In this lecture we will look at linear image restoration techniques
- Differentiation of matrices and vectors
- Linear space invariant degradation
- Restoration in absence of noise
  - Inverse filter
  - Pseudo-inverse filter
- Restoration in presence of noise
  - Inverse filter
  - Wiener filter
  - Constrained least squares filter
Notation:
\( A \) is a \( M \times N \) matrix with elements \( a_{ij} \).
\( x \) is a \( N \times 1 \) vector with elements \( x_i \).
\( f(x) \) is a scalar function of vector \( x \).
\( g(x) \) is a \( M \times 1 \) vector valued function of vector \( x \).

Scalar derivative of a matrix.
\( A \) is a \( M \times N \) matrix with elements \( a_{ij} \).
\[
\frac{\partial A}{\partial t} = \begin{pmatrix}
\frac{\partial a_{11}}{\partial t} & \cdots & \frac{\partial a_{1N}}{\partial t} \\
\vdots & \ddots & \vdots \\
\frac{\partial a_{M1}}{\partial t} & \cdots & \frac{\partial a_{MN}}{\partial t}
\end{pmatrix}
\]
Differentiation of Matrices and Vectors (cont...)

Vector derivative of a function (gradient).
x is a \( N \times 1 \) vector with elements \( x_i \).
f(\( \mathbf{x} \)) is a scalar function of vector \( \mathbf{x} \).

\[
\frac{\partial f}{\partial \mathbf{x}} = \nabla_{\mathbf{x}} f = \left( \frac{\partial f}{\partial x_1} \ldots \frac{\partial f}{\partial x_N} \right)^T
\]

Differentiation of Matrices and Vectors (cont...)

Vector derivative of a vector (Jacobian):
x is a \( N \times 1 \) vector with elements \( x_i \).
g(\( \mathbf{x} \)) is a \( M \times 1 \) vector valued function of vector \( \mathbf{x} \).

\[
\frac{\partial \mathbf{g}}{\partial \mathbf{x}} = \left( \begin{array}{ccc}
\frac{\partial g_1}{\partial x_1} & \ldots & \frac{\partial g_1}{\partial x_N} \\
\frac{\partial g_2}{\partial x_1} & \ldots & \frac{\partial g_2}{\partial x_N} \\
\vdots & \ddots & \vdots \\
\frac{\partial g_M}{\partial x_1} & \ldots & \frac{\partial g_M}{\partial x_N}
\end{array} \right)
\]
Some useful derivatives.

\( \mathbf{x} \) is a \( N \times 1 \) vector with elements \( x_i \).

\( \mathbf{b} \) is a \( N \times 1 \) vector with elements \( b_i \).

\[ \frac{\partial}{\partial \mathbf{x}} (\mathbf{b}^T \mathbf{x}) = \mathbf{b} \]

It is the derivative of the scalar valued function \( \mathbf{b}^T \mathbf{x} \) with respect to vector \( \mathbf{x} \).

---

Some useful derivatives.

\( \mathbf{x} \) is a \( N \times 1 \) vector with elements \( x_i \).

\( \mathbf{A} \) is a \( M \times N \) matrix with elements \( a_{ij} \).

\[ \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{A} \mathbf{x}) = \left( \mathbf{A} + \mathbf{A}^T \right) \mathbf{x} \]

If \( \mathbf{A} \) is symmetric:

\[ \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{A} \mathbf{x}) = 2 \mathbf{A} \mathbf{x} \]
Differentiation of Matrices and Vectors (cont...)

Some useful derivatives.

\( \mathbf{x} \) is a \( N \times 1 \) vector with elements \( x_i \).

\( \mathbf{b} \) is a \( N \times 1 \) vector with elements \( b_i \).

\( \mathbf{A} \) is a \( M \times N \) matrix with elements \( a_{ij} \).

\[
\frac{\partial}{\partial \mathbf{x}} \| \mathbf{Ax} + \mathbf{b} \|^2 = 2 \mathbf{A}^T (\mathbf{Ax} + \mathbf{b})
\]

It may be proved using the previous properties.

Linear, Position-Invariant Degradation

We now consider a degraded image to be modelled by:

\[
g(x, y) = h(x, y) * f(x, y) + \eta(x, y)
\]

where \( h(x, y) \) is the impulse response of the degradation function (i.e. point spread function blurring the image).

The convolution implies that the degradation mechanism is linear and position invariant (it depends only on image values and not on location).
In the Fourier domain:

\[ G(k, l) = H(k, l)F(k, l) + N(k, l) \]

where multiplication is element-wise.

In matrix-vector form:

\[ g = Hf + \eta \]

where \( H \) is a doubly block circulant matrix and \( f, g, \) and \( \eta \) are vectors (lexicographic ordering).

If the degradation function is unknown the problem of simultaneously recovering \( f(x, y) \) and \( h(x, y) \) is called blind deconvolution.
Linear Restoration

Using the imaging system

\[ g = Hf + \eta \]

we want to estimate the true image from the degraded observation with known degradation \( H \).

A linear method applies an operator (a matrix) \( P \) to the observation \( g \) to estimate the unobserved noise-free image \( f \):

\[ \hat{f} = Pg \]

Restoration in Absence of Noise

The Inverse Filter

When there is no noise:

\[ g = Hf \]

an obvious solution would be to use the inverse filter:

\[ P = H^{-1} \]

yielding

\[ \hat{f} = Pg = H^{-1}g = H^{-1}Hf = f \]
Restoration in Absence of Noise
The Inverse Filter (cont...)

\[ \hat{f} = H^{-1}g \]

For a \( N \times N \) image, \( H \) is a \( N^2 \times N^2 \) matrix!

To tackle the problem we transform it to the Fourier domain.

\( H \) is doubly block circulant and therefore it may be diagonalized by the 2D DFT matrix \( W \):

\[ H = W^{-1} \Lambda W \]

Restoration in Absence of Noise
The Inverse Filter (cont...)

where

\[ \Lambda = \text{diag}\{H(1,1),\ldots,H(N,1),H(1,2),\ldots,H(N,N)\} \]

Therefore:

\[ \hat{f} = Pg \iff \hat{f} = H^{-1}g \iff \hat{f} = (W^{-1} \Lambda W)^{-1} g \]

\[ \iff \hat{f} = W^{-1} \Lambda^{-1} W g \iff W\hat{f} = WW^{-1} \Lambda^{-1} W g \]

\[ \iff \hat{F} = \Lambda^{-1} G \]
Restoration in Absence of Noise
The Inverse Filter (cont...)

Which is the vectorized form of the DFT of the image:
\[ \hat{F} = \Lambda^{-1} G \iff \hat{F}(k,l) = \frac{G(k,l)}{H(k,l)} \]

Take the inverse DFT and obtain \( f(m,n) \).
Problem: what happens if \( H(k,l) \) has zero values?
Cannot perform inverse filtering!

Restoration in Absence of Noise
The Pseudo-inverse Filter

A solution is to set:
\[ \hat{F}(k,l) = \begin{cases} 
\frac{G(k,l)}{H(k,l)}, & H(k,l) \neq 0 \\
0, & H(k,l) = 0 
\end{cases} \]

which is a type of pseudo-inversion.
Notice that the signal cannot be restored at locations where \( H(k,l)=0 \).
A pseudo-inverse filter also arises by the unconstrained least squares approach.

Find the image $f$, that, when it is blurred by $H$, it will provide an observation as close as possible to $g$, i.e. It minimizes the distance between $Hf$ and $g$.

This distance is expressed by the norm:

$$J(f) = \|Hf - g\|^2$$

$$\min_f \{J(f)\} \iff \frac{\partial J}{\partial f} = 0 \iff \frac{\partial}{\partial f} \left(\|Hf - g\|^2\right) = 0$$

$$\iff 2H^T(Hf - g) = 0 \iff 2H^THf = 2H^Tg$$

$$\iff f = (H^TH)^{-1}H^Tg$$
Recall the imaging model with spatially invariant degradation and noise

\[ g(x, y) = h(x, y) \ast f(x, y) + \eta(x, y) \]

\[ G(k, l) = H(k, l)F(k, l) + N(k, l) \]

\[ g = Hf + \eta \]

Applying the inverse filter in the Fourier domain:

\[ \hat{F}(k, l) = F(k, l) + \frac{N(k, l)}{H(k, l)} \]

Even if we know \( H(k, l) \) we cannot recover \( F(k, l) \) due to the second term.

If \( H(k, l) \) has small values the second term dominates (it goes to infinity if \( H(k, l) = 0 \)).
One approach to get around the problem is to limit the ratio $G(k,l) / H(k,l)$ to frequencies near the origin that have lower probability of being zero.

We know that $H(0,0)$ is usually the highest value of the DFT. Thus, by limiting the analysis to frequencies near the origin we reduce the probability of encountering zero values.

Blurring degradation

C. Nikou – Digital Image Processing (E12)
The Inverse Filter (cont...)

Inverse filter with cut-off

So far we assumed nothing about the statistical properties of the image and noise. We now consider image and noise as random variables and the objective is to find an estimate of the uncorrupted image $\hat{f}$ such that the mean square error between the estimate and the image is minimized:

$$\min_{\hat{f}} \{ E[(f - \hat{f})^2] \}$$

where $E[x]$ is the expected value of vector $x$. 
Recall also the definition of the correlation matrix between two vectors $x$ and $y$:

$$R_{xy} = E[x_1 y_1, \ldots, x_N y_N] = \begin{bmatrix} E[x_1 y_1] & E[x_1 y_2] & \cdots & E[x_1 y_N] \\ E[x_2 y_1] & E[x_2 y_2] & \cdots & E[x_2 y_N] \\ \vdots & \vdots & \ddots & \vdots \\ E[x_N y_1] & E[x_N y_2] & \cdots & E[x_N y_N] \end{bmatrix}$$

We assume that the image and the noise are uncorrelated:

$$R_{nf} = R_{fn} = 0$$

We are looking for the best estimate:

$$\min_{\hat{f}} \left\{ E \left[ (f - \hat{f})^2 \right] \right\}$$

Let’s confine our estimate to be obtainable by a linear operator on the observation:

$$\hat{f} = Pg$$

and the goal is to find the best matrix $P$. 
Restoration in Presence of Noise
Wiener Filter (cont...)

\[ J(\hat{f}) = E\left[ (f - \hat{f})^2 \right] = E\left[ \|f - \hat{f}\|^2 \right] = E\left[ \|f - Pg\|^2 \right] \]

Denoting by \( p_n^T \) the \( n \)-th row of \( P \):

\[ J(\hat{f}) = E\left[ \sum_n (f_n - p_n^T g)^2 \right] = \sum_n E\left[ (f_n - p_n^T g)^2 \right] \]

\[ J(\hat{f}) = \sum_n E\left[ (f_n - p_n^T g) (f_n - p_n^T g)^T \right] \]

\[ = \sum_n E\left[ f_n f_n^T - p_n^T g f_n^T - f_n^T g p_n + p_n^T g g^T p_n \right] \]

\[ = \sum_n E\left[ f_n f_n^T \right] - p_n^T E\left[ g f_n^T \right] - E\left[ f_n^T g \right] p_n + p_n^T E\left[ g g^T \right] p_n \]

\[ = \sum_n R_{f_n f_n} - 2p_n^T R_{g f_n} + p_n^T R_{g g} p_n \]
Restoration in Presence of Noise
Wiener Filter (cont...)

We can now minimize the sum with respect to each term:

\[
\frac{\partial}{\partial p_n} \left( R_{f,f_n} - 2 p_n^T R_{g,f_n} + p_n^T R_{g,g} p_n \right) = 0
\]

\[\iff -2R_{g,f_n} + 2R_{g,g} p_n = 0 \iff p_n = R_{g,g}^{-1} R_{g,f_n}
\]

\[\iff p_n^T = R_{f,g} R_{g,g}^{-1}
\]

Assembling the rows together:

\[P = R_{f,g} R_{g,g}^{-1}
\]

We have to compute the two matrices:

\[R_{g,g} = \mathbb{E}[gg^T] = \mathbb{E}[(Hf + \eta)(Hf + \eta)^T]
\]

\[= \mathbb{E}[Hff^T + Hf\eta + \eta f^T H^T + \eta^T \eta]
\]

\[= HR_{f,f} H^T + HR_{f,\eta} + R_{\eta f} H^T + R_{\eta,\eta}
\]
Assuming noise is uncorrelated with image:

\[ R_{gg} = HR_{ff}^T + R_{\eta \eta} \]

Also,

\[ R_{fg} = E\left[ f g^T \right] = E\left[ f (Hf + \eta)^T \right] = \ldots = R_{ff} H^T \]

Finally the matrix we are looking for is

\[ P = R_{fg} R_{gg}^{-1} = R_{ff} H^T \left( HR_{ff} H^T + R_{\eta \eta} \right)^{-1} \]

The estimated uncorrupted image is

\[ \hat{f} = Pg \Leftrightarrow \hat{f} = R_{ff} H^T \left( HR_{ff} H^T + R_{\eta \eta} \right)^{-1} g \]

which may be also expressed as

\[ \hat{f} = \left( H^T R_{\eta \eta}^{-1} H + R_{ff}^{-1} \right)^{-1} H^T R_{\eta \eta}^{-1} g \]

This result is known as the Wiener filter or the minimum mean square error (MMSE) filter.
Restoration in Presence of Noise
Wiener Filter (cont...)

Special cases:
No blur ($H=I$, $g=f+\eta$): 
\[ \hat{f} = R_{ff} \left( R_{ff} + R_{\eta\eta} \right)^{-1} g \]

No noise ($R_{\eta\eta}=0$, $g=Hf$): 
\[ \hat{f} = H^{-1} g \]
This is the inverse filter.

No blur, no noise ($H=I$, $R_{\eta\eta}=0$): 
\[ \hat{f} = g \]
Do nothing on the observation.

The size of the matrix to be inverted poses difficulties and Wiener filter is implemented in the Fourier domain.
This occurs when $H$ is doubly block circulant (represents convolution) and the image $f$ and noise $\eta$ are wide-sense stationary (w.s.s).

Definition of a w.s.s. signal:
1) $E[f(m,n)]=\mu$, independent of $m,n$.
2) $E[f(m,n)f(k,l)]=r(m-k,n-l)$, independent of location.
Restoration in Presence of Noise
Wiener Filter (cont...)
The Wiener solution is now transformed to the Fourier domain:

\[
\hat{f} = R_{ff} H^T \left( R_{hh} H^T + R_{\eta\eta} \right)^{-1} g \\
= (W^{-1} \Lambda_{ff} W)(W^{-1} \Lambda_{ff}^* W) \left[ (W^{-1} \Lambda_{ff} W)(W^{-1} \Lambda_{ff} W)(W^{-1} \Lambda_{ff} W) + (W^{-1} \Lambda_{\eta\eta} W) \right]^{-1} g \\
= W^{-1} \Lambda_{ff} \Lambda_{ff}^* W \left[ W^{-1} (\Lambda_{ff} \Lambda_{ff}^* \Lambda_{ff} + \Lambda_{\eta\eta}) \right]^{-1} g
\]

\[
\Leftrightarrow W \hat{f} = \Lambda_{ff} \Lambda_{ff}^* (\Lambda_{ff} \Lambda_{ff}^* \Lambda_{ff} + \Lambda_{\eta\eta})^{-1} W g
\]

Notice that the matrices are diagonal.

\[
S_{ff}(k,l) = \text{DFT}(R_{ff}(m,n)) \text{ is the power spectrum of the image } f(m,n).
\]

\[
S_{\eta\eta}(k,l) = \text{DFT}(R_{\eta\eta}(m,n)) \text{ is the power spectrum of the noise } \eta(m,n).
\]
Restoration in Presence of Noise
Wiener Filter (cont...)

If \( S_{ff}(k,l) \) is not zero we may define the Signal to Noise Ratio in the frequency domain:

\[
\text{SNR}(k,l) = \frac{S_{ff}(k,l)}{S_{\eta\eta}(k,l)}
\]

and the Wiener filter becomes:

\[
F(k,l) = \frac{H^*(k,l)}{|H(k,l)|^2 + \text{SNR}^{-1}(k,l)} G(k,l)
\]

A well known estimate of \( S_{ff}(k,l) \) is the periodogram (the ML estimate of \( R_{ff} \) when \( f \) is assumed Gaussian):

\[
S_{ff}(k,l) = \frac{1}{MN} |F(k,l)|^2
\]

In practice, as \( F(k,l) \) is unknown, we use

\[
\hat{S}_{ff}(k,l) = \frac{1}{MN} |G(k,l)|^2
\]
Restoration in Presence of Noise
Wiener Filter (cont...)

FIGURE 5.28 Comparison of inverse and Wiener filtering (a) Result of full inverse filtering of Fig. 5.25(b). (b) Radially limited inverse filter result. (c) Wiener filter result.

Noise variance one order of magnitude less.

Noise variance ten orders of magnitude less.
When we do not have information on the power spectra, the Wiener filter is not optimal. Another idea is to introduce a smoothness term in our criterion. We define smoothness by the quantity $\|Qf\|^2$ where $Q$ is a high-pass filter operator, e.g., the Laplacian, $(Q$ is a doubly block circulant matrix representing convolution by the Laplacian). We look for smooth solutions minimizing $\|Qf\|^2$.

We have the following constrained least squares (CLS) optimization problem:

Minimize $\|Qf\|^2$ subject to $Hf = g$

yielding the Lagrange multiplier minimization of the function:

$$J(f, \lambda) = \|Hf - g\|^2 + \lambda \|Qf\|^2$$

Data fidelity term Smoothness term
Parameter $\lambda$ controls the degree of smoothness:

- $\lambda = 0$, $f = (H^T H)^{-1} H^T g$ pseudo-inverse, ultra rough solution
- $\lambda \to \infty$, $f = 0$ ultra smooth solution

In the Fourier domain, the constrained least squares filter becomes:

$$F(k, l) = \frac{H^*(k, l)}{\|H(k, l)\|^2 + \lambda \|Q(k, l)\|^2} G(k, l)$$

Keep always in mind to zero-pad the images properly.
Restoration in Presence of Noise
Constrained Least Squares Filter (cont...)

Low noise: Wiener and CLS generate equal results. High noise: CLS outperforms Wiener if $\lambda$ is properly selected. It is easier to select the scalar value for $\lambda$ than to approximate the SNR which is seldom constant.

Restoration Performance Measures

Original image $f$ and restored image $\hat{f}$

Mean square error (MSE):

$$MSE = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[ f(m,n) - \hat{f}(m,n) \right]^2$$
The Signal to Noise Ratio (SNR) considers the difference between the two images as noise:

\[
\text{SNR} = \frac{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \hat{f}(m,n)^2}{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[ f(m,n) - \hat{f}(m,n) \right]^2}
\]