Signal and Image Processing  (236327)

Tutorial 13

Wiener Filtering and Constrained Deconvolution
Wiener Filtering for Signal Restoration: Discrete Problem Settings

A discrete signal, denoted as the $N$-length column vector $\phi$, is deteriorated according to the model:

$$\phi_{data} = H\phi + \n$$

where

- $H$ is a known $N \times N$ matrix representing a linear degradation operator.
- $\n$ is an additive noise vector, considered as a realization of an $N$-length random vector, having i.i.d components that follow the properties:

$$\mu_n \triangleq E\n = 0 \quad \text{and} \quad R_n \triangleq E\n\n^T = \sigma_n^2 I$$

- $\phi_{data}$ is the given degraded signal (a $N$-length column vector).

The task is to estimate the unknown signal $\phi$. 
Wiener Filtering for Signal Restoration: Discrete Problem Settings

The signal $\bar{\phi}$ is considered here as a realization of a random vector, associated with a class of signals, having the second-order statistics:

- Mean vector $\bar{\mu}_\phi \triangleq E\{\bar{\phi}\} = 0$ (we consider zero mean for simplicity)
- An autocorrelation matrix $R_\phi \triangleq E\{\bar{\phi}\bar{\phi}^T\}$

Then, the Wiener filter is the matrix

$$W = R_\phi H^T(HR_\phi H^T + \sigma_n^2 I)^{-1}$$

and the signal estimate is

$$\bar{\phi}_{est}^{opt} = W\bar{\phi}_{data} = R_\phi H^T(HR_\phi H^T + \sigma_n^2 I)^{-1}\bar{\phi}_{data}$$
Wiener Filtering for Signal Denoising: An Example

Consider a discrete signal, denoted as the $N$-length column vector $\bar{\varphi}$, being a realization of a (statistical) class of signals defined as:

$$\bar{\varphi} = [L_1, ..., L_1, L_2, ..., L_2]^T$$

where $L_1, L_2$ and $K$ are statistically independent random variables:

- $K$ is uniformly distributed among the integers $1, ..., N$.
- $L_1$ and $L_2$ follow $E\{L_1\} = E\{L_2\} = 0$ and $\text{var}\{L_1\} = \text{var}\{L_2\} = \sigma_L^2$. 

Wiener Filtering for Signal Denoising: An Example

The signal $\bar{\varphi}$ is deteriorated by an i.i.d additive noise:

$$\bar{\varphi}_{data} = \bar{\varphi} + \bar{n}$$

where

$\bar{n}$ follows the second-order statistics:

$$\bar{\mu}_n \triangleq E\bar{n} = \bar{0} \quad \text{and} \quad R_n \triangleq E\bar{n}\bar{n}^T = \sigma_n^2 I$$

and $\bar{\varphi}_{data}$ is the given noisy signal.

• The task is to denoise $\bar{\varphi}_{data}$ by estimating the unknown signal $\bar{\varphi}$.

• Formulate the Wiener filter for the above defined denoising problem.
Formulating the Wiener filter requires the second-order statistics of the signal class:

In general we write \( \bar{\phi} = [\varphi_1, \ldots, \varphi_N]^T \)

then, the mean is

\[
E\{\bar{\phi}\} = [E\{\varphi_1\}, \ldots, E\{\varphi_N\}]^T
\]

The mean of the \( i^{th} \) signal sample is

\[
E\{\varphi_i\} = E\{\varphi_i| i \leq K\} \cdot P\{i \leq K\} + E\{\varphi_i| i > K\} \cdot P\{i > K\}
\]

\[
= E\{L_1| i \leq K\} \cdot P\{i \leq K\} + E\{L_2| i > K\} \cdot P\{i > K\}
\]

(by the signal definition)

\[
= E\{L_1\} \cdot P\{i \leq K\} + E\{L_2\} \cdot P\{i > K\}
\]

(Since \( L_1 \) and \( L_2 \) are independent of \( K \))

\[
= 0 \cdot P\{i \leq K\} + 0 \cdot P\{i > K\}
\]

\[
= 0
\]

This implies that \( E\{\bar{\phi}\} = \bar{0} \) (the signal has zero mean).
Wiener Filtering for Signal Denoising: An Example

The signal autocorrelation matrix is \( \mathbf{R}_\varphi = E\{\varphi \varphi^T\} \)

and the \((i, j)\) component of \( \mathbf{R}_\varphi \) is \( r_{ij} = E\{\varphi_i \varphi_j\} \).

Let us consider \( i \leq j \):

\[
E\{\varphi_i \varphi_j\} = E\{\varphi_i \varphi_j | i > K\} \cdot P\{i > K\} + E\{\varphi_i \varphi_j | j \leq K\} \cdot P\{j \leq K\} + E\{\varphi_i \varphi_j | i \leq K < j\} \cdot P\{i \leq K < j\}
\]

We evaluate the three cases as:

\[
E\{\varphi_i \varphi_j | i > K\} = E\{L_2^2 | i > K\} = E\{L_2^2\} = \sigma_L^2
\]

\[
E\{\varphi_i \varphi_j | j \leq K\} = E\{L_1^2 | j \leq K\} = E\{L_1^2\} = \sigma_L^2
\]

\[
E\{\varphi_i \varphi_j | i \leq K < j\} = E\{L_1 L_2 | i \leq K < j\} = E\{L_1 L_2\} = E\{L_1\}E\{L_2\} = 0
\]

Leading to \( E\{\varphi_i \varphi_j\} = \sigma_L^2 \cdot P\{i > K\} + \sigma_L^2 \cdot P\{j \leq K\} \)
Wiener Filtering for Signal Denoising: An Example

We got that $E\{\phi_i \phi_j\} = \sigma_L^2 \cdot (P\{i > K\} + P\{j \leq K\})$

and due to the uniform distribution of $K$:

$$P\{i > K\} = \frac{i-1}{N} \quad \text{and} \quad P\{j \leq K\} = \frac{N-j+1}{N}$$

Then, for $i \leq j$:

$$E\{\phi_i \phi_j\} = \sigma_L^2 \cdot \left(\frac{i-1}{N} + \frac{N - j + 1}{N}\right) = \sigma_L^2 \cdot \frac{N - (j - i)}{N}$$

Similar developments for $i \geq j$ show:

$$E\{\phi_i \phi_j\} = \sigma_L^2 \cdot \frac{N - (i - j)}{N}$$

Hence, for any $i, j$:

$$E\{\phi_i \phi_j\} = \sigma_L^2 \cdot \frac{N - |j - i|}{N}$$
Wiener Filtering for Signal Denoising: An Example

Since  \( E\{\varphi_i \varphi_j\} = \sigma_L^2 \cdot \frac{N-|j-i|}{N} \) the signal autocorrelation matrix is

\[
\mathbf{R}_{\varphi} = \sigma_L^2 \cdot \begin{bmatrix}
1 & \frac{N-1}{N} & \cdots & \frac{1}{N} \\
\frac{N-1}{N} & 1 & \cdots & \vdots \\
\vdots & \vdots & \ddots & \frac{N-1}{N} \\
\frac{1}{N} & \cdots & \frac{N-1}{N} & 1
\end{bmatrix}
\]

The Wiener filter, \( \mathbf{W} \), for the described denoising problem is obtained by setting \( \mathbf{R}_{\varphi} \) in

\[
\mathbf{W} = \mathbf{R}_{\varphi} \left( \mathbf{R}_{\varphi} + \sigma_n^2 \mathbf{I} \right)^{-1}
\]

The corresponding signal estimate (denoised signal):

\[
\varphi_{est}^{opt} = \mathbf{R}_{\varphi} \left( \mathbf{R}_{\varphi} + \sigma_n^2 \mathbf{I} \right)^{-1} \varphi_{data}
\]
Constrained Deconvolution: Discrete Problem Settings

A discrete signal, denoted as the $N$-length column vector $\bar{\phi}$, is deteriorated according to the model:

$$\bar{\phi}_{data} = H\bar{\phi} + \bar{n}$$

where

- $H$ is a known $N \times N$ matrix representing a linear degradation operator.
- $\bar{n}$ is an additive noise vector, considered as a realization of an $N$-length random vector, having i.i.d components that follow the properties:
  $$\bar{\mu}_n \triangleq E\bar{n} = 0 \quad \text{and} \quad R_n \triangleq E\bar{n}\bar{n}^T = \sigma_n^2I$$
- $\bar{\phi}_{data}$ is the given degraded signal (a $N$-length column vector).

The task is to estimate the unknown signal $\bar{\phi}$. 
Constrained Deconvolution: Discrete Problem Settings

The degradation model:

$$\bar{\varphi}_{data} = H\bar{\varphi} + n$$

The signal $\bar{\varphi}$ is considered here via a **deterministic** model assuming that $A\bar{\varphi}$ is a short vector, i.e., the quantity

$$\|A\bar{\varphi}\|_2^2 = \bar{\varphi}^T A^T A \bar{\varphi}$$

is small.

Here, $A$ is an $N \times N$ matrix.

**For example:**

$$A = \begin{bmatrix} -1 & 1 \\ -1 & 1 \\ \vdots & \vdots \\ -1 & 1 \\ 1 & -1 \end{bmatrix}$$

for this $A$, the vector $A\bar{\varphi}$ is the discrete derivative of the signal. Then, small $\|A\bar{\varphi}\|_2^2$ implies that $\bar{\varphi}$ is slowly varying (and, hence, relatively smooth).
Constrained Deconvolution: Discrete Problem Settings

The degradation model:

\[ \overline{\varphi}_{data} = H\overline{\varphi} + \bar{n} \]

It's important to note the difference from the Wiener filter settings presented in the lecture, where the signal \( \overline{\varphi} \) is considered as a realization of a random vector defined by its second-order statistics.

In the settings presented here for the constrained deconvolution:

- The additive noise is still considered as a realization of a random vector.
- The signal is characterized deterministically by having small \( \|A\overline{\varphi}\|_2^2 \).
Constrained Deconvolution: 
The Signal Estimate

Let us denote the signal estimate as $\Phi_{est}$.

Obviously, we would like the estimate to satisfy (or at least approximate) the degradation model via

$$
\Phi_{data} \approx H\Phi_{est} + \bar{n}
$$

leading to

$$
\Phi_{data} - H\Phi_{est} \approx \bar{n}
$$

We notice that

$$
\|\Phi_{data} - H\Phi_{est}\|_2^2 \approx \|\bar{n}\|_2^2
$$
Constrained Deconvolution: The Signal Estimate

$$\|\Phi_{data} - H\Phi_{est}\|^2 \approx \|\bar{n}\|^2 = \bar{n}^T \bar{n} = \sum_{i=1}^{N} n_i^2 \approx N \sigma_n^2$$

Recall that all the noise components are identically distributed, allowing the last transition as an empirical approximation of a noise-sample variance (the approximation has a reasonable accuracy for a large $N$).

Empirical approximation of variance:

In general, consider a zero-mean random variable $x$.

Given $M$ realizations of $x$, denoted as $x_1, ..., x_M$, the variance of $x$ can be estimated for a large $M$ as

$$\text{var}\{x\} \approx \frac{1}{M} \sum_{i=1}^{M} x_i^2$$
Constrained Deconvolution: Optimization Formulation

Following the presented properties the signal estimate as $\bar{\varphi}_{est}$, we pose the (constrained) optimization problem as

$$\begin{align*}
\text{minimize} \quad & \|A\bar{\varphi}_{est}\|_2^2 \\
\text{subject to} \quad & \|\bar{\varphi}_{data} - H\bar{\varphi}_{est}\|_2^2 = N\sigma_n^2
\end{align*}$$

having the \textbf{unconstrained} Lagrangian form of

$$\begin{align*}
\text{minimize} \quad & \|A\bar{\varphi}_{est}\|_2^2 + \lambda \|\bar{\varphi}_{data} - H\bar{\varphi}_{est}\|_2^2 \\
\text{where} \quad & \lambda \geq 0 \text{ is a Lagrange multiplier that leads to an estimate satisfying } \|\bar{\varphi}_{data} - H\bar{\varphi}_{est}\|_2^2 = N\sigma_n^2.
\end{align*}$$
Constrained Deconvolution: Optimization Formulation

The Lagrangian cost function to be minimized is defined as

\[ \Psi(\bar{\varphi}_{est}) = \| A \bar{\varphi}_{est} \|_2^2 + \lambda \| \bar{\varphi}_{data} - H \bar{\varphi}_{est} \|_2^2 \]

For some \( \lambda \geq 0 \) we can minimize the above cost via

\[ \frac{\partial}{\partial \bar{\varphi}_{est}} \Psi(\bar{\varphi}_{est}) = 0 \]

We notice that

\[ \Psi(\bar{\varphi}_{est}) = \bar{\varphi}_{est}^T A^T A \bar{\varphi}_{est} + \lambda (\bar{\varphi}_{data} - H \bar{\varphi}_{est})^T (\bar{\varphi}_{data} - H \bar{\varphi}_{est}) \]
Constrained Deconvolution: Solving the Optimization

Developing further:

\[ \Psi(\phi_{est}) = \phi_{est}^T A^T A \phi_{est} + \lambda (\phi_{data} - H \phi_{est})^T (\phi_{data} - H \phi_{est}) \]

\[ = \phi_{est}^T A^T A \phi_{est} + \lambda \phi_{data}^T \phi_{data} - \lambda \phi_{data}^T H \phi_{est} - \lambda \phi_{est}^T H^T \phi_{data} + \lambda \phi_{est}^T H^T H \phi_{est} \]

and the cost derivative is

\[ \frac{\partial}{\partial \phi_{est}} \Psi(\phi_{est}) = \]

\[ = \frac{\partial}{\partial \phi_{est}} \{ \phi_{est}^T A^T A \phi_{est} \} - \lambda \frac{\partial}{\partial \phi_{est}} \{ \phi_{data}^T H \phi_{est} \} - \lambda \frac{\partial}{\partial \phi_{est}} \{ \phi_{est}^T H^T \phi_{data} \} + \lambda \frac{\partial}{\partial \phi_{est}} \{ \phi_{est}^T H^T H \phi_{est} \} \]
Constrained Deconvolution: Solving the Optimization

The following derivation (by a vector) formulas are useful here:

- \( \frac{\partial}{\partial \bar{x}} \{ \bar{v}^T \bar{x} \} = \frac{\partial}{\partial \bar{x}} \{ \bar{x}^T \bar{v} \} = \bar{v} \) where \( \bar{x} \) and \( \bar{v} \) are two column vectors of the same size.

- \( \frac{\partial}{\partial \bar{x}} \{ \bar{x}^T Z \bar{x} \} = Z \bar{x} + Z^T \bar{x} = (Z + Z^T) \bar{x} \)
  where \( \bar{x} \) is \( L \)-length column vector and \( Z \) is a matrix of size \( L \times L \).

Using the above formulas we get

\[
\frac{\partial}{\partial \bar{\varphi}_{est}} \Psi(\bar{\varphi}_{est}) = \frac{\partial}{\partial \bar{\varphi}_{est}} \{ \bar{\varphi}_{est}^T A^T A \bar{\varphi}_{est} \} - \lambda \frac{\partial}{\partial \bar{\varphi}_{est}} \{ \bar{\varphi}_{data} H \bar{\varphi}_{est} \} - \lambda \frac{\partial}{\partial \bar{\varphi}_{est}} \{ \bar{\varphi}_{est}^T H^T \bar{\varphi}_{data} \} + \lambda \frac{\partial}{\partial \bar{\varphi}_{est}} \{ \bar{\varphi}_{est}^T H^T H \bar{\varphi}_{est} \}
\]

\[
= 2A^T A \bar{\varphi}_{est} - \lambda H^T \bar{\varphi}_{data} - \lambda H^T H \bar{\varphi}_{est} + \lambda \cdot 2H^T H \bar{\varphi}_{est}
\]

\[
= 2(A^T A + \lambda H^T H) \bar{\varphi}_{est} - 2\lambda H^T \bar{\varphi}_{data}
\]

\[
\frac{\partial}{\partial \bar{\varphi}_{est}} \Psi(\bar{\varphi}_{est}) = 0 \quad \rightarrow 2(A^T A + \lambda H^T H) \bar{\varphi}_{est} - 2\lambda H^T \bar{\varphi}_{data} = 0
\]
Constrained Deconvolution: Optimal Estimate

Setting the last result in
\[ \frac{\partial}{\partial \bar{\varphi}_{est}} \Psi(\bar{\varphi}_{est}) = 0 \]
gives
\[ 2( \mathbf{A}^T \mathbf{A} + \lambda \mathbf{H}^T \mathbf{H}) \bar{\varphi}_{est} - 2\lambda \mathbf{H}^T \bar{\varphi}_{data} = 0 \]

and the optimal estimate is
\[ \bar{\varphi}_{est}^{opt} = ( \mathbf{A}^T \mathbf{A} + \lambda \mathbf{H}^T \mathbf{H})^{-1} \lambda \mathbf{H}^T \bar{\varphi}_{data} \]

Note that the above optimal estimate is for a given \( \lambda \geq 0 \).

Hence, we consider the optimal estimate as a function of \( \lambda \):
\[ \bar{\varphi}_{est}^{opt}(\lambda) \]

Accordingly, we should find a \( \lambda \) that satisfies (or at least approximates)
\[ \| \bar{\varphi}_{data} - \mathbf{H} \bar{\varphi}_{est}^{opt}(\lambda) \|_2^2 \approx N\sigma_n^2 \]

When \( \mathbf{H} \) and \( \mathbf{A} \) are linear shift-invariant operators, it can be shown that \( \lambda \) can be found by a (conceptually) simple procedure.
## Constrained Deconvolution and Wiener Filtering: A Comparison

<table>
<thead>
<tr>
<th></th>
<th>Constrained Deconvolution</th>
<th>Wiener Filtering (proved in lecture)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Degradation Model</strong></td>
<td>$\overline{\phi}_{data} = H\overline{\phi} + \overline{n}$</td>
<td></td>
</tr>
<tr>
<td><strong>Signal Model</strong></td>
<td>Deterministic model: For a linear operator $A$, $|A\overline{\phi}|^2_2$ is small.</td>
<td>Stochastic model: $\overline{\phi}$ is a realization of a random vector with second order statistics: Mean vector $\overline{\mu}<em>\phi \triangleq E{\overline{\phi}} = 0$ (zero mean for simplicity) and an autocorrelation matrix $R</em>\phi \triangleq E{\overline{\phi}\overline{\phi}^T}$</td>
</tr>
<tr>
<td><strong>Noise Model</strong></td>
<td>Zero-mean additive white noise with second-order statistics $\overline{\mu}_n \triangleq E\overline{n} = 0$, $R_n \triangleq E\overline{n}\overline{n}^T = \sigma_n^2 I$</td>
<td></td>
</tr>
<tr>
<td><strong>Optimal Estimate</strong></td>
<td>$\overline{\phi}<em>{est}^{opt} = (A^T A + \lambda H^T H)^{-1} \lambda H^T \overline{\phi}</em>{data}$</td>
<td>$\overline{\phi}<em>{est}^{opt} = R</em>\phi H^T (HR_\phi H^T + \sigma_n^2 I)^{-1} \overline{\phi}_{data}$</td>
</tr>
</tbody>
</table>

An appropriate $\lambda$ should be determined with respect to $\sigma_n^2$. 

- $\lambda$ represents the trade-off between the fit to the data and the smoothness of the estimated signal.
- $\sigma_n^2$ is the variance of the noise.
- $A$ is the degradation operator.
- $H$ is the system transfer function.
- $\overline{\phi}$ is the true signal.
- $\overline{n}$ is the noise term.