized inverse filter with a threshold of 10^{-3}, ISNR $= -36.6$ dB

Psedoinverse restoration examples

Figure 3: Degraded by a 7×7 pill-box blur, 20 dB BSNR

Figure 5: Degraded by a 5×5 Gaussian blur ($\sigma^2 = 1$), 20 dB BSNR

Figure 13: Result of Figure 3 restored by a generalized inverse filter with a threshold of 10^{-1}, ISNR $= 0.61$ dB

Figure 19: Result of Figure 5 restored by a generalized inverse filter with a threshold of 10^{-1}, ISNR $= -1.8$ dB

Constrained Least Squares (CLS) Restoration

- By introducing a so called **Lagrange multiplier** or **regularisation parameter** α, we transform the constrained optimisation problem to an unconstrained one as follows.

- The problem

$$\underset{\mathbf{f}}{\textbf{minimise}}\, J(\mathbf{f}) = \|\mathbf{y}-\mathbf{H}\mathbf{f}\|^2$$
$$\textbf{subject to } \|\mathbf{C}\mathbf{f}\|^2 < \varepsilon$$

  is equivalent to

$$\underset{\mathbf{f}}{\textbf{minimise}}\, J(\mathbf{f}) = \|\mathbf{y}-\mathbf{H}\mathbf{f}\|^2 + \alpha\|\mathbf{C}\mathbf{f}\|^2$$

- The imposed constraint implies that the energy of the restored image at high frequencies is below a threshold.

- It is basically a smoothness constraint.
  C a high pass filter operator
  Cf a high pass filtered version of the image

Constrained Least Squares (CLS) Restoration

- We formulate an unconstrained optimisation problem as follows:

$$\underset{\mathbf{f}}{\text{minimise}}\, J(\mathbf{f}) = \|\mathbf{y}-\mathbf{Hf}\|^2 + \alpha\,\|\mathbf{Cf}\|^2$$

$$\|\mathbf{y}-\mathbf{Hf}\|^2 + \alpha\,\|\mathbf{Cf}\|^2 = (\mathbf{y}-\mathbf{Hf})^T(\mathbf{y}-\mathbf{Hf}) + \alpha(\mathbf{Cf})^T(\mathbf{Cf})$$
$$= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{Hf} - \mathbf{f}^T\mathbf{H}^T\mathbf{y} + \mathbf{f}^T\mathbf{H}^T\mathbf{Hf} + \alpha\mathbf{f}^T\mathbf{C}^T\mathbf{Cf}$$

- We set the first derivative of $J(\mathbf{f})$ equal to 0.

$$\frac{\partial J(\mathbf{f})}{\partial \mathbf{f}} = \mathbf{0} \Rightarrow -2\mathbf{H}^T\mathbf{y} + 2\mathbf{H}^T\mathbf{Hf} + 2\alpha\mathbf{C}^T\mathbf{Cf} = \mathbf{0}$$

- Therefore,

$$\frac{\partial J(\mathbf{f})}{\partial \mathbf{f}} = \mathbf{0} \Rightarrow -2\frac{\partial \mathbf{f}^T\mathbf{H}^T\mathbf{y}}{\partial \mathbf{f}} + \frac{\partial \mathbf{f}^T\mathbf{H}^T\mathbf{Hf}}{\partial \mathbf{f}} = \mathbf{0} \Rightarrow -2\mathbf{H}^T\mathbf{y} + 2\mathbf{H}^T\mathbf{Hf} = 0 \Rightarrow$$

$$(\mathbf{H}^T\mathbf{H} + \alpha\mathbf{C}^T\mathbf{C})\mathbf{f} = \mathbf{H}^T y$$

- In frequency domain and under the presence of noise we have:

$$\hat{F}(u,v) = \frac{H^*(u,v)(Y(u,v) - N(u,v))}{|H(u,v)|^2 + \alpha|C(u,v)|^2}$$

Constrained Least Squares (CLS) Restoration: Observations

- In frequency domain and under the presence of noise we have:

$$\hat{F}(u,v) = \frac{H^*(u,v)(Y(u,v) - N(u,v))}{|H(u,v)|^2 + \alpha|C(u,v)|^2}$$

- The regularisation parameter α controls the contribution between the terms $||\mathbf{y} - \mathbf{Hf}||^2$ and $||\mathbf{Cf}||^2$.
- Small α implies that emphasis is given to the minimisation function $||\mathbf{y} - \mathbf{Hf}||^2$.
  - Note that in the extreme case where $\alpha = 0$, CLS becomes Inverse Filtering.
  - Note that with smaller values of α, the restored image tends to have more amplified noise effects.
- Large α implies that emphasis is given to the minimisation function $||\mathbf{Cf}||^2$. A large a should be chosen if the noise is high.
  - Note that with larger values of α, and thus more regularisation, the restored image tends to have more ringing.

Choice of α cont.

- The problem of the choice of α has been attempted in a large number of studies and different techniques have been proposed.
- One possible choice is based on a **set theoretic approach.**
  - A restored image is approximated by an image which lies in the intersection of the two ellipsoids defined by

    $Q_{f|y} = \{f \mid \|y - Hf\|^2 \leq E^2\}$ and $Q_f = \{f \mid \|Cf\|^2 \leq \varepsilon^2\}$
  - The center of one of the ellipsoids which bounds the intersection of $Q_{f|y}$ and Q_f, is given by the equation

    $$f = \left(H^T H + \alpha C^T C\right)^{-1} H^T y$$

    with $\alpha = (E/\varepsilon)^2$.
- Finally, a choice for α is also:

$$\alpha = \frac{1}{\text{BSNR}}$$

Choice of α

- The variance and bias of the error image in frequency domain are

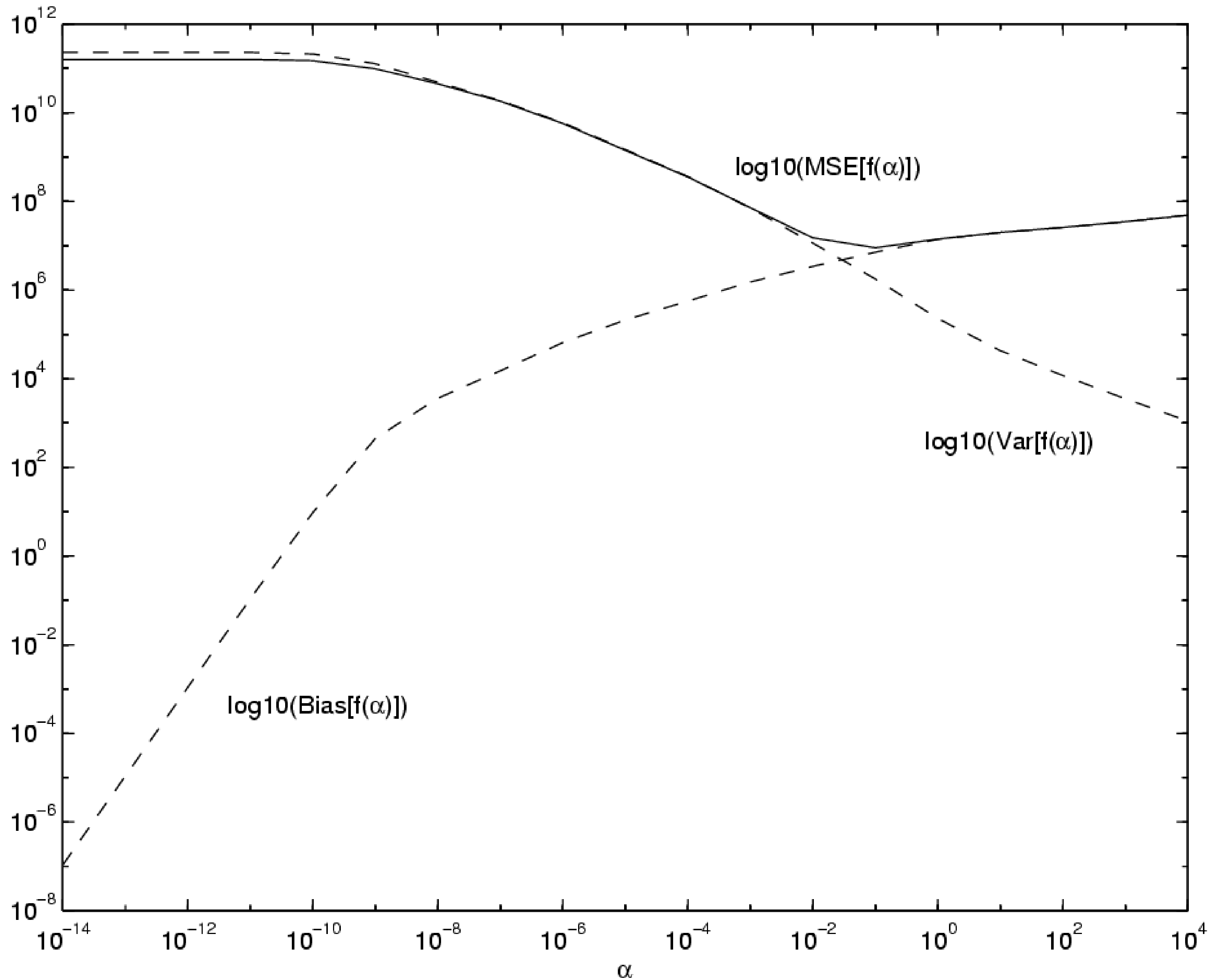
$$\mathrm{Var}(\hat{f}(a)) = \sigma_n^2 \sum_{u=0}^{M} \sum_{v=0}^{N} \frac{|H(u,v)|^2}{(|H(u,v)|^2 + \alpha|C(u,v)|^2)^2}$$

$$\mathrm{Bias}(\hat{f}(a)) = \sigma_n^2 \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \frac{|F(u,v)|^2 \alpha^2 |C(u,v)|^4}{(|H(u,v)|^2 + \alpha|C(u,v)|^2)^2}$$

- Note that the Mean Squared Error (MSE) $E(\alpha)$ in this problem is the expected value of the Euclidian norm of the difference between the true original image f and the estimated original image $\hat{f}(a)$, i.e., $E\{\|\hat{f}(a) - f\|^2\}$.
- It has been shown that the minimum Mean Squared Error (solid curve in next slide) is encountered close to the intersection of the above functions and is equal to

$$E\{\|\hat{f}(a) - f\|^2\} = \mathrm{Bias}(\hat{f}(a) + \mathrm{Var}(\hat{f}(a))$$

- Observing the graphs for the variance and bias of the error in the next slide we can say that another good choice of α is one that gives the best compromise between the variance and bias of the error image.



log10(MSE[f(α)])
log10(Var[f(α)])
log10(Bias[f(α)])
10^{12}
10^{10}
10^{8}
10^{6}
10^{4}
10^{2}
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-14}
10^{-12}
10^{-10}
10^{-8}
10^{-6}
10^{-4}
10^{-2}
10^{0}
10^{2}
10^{4}
α

Figure 3: Degraded by a 7×7 pill-box blur, 20 dB BSNR

Figure 5: Degraded by a 5×5 Gaussian blur ($\sigma^2 = 1$), 20 dB BSNR

Figure 26: CLS restoration of Figure 3 with $\alpha = 1$, ISNR $= 2.5$ dB

Figure 40: CLS restoration of Figure 5 with $\alpha = 1$, ISNR $= 1.3$ dB

Figure 3: Degraded by a 7×7 pill-box blur, 20 dB BSNR

Figure 5: Degraded by a 5×5 Gaussian blur ($\sigma^2 = 1$), 20 dB BSNR

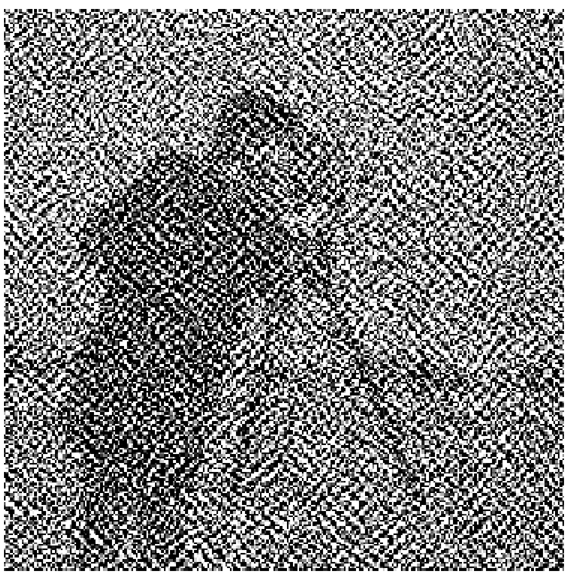


Figure 30: CLS restoration of Figure 3 with $\alpha = 0.0001$, ISNR = -21.9 dB

e 44: CLS restoration of Figure 5 with $\alpha = 0.0001$, ISNR = -22.1 dB

Figure 3: Degraded by a 7×7 pill-box blur, 20 dB BSNR

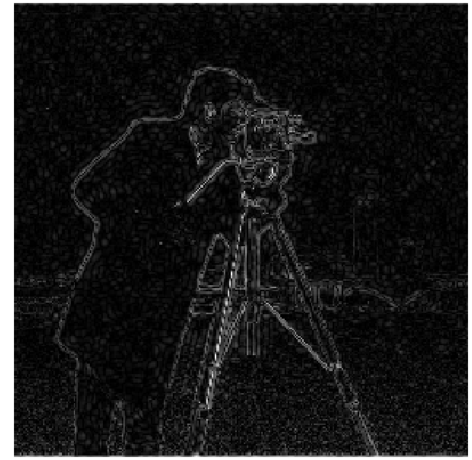


Figure 27: Corresponding error image for Figure 26 ($|\text{original} - \text{restored}|$, scaled for display)

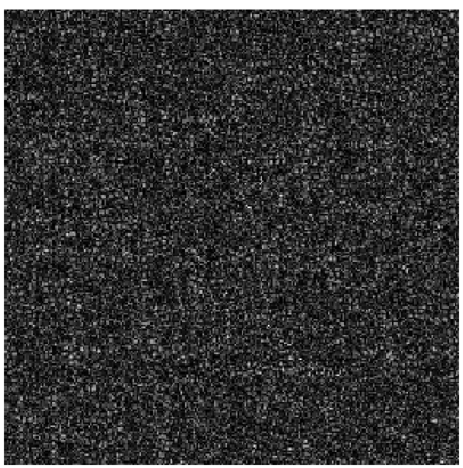


Figure 29: Corresponding error image for Figure 28 ($|\text{original} - \text{restored}|$, scaled for display)

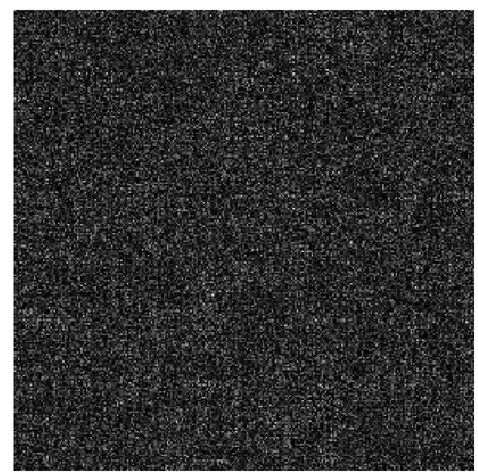
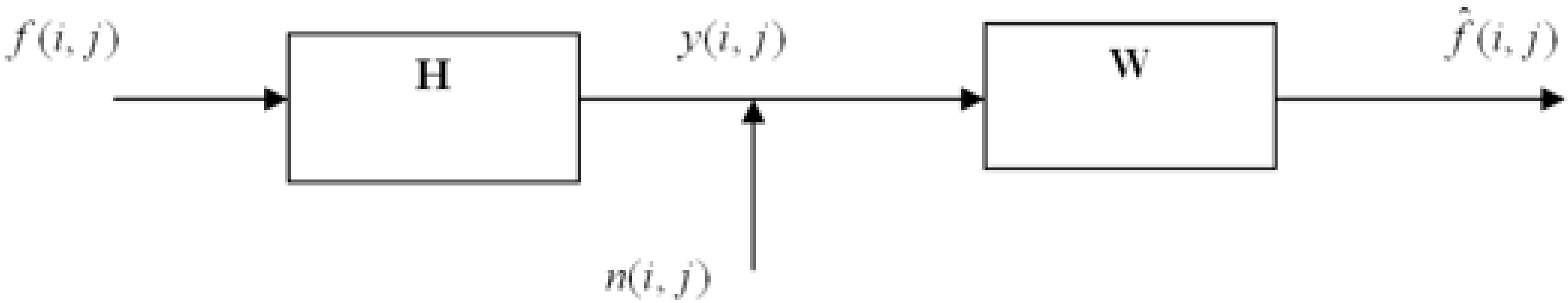


Figure 31: Corresponding error image for Figure 30 ($|\text{original} - \text{restored}|$, scaled for display)

Wiener Filter Estimator (Stochastic Regularisation)

- The image restoration problem can be viewed as a system identification problem as follows:



- The objective is to minimize the expected value of the Euclidian norm of the error:

$$E\{(f-\hat{f})^T(f-\hat{f})\}$$

To do so the following conditions should hold:

i. $E\{\hat{f}\} = E\{f\} \Rightarrow E\{f\} = WE\{y\}$

ii. The error must be orthogonal to the observation about the mean
$E\{(\hat{f}-f)(y-E\{y\})^T\} = 0$

Wiener Filter Estimator (Stochastic Regularisation)

The following conditions should hold:

i. $E\{\hat{f}\} = E\{f\} \Rightarrow E\{f\} = WE\{y\}$

ii. The error must be orthogonal to the observation about the mean
$E\{(\hat{f} - f)(y - E\{y\})^T\} = 0$

From i. and ii. we have that

$$E\{(Wy - f)(y - E\{y\})^T\} = 0 \Rightarrow E\{(Wy + E\{f\} - WE\{y\} - f)(y - E\{y\})^T\} = 0$$
$$\Rightarrow E\{[W(y - E\{y\}) - (f - E\{f\})](y - E\{y\})^T\} = 0$$

If $\tilde{y} = y - E\{y\}$ and $\tilde{f} = f - E\{f\}$ then
$$E\{(W\tilde{y} - \tilde{f})\tilde{y}^T\} = 0 \Rightarrow E\{W\tilde{y}\tilde{y}^T\} = E\{\tilde{f}\tilde{y}^T\} \Rightarrow WE\{\tilde{y}\tilde{y}^T\} = E\{\tilde{f}\tilde{y}^T\} \Rightarrow WR_{\tilde{y}\tilde{y}}$$

Wiener Filter Estimator (Stochastic Regularisation)

- If the original and the degraded image are both zero mean then
$$R_{\tilde{y}\tilde{y}} = R_{yy} \text{ and } R_{\tilde{f}\tilde{y}} = R_{fy}$$
In that case we have that $WR_{yy} = R_{fy}$.

- If we go back to the degradation model and find the autocorrelation matrix of the degraded image then we get that
$$y = Hf + n \Rightarrow y^T = f^T H^T + n^T$$
$$E\{yy^T\} = HR_{ff}H^T + R_{nn} = R_{yy}$$
$$E\{fy^T\} = R_{ff}H^T = R_{fy}$$

- From the above we get the following result
$$W = R_{fy}R_{yy}^{-1} = R_{ff}H^T(HR_{ff}H^T + R_{nn})^{-1}$$
and the estimate for the original image is
$$\hat{f} = R_{ff}H^T(HR_{ff}H^T + R_{nn})^{-1}y$$

- Note that knowledge of R_{ff} and R_{nn} is assumed.

Imperial College London

Wiener Filter Estimator (Stochastic Regularisation)

In frequency domain

$$W(u,v) = \frac{S_{ff}(u,v)H^*(u,v)}{S_{ff}(u,v)|H(u,v)|^2 + S_{nn}(u,v)}$$

$$\hat{F}(u,v) = \frac{S_{ff}(u,v)H^*(u,v)}{S_{ff}(u,v)|H(u,v)|^2 + S_{nn}(u,v)} Y(u,v)$$

- $S_{ff}(u,v) = |F(u,v)|^2$ is the Power Spectral Density of $f(i,j)$
- $S_{nn}(u,v) = |N(u,v)|^2$ is the Power Spectral Density of $n(i,j)$

Computational issues

- The noise variance has to be known, otherwise it is estimated from a flat region of the observed image.
- In practical cases where a single copy of the degraded image is available, it is quite common to use $S_{yy}(u,v)$ as an estimate of $S_{ff}(u,v)$. **This is very often a poor estimate.**

Wiener Smoothing Filter

In the absence of any blur, $H(u,v) = 1$ and

$$W(u,v) = \frac{S_{ff}(u,v)}{S_{ff}(u,v) + S_{nn}(u,v)} = \frac{(SNR)}{(SNR) + 1}$$

- $(SNR) >> 1 \Rightarrow W(u,v) \cong 1$
- $(SNR) << 1 \Rightarrow W(u,v) \cong (SNR)$

(SNR) is high in low spatial frequencies and low in high spatial frequencies so $W(u,v)$ can be implemented with a lowpass (smoothing) filter.

Relation with Inverse Filtering

If $S_{nn}(u,v) = 0 \Rightarrow W(u,v) = \frac{1}{H(u,v)}$ which is the inverse filter

If $S_{nn}(u,v) \to 0$

$$\lim_{S_{nn}\to 0} W(u,v) = \begin{cases} \dfrac{1}{H(u,v)} & H(u,v) \neq 0 \\ 0 & H(u,v) = 0 \end{cases}$$

which is the pseudoinverse filter.