Imperial College London

Digital Image Processing

Image Restoration

DR TANIA STATHAKI

READER (ASSOCIATE PROFESSOR) IN SIGNAL PROCESSING

IMPERIAL COLLEGE LONDON

What is Image Restoration?

Image Restoration refers to a class of methods that aim at reducing or removing various types of distortions of an image under consideration. These can be:

- Distortion due to sensor noise
- Out-of-focus camera
- Motion blur
- Weather conditions
- Scratches, holes, cracks caused by aging of the image
- Others




Classification of restoration methods

Image restoration methods can be classified according to the type and amount of information related to the images involved in the problem and also type and amount of information related to the distortion.

- **Deterministic** or **stochastic** methods.
  - In deterministic methods, we work directly with the image values in either space or frequency domain.
  - In stochastic methods, we work with the statistical properties of the image of interest (autocorrelation function, covariance function, variance, mean etc.)

- **Non-blind** or **semi-blind** or **blind** methods.
  - In non-blind methods the degradation process is known. This is a typical so-called **Inverse Problem**.
  - In semi-blind methods the degradation process is partly-known.
  - In blind methods the degradation process is unknown.

Classification of implementation types of restoration methods

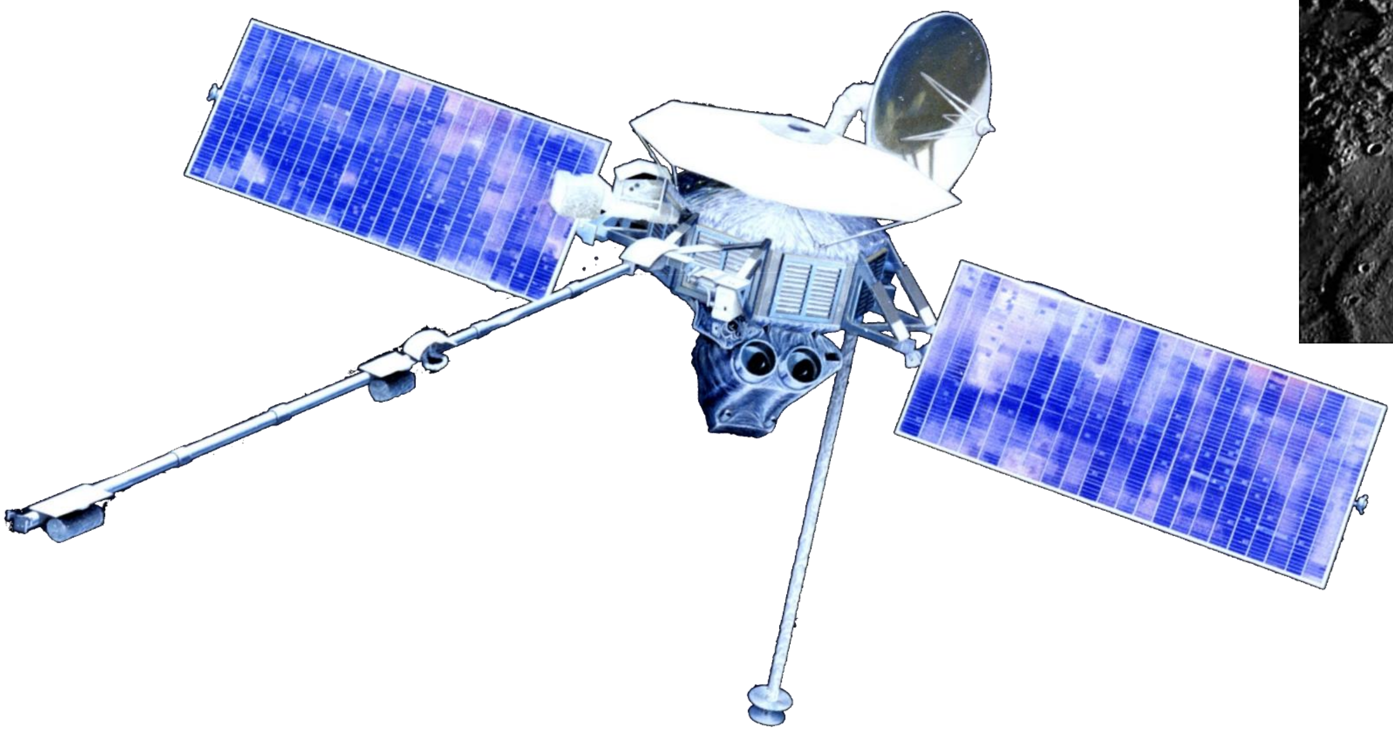
- **Direct** methods. The signals we are looking for (original undistorted image and degradation model) are obtained through a single, closed-form expression.

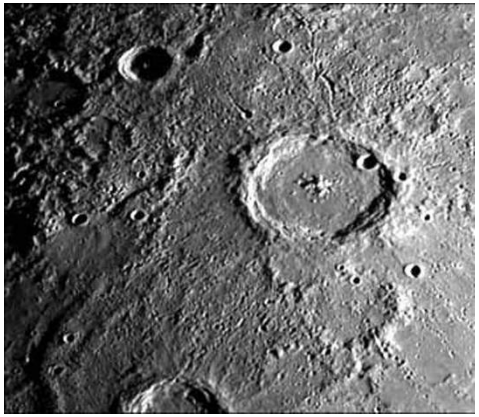
- **Iterative** methods. The signals we are looking for are obtained through a mathematical procedure that generates a sequence of gradually improving approximate solutions to the problem.

Imperial College London

Historical notes

- US and former Soviet Union space programs in 1950s and 1960s.
  - The 22 images produced during the Mariner IV flight to Mars in 1964 cost $10M.
  - These were very valuable images which underwent various restoration techniques.





Historical notes cont.

- Image Restoration is an active area of R&D.
  - Activity driven by mishaps (i.e., Humble Space Telescope - HSP).
  - New applications (i.e., SR of images and videos for Ultra High Definition - UHD displays), and new mathematical developments (i.e., sparsity).

- Notoriety in the media
  - “JFK” – Zapruder 8mm film underwent a number of restorations.
  - “No Way Out”, 1987, “Rising Sun”, 1993. These films’ whole plot largely relies on the successful restoration of some surveillance video.

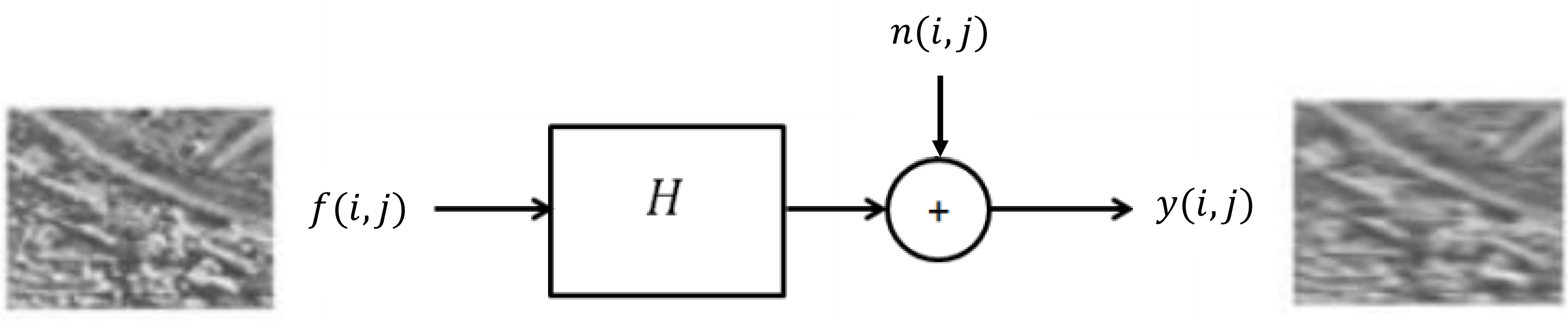
A generic model of an image restoration system

- A generic common model for an image restoration system is given by the following mathematical equation:

$$y(i,j) = H[f(i,j)] + n(i,j)$$

where:

  - (i,j) are the space coordinates
  - $f(i,j)$ is the original (undistorted) image
  - $H[\cdot]$ is a generic representation of the degradation function which is imposed onto the original image
  - $n(i,j)$ is a noise signal added to the distorted image



Linear and space-invariant (LSI) degradation model

In the degradation model:

$$y(i,j) = H[f(i,j)] + n(i,j)$$

we are interested in the definitions of linearity and space-invariance.

- The degradation model is **linear** if
  $H[k_1 f_1(i,j) + k_2\, f_2(i,j)] = k_1 H[f_1(i,j)] + k_2 H[f_2(i,j)]$
- The degradation model is **space-** or **position-invariant** if
  $H[f(i-i_0, j-j_0)] = y(i-i_0, j-j_0)$
- In the above definitions we ignore the presence of external noise.
- In real life scenarios, various types of degradations can be approximated by linear, space-invariant operators.

Advantages and drawbacks of LSI assumptions

Advantages

- It is much easier to deal with linear and space-invariant models because mathematics are easier.
- The distorted image is the convolution of the original image and the distortion system's function.
- Software tools are available.

Drawbacks

For various realistic types of image degradations, assumptions for linearity and space-invariance are too strict and significantly deviate from the true degradation model.

Linear and space invariant (LSI) degradation model

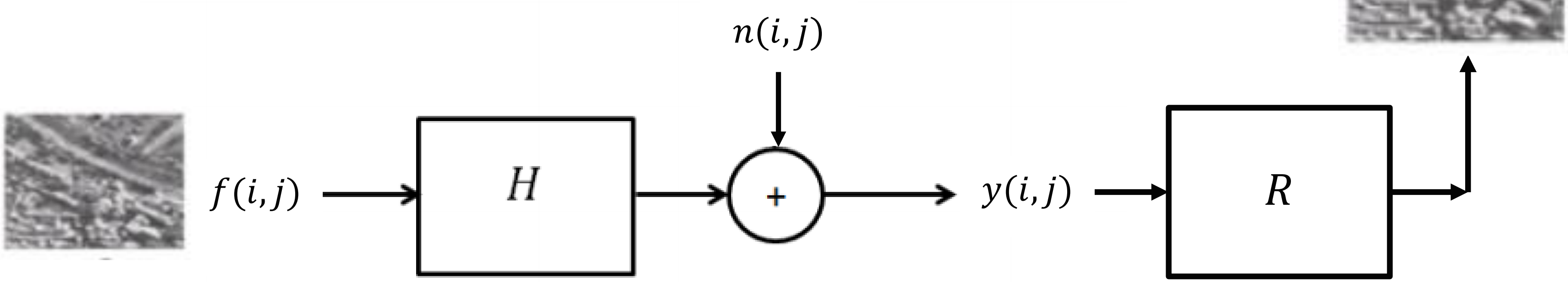
- In the case of a LSI degradation model the output is the convolution between the input and the degradation model as follows:

$$y(i,j) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} f(k,l)h(i-k,j-l) + n(i,j) = f(i,j) ** h(i,j) + n(i,j)$$

- Solving for $f(i,j)$, knowing the impulse response of the system $h(i,j)$ and the available data $y(i,j)$, is a deconvolution problem.
- The objective of restoration is to design a system R which will operate on the observation $y(i,j)$ and give us an estimate $\hat{f}$ of the original image.
- $\hat{f}$ should be as close as possible to the original image f subject to an optimisation criterion.
- In the subsequent methods, we assume that the following information is available:
  - The degradation model H
  - The noise statistics
  - Possibly some prior knowledge of f

A generic model of an image restoration system

- The objective of restoration is to design a system R which will operate on the observation $y(i,j)$ and give us an estimate $\hat{f}$ of the original image.
- $\hat{f}$ should be as close as possible to the original image f subject to an optimisation criterion.
- In most of the subsequent methods, we assume that the following information is available:
  - The degradation model H
  - The noise statistics
  - Possibly some prior knowledge of f



Motion blur: A typical type of degradation

Types of motion blur

- 1-D uniform motion blur

$$h(i) = \begin{cases} \dfrac{1}{2L+1} & -L \le i \le L \\ 0 & \text{otherwise} \end{cases}$$

- 2-D uniform motion blur

$$h(i,j) = \begin{cases} \dfrac{1}{(2L+1)^2} & -L \le i,j \le L \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{1}{9} \times \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array} \qquad \frac{1}{25} \times \begin{array}{|c|c|c|c|c|} \hline 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 \\ \hline \end{array} \qquad \frac{1}{49} \times \begin{array}{|c|c|c|c|c|c|c|} \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline \end{array}$$

Atmospheric turbulence: A typical type of degradation

Typical model for atmospheric turbulence

- Typical model for atmospheric turbulence: $h(i,j) = e^{-K(i^2+j^2)^{5/6}}$

R. E. Hufnagel and N. R. Stanley, Modulation Transfer Function Associated with Image Transmission through Turbulence Media, Optical Society of America Journal A, vol. 54, pp. 52–61, 1964.

negligible turbulence ⇒

severe
$\Leftarrow K = 0.0025$

mild
$K = 0.001 \Rightarrow$

low
$\Leftarrow K = 0.00025$

Uniform out-of-focus blur: A typical type of degradation

- 2-D out-of-focus blur

$$h(i,j) = \begin{cases} \dfrac{1}{\pi R^2} & i^2 + j^2 \leq R^2 \\ 0 & \text{otherwise} \end{cases}$$

- Note that the model is defined within a circular disc.

Imperial College London

An objective degradation metric: Blurred Signal-to-Noise-Ratio

- This is a metric that reflects the severity of additive noise $n(i,j)$ in relation to the blurred image.
- $z(i,j) = y(i,j) - n(i,j) = \sum_{k=0}^{M-1}\sum_{l=0}^{N-1} f(k,l)h(i-k,j-l)$ is the distorted image due to the degradation $h(i,j)$ only.
- $\bar{z}(i,j) = E\{z(i,j)\}$ is the expected value of $z(i,j)$.
- The BSNR is defined as follows:

$$\mathrm{BSNR} = 10\log_{10}\left\{\frac{\frac{1}{MN}\sum_i\sum_j[z(i,j)-\bar{z}(i,j)]^2}{\sigma_n^2}\right\}$$

  σ_n^2 is the variance of additive noise
- The numerator is the variance of the zero-mean, blurred but noiseless image $z(i,j) = \sum_{k=0}^{M-1}\sum_{l=0}^{N-1} f(k,l)h(i-k,j-l)$.

An objective restoration metric
Improvement in Signal-to-Noise Ratio (ISNR)

- This is a metric that reflects the quality of the restored image.
- $\hat{f}(i,j)$ is the estimated original image after applying an image restoration algorithm.
- The ISNR is defined as:

$$\mathrm{ISNR} = 10\log_{10}\left\{\frac{\sum_i\sum_j[f(i,j)-y(i,j)]^2}{\sum_i\sum_j\left[f(i,j)-\hat{f}(i,j)\right]^2}\right\}$$

- If $\sum_i\sum_j[f(i,j)-y(i,j)]^2 < \sum_i\sum_j\left[f(i,j)-\hat{f}(i,j)\right]^2$ then $\mathrm{ISNR} < 0$.

  In practice this implies that the restored image $\hat{f}(i,j)$ deviates from the true image more than the blurred image.
  - The above implies that not only we didn't achieve anything by applying restoration but we created an image "worse" compared to the one that we have already got.

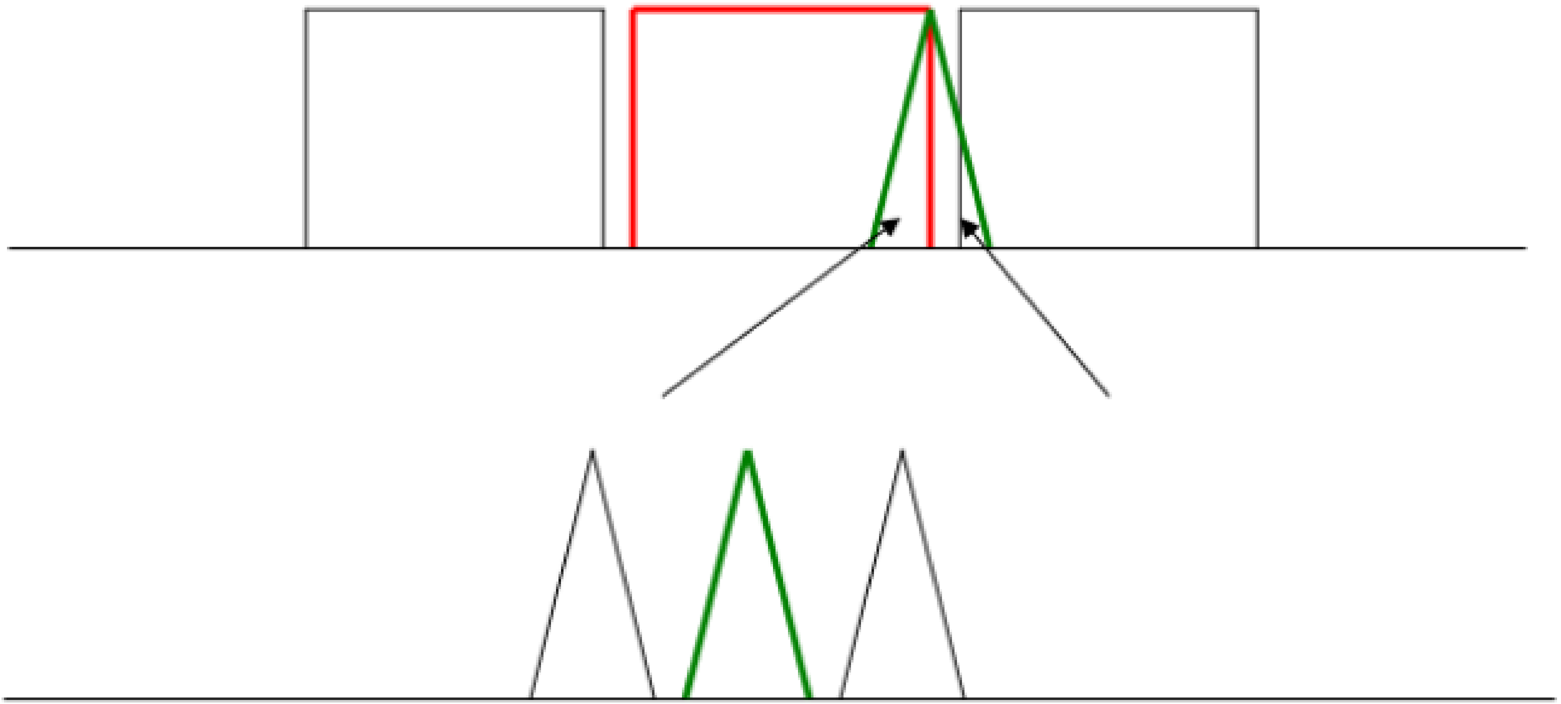
Periodic extension of images and degradation model

- In image restoration we often work with Discrete Fourier Transforms (DFTs).
- DFT assumes periodicity of the signal in both time or space **and** frequency domains.
- Therefore, periodic extension of images is implied.
- Using DFTs yields a distorted image that is the convolution of the original image and the distortion model. We are able to assume this because of the linearity and space invariance assumptions.
- Convolution increases the size of signals.
- Periodic extension must take into consideration the process of convolution which extends the signal lengths: zero-padding is required.
- Every signal involved in an image restoration system must be extended by zero-padding and also treated as virtually periodic.

Zero padding in 1D

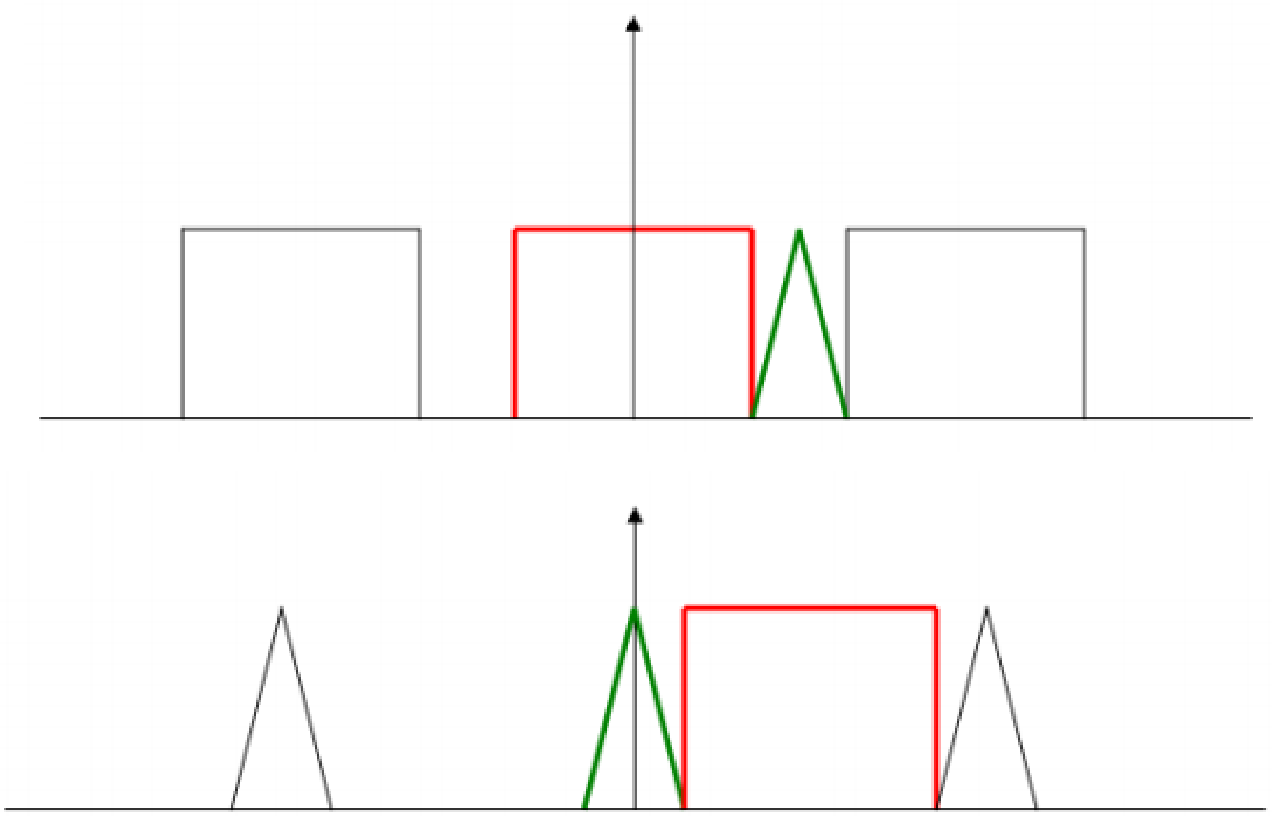
- Consider a system with input $x[n]$ of M samples and impulse response $h[n]$ of N samples.
- We pad both $x[n]$ and $h[n]$ with **dummy zero** samples so that their length becomes $M+N-1$.
- We now apply linear convolution and keep only the first $M+N-1$ samples.
- When we pad both $x[n]$ and $h[n]$ with dummy zero samples so that their length becomes $M+N-1$, their convolution is a signal of size $(M+N-1)+(M+N-1)-1=2M+2N-3$ samples.
- However, only the first $M+N-1$ samples of the original convolution are non-zero. The rest of the samples are 0!!!
- In that case the circular convolution of size $M+N-1$ samples is identical to the linear convolution of the original, non zero-padded signals!!!

Wrong periodic extension of signals
Red and green signal are convolved



Imperial College London

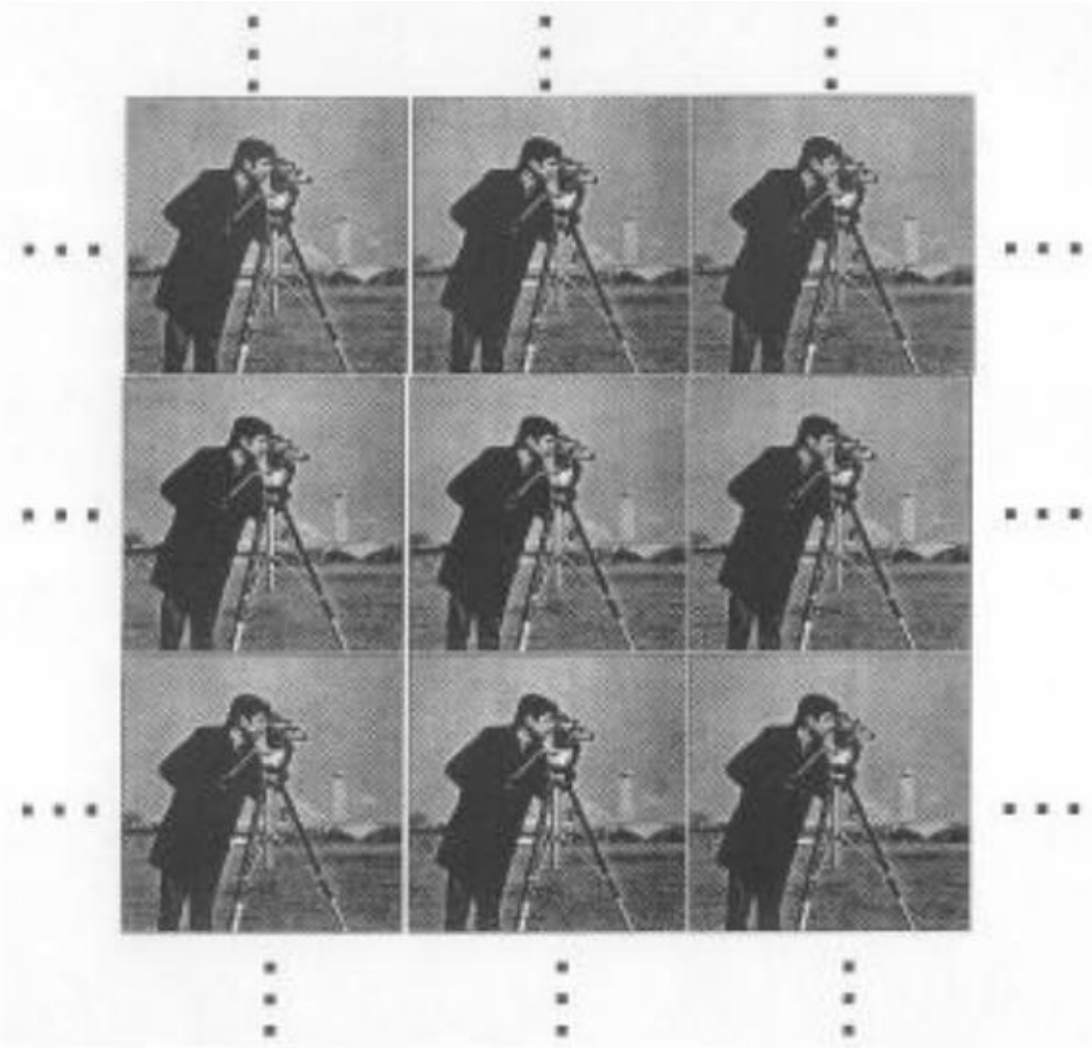
Correct periodic extension of signals
Red and green signal are convolved



Correct periodic extension of images and degradation model

- The original image $f(x,y)$ is of size $A \times B$.
- The degradation model $h(x,y)$ is of size $C \times D$.
- We form the extended versions of $f(x,y)$ and $h(x, y)$ by zero padding, both of size $M \times N$.
  - $M \geq A + C - 1$
  - $N \geq B + D - 1$
- Example
  - Image 256×256
  - Degradation 3×3
  - With extension by zero padding both images have dimension at least $(256 + 3 - 1) \times (256 + 3 - 1) = 258 \times 258$.
  - They are also assume to be periodic.

Correct periodic extension of images and degradation model



Circular convolution in two-dimensional signals (images)

- Suppose we have a two-dimensional discrete image $f(i,j)$ of size $A \times B$ samples which is due to a degradation process.
- The degradation can now be modeled by a two-dimensional discrete impulse response $h(i,j)$ of size $C \times D$ samples.
    - Recall the spatial masks we learnt.
- We form the extended versions of $f(i,j)$ and $h(i,j)$, both of size $M \times N$, where $M \geq A + C - 1$ and $N \geq B + D - 1$, and periodic with period $M \times N$. These can be denoted as $f_e(i,j)$ and $h_e(i,j)$.
- From this point forward we will assume that additive noise $n_e(i,j)$ is present in the model.
- For a linear, space-invariant degradation process we obtain

$$y_e(i,j) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} f_e(k,l) h_e(i-k, j-l) + n_e(i,j)$$

Lexicographic ordering

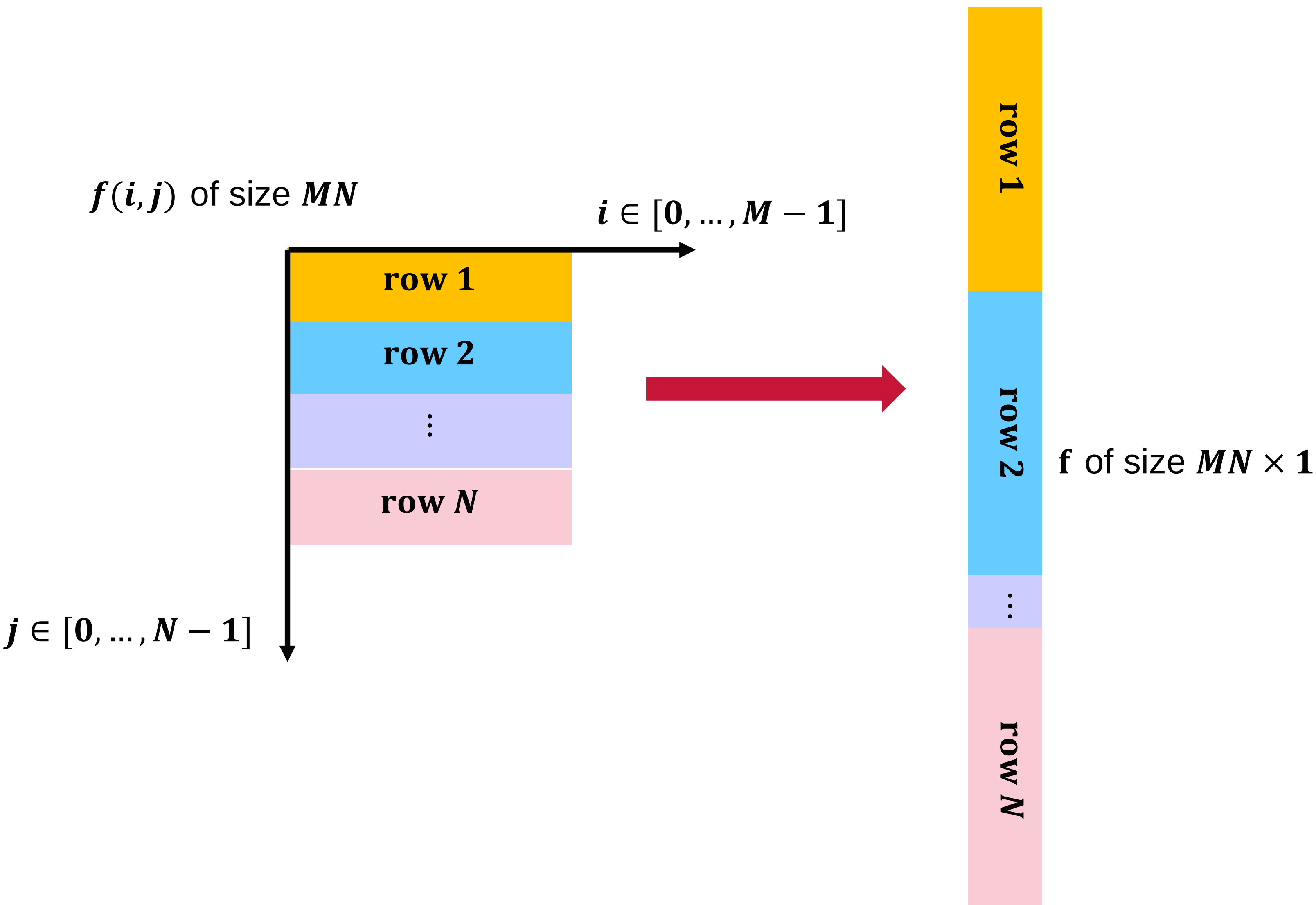
- In this section we use a mathematical tool for representing all the pixels in image as a vector.
- This can be done by stacking the rows of the image on top of each other. It is called **lexicographic ordering** of an image.
  - **Please observe next slide.**
- This introduces an alternative way for representing the degradation equation in the restoration problem; it becomes a vector-matrix equation.
- The input-output relationship is written as:

$$\mathbf{y} = \mathbf{Hf} + \mathbf{n}$$

where f and y are MN $-$dimensional column vectors that represent the lexicographic ordering of images $f_e(i,j)$ and $h_e(i,j)$, respectively.

- If the degradation is LSI, then this matrix has a specific form; it is a so-called **block-circulant** matrix.
- It is straightforward to describe this matrix in the spectral domain, by finding its eigenvalues and eigenvectors.

Lexicographic ordering of an image



The degradation model: block-circulant matrices

- As mentioned, in two dimensions the input-output relationship is written as:

$$\mathbf{y} = \mathbf{H}\mathbf{f} + \mathbf{n}$$

where the matrix $\mathbf{H}$ is a **block-circulant** matrix.

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_{M-1} & \dots & \mathbf{H}_1 \\ \mathbf{H}_1 & \mathbf{H}_0 & \dots & \mathbf{H}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{M-1} & \mathbf{H}_{M-2} & \dots & \mathbf{H}_0 \end{bmatrix}$$

- Each sub-matrix of a block-circulant matrix is a circular matrix itself.

$$\mathbf{H}_j = \begin{bmatrix} h_e(j,0) & h_e(j,N-1) & \dots & h_e(j,1) \\ h_e(j,1) & h_e(j,0) & \dots & h_e(j,2) \\ \vdots & \vdots & \ddots & \vdots \\ h_e(j,N-1) & h_e(j,N-2) & \dots & h_e(j,0) \end{bmatrix}$$

- Circulant and block-circulant matrices are very useful because it is very easy to diagonalize them:
  - More specifically, they can be decomposed in a sequence of matrices whose elements are related to the DFT values of their first row.

Restoration model in the frequency domain

- The restoration model can be transformed in the frequency domain.
- In that case the complicated large-scale matrix problem consists of a set of $M \times N$ scalar problems.

$$Y(u,v) = MNH(u,v)F(u,v) + N(u,v)$$

$u = 0,1, \ldots, M-1, v = 0,1, \ldots, N-1$

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 $y(i, j)$.

Inverse filtering for image restoration

- We formulate an unconstrained optimisation problem as follows:

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

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

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

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

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

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

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

  - 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.

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

$$= \Im^{-1}\left[\frac{Y(u,v)}{H(u,v)}\right]$$

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(s) (u,v) or for a whole region in the (u,v) plane. In that region inverse filtering cannot be applied.
  - If these points are known they can be neglected in the computation of $F(u,v)$.

- Note that in most real applications $H(u,v)$ drops off rapidly as a function of distance from the origin. **This is because it has the characteristics of a lowpass filter.**

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.
- In that case we have the so-called **noise amplification** effect.

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

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

or

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

Psedoinverse restoration examples

Figure 1: Degraded by a 7×7 pill-box blur, 20 dB BSNR

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

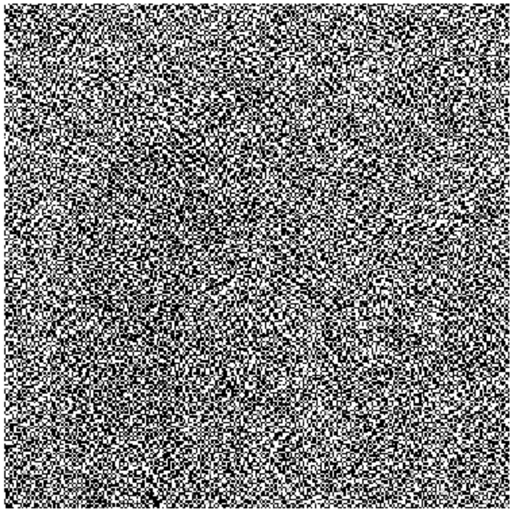


Figure 3 : Result of Figure 3 restored by a generalized inverse filter with a threshold of 10^{-3}, ISNR $= -32.9$ dB

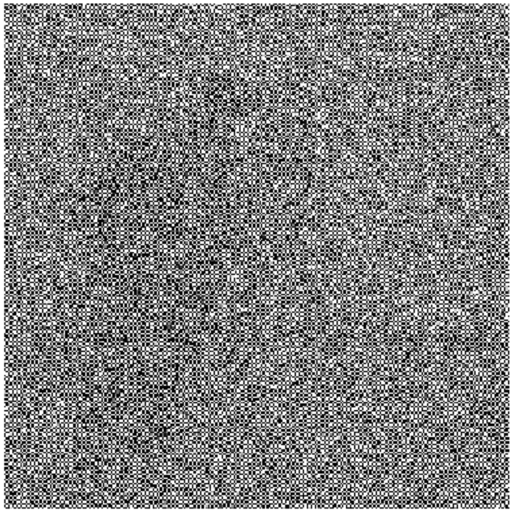


Figure 4 : Result of Figure 5 restored by a generalized inverse filter with a threshold of 10^{-3}, ISNR $= -36.6$ dB

Psedoinverse restoration examples

Figure 1: Degraded by a 7×7 pill-box blur, 20 dB BSNR

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

Figure 5 : Result of Figure 3 restored by a generalized inverse filter with a threshold of 10^{-1}, ISNR = 0.61 dB

Figure 6 : 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

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

  is equivalent to

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

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

- It is basically a smoothness constraint.
  $\mathbf{C}$ a high pass filter operator
  $\mathbf{Cf}$ 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,

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

additive noise signal
is not shown in this expression

- 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

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

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

- When $|H(u,v)|$ is zero or very small the danger of the denominator being zero or very small is eliminated due to the term $\alpha|C(u,v)|^2$.

C(u,v) is the frequency response of a high-pass filter.
For frequency pairs far from the origin where H(u,v) is small, C(u,v) will not be small.

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 loose its sharpness.

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:

$$\alpha = \frac{1}{\text{BSNR}} = \begin{array}{l}\text{large for small BSNR (heavy noise)}\\ \text{and small for large BSNR (light noise)}\end{array}$$

Choice of α

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

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

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

In statistics, the bias of an estimator is the difference between this estimator's expected value and the true value of the parameter being estimated.

- 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.

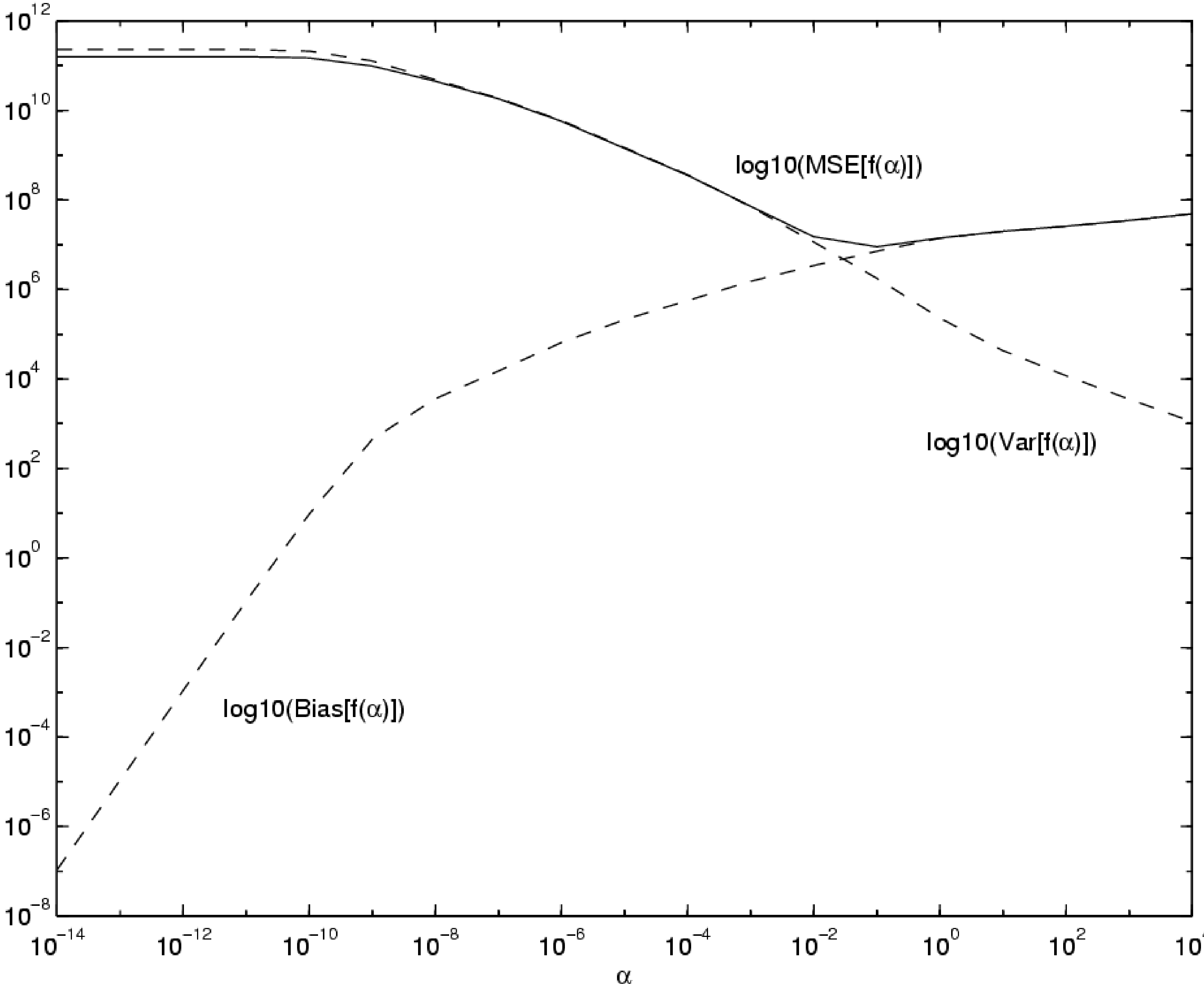

log10(MSE[f(α)])
log10(Var[f(α)])
log10(Bias[f(α)])
10^{12}
10^{10}
10^{8}
10^{6}
10^{4}
10^{2}
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-14}
10^{-12}
10^{-10}
10^{-8}
10^{-6}
10^{-4}
10^{-2}
10^{0}
10^{2}
10^{4}
α

Figure 1: Degraded by a 7×7 pill-box blur, 20 dB BSNR

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

Figure 7 : CLS restoration of Figure 3 with $\alpha = 1$, ISNR = 2.5 dB

Figure 8 : CLS restoration of Figure 5 with $\alpha = 1$, ISNR = 1.3 dB

Figure1 : Degraded by a 7×7 pill-box blur, 20 dB BSNR

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

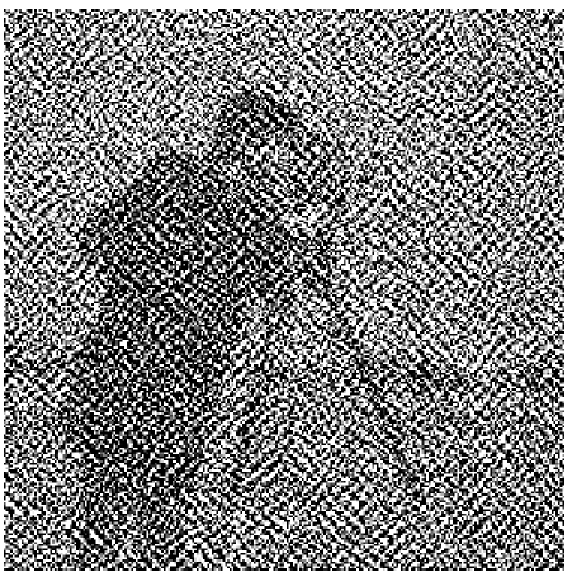


Figure 9 : CLS restoration of Figure 3 with $\alpha = 0.0001$, ISNR $= -22$ dB

Figure10 : CLS restoration of Figure 5 with $\alpha = 0.0001$, ISNR $= -22.1$ dB

Figure 1: Degraded by a 7×7 pill-box blur, 20 dB BSNR

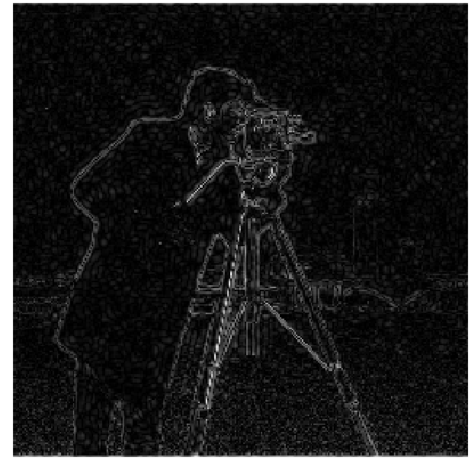


Figure 11: Corresponding error image for Figure 7 (|original − restored|, scaled for display)

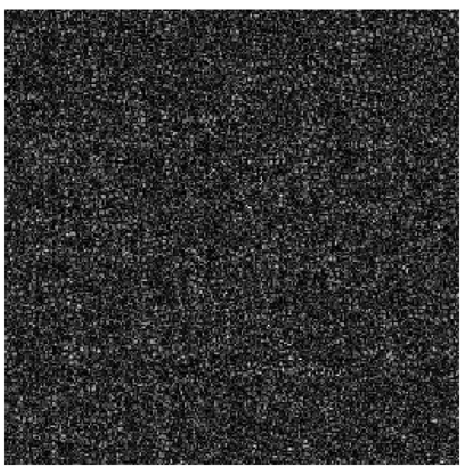


Figure 12: Corresponding error image for Figure 9 (|original − restored|, scaled for display)

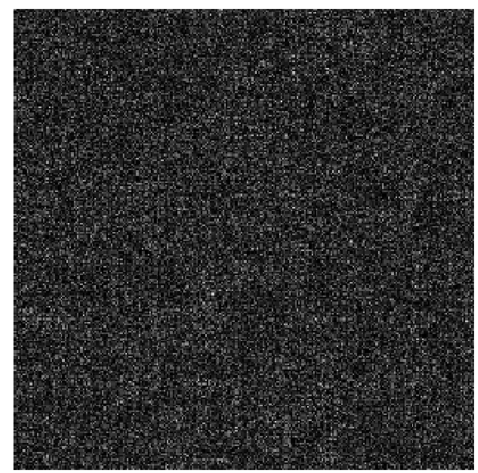
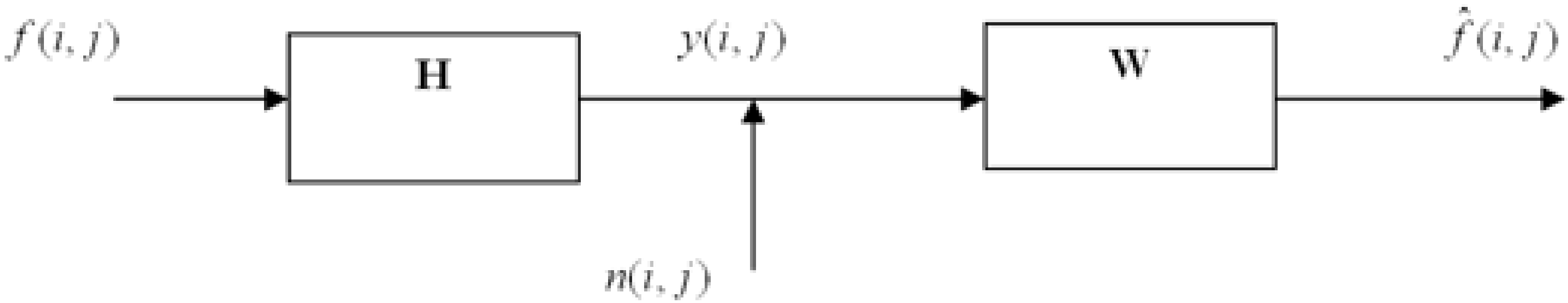


Figure 13: Corresponding error image for Figure 10 (|original − restored|, scaled for display)

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

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

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.

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

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

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

(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 $S_{nn}(u,v) \rightarrow 0$ then

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

will tend to the following:

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

The last relationship is the pseudoinverse filter.