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 $\overline{\varphi}$, is deteriorated according to the model:

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

where

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

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

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

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

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

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

Then, the Wiener filter is the matrix

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

and the signal estimate is

$$\bar{\varphi}_{est}^{opt} = W \bar{\varphi}_{data}$$

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

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

\[
\bar{\phi} = [L_1, \ldots, L_1, L_2, \ldots, L_2]^T
\]

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

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

$N-K$ components of value $L_1$

$K$ components of value $L_2$
Wiener Filtering for Signal Denoising: An Example

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

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

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

- The task is to denoise $\bar{\phi}_{data}$ by estimating the unknown signal $\bar{\phi}$.
- Formulate the Wiener filter for the above defined denoising problem.
Wiener Filtering for Signal Denoising: An Example

Formulating the Wiener filter requires the second-order statistics of the signal class:

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

then, the mean is

\[
E\{\overline{\varphi}\} = [E\{\varphi_1\}, \ldots, E\{\varphi_N\}]^T
\]

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 \( E\{\overline{\varphi}\} = \overline{0} \) (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 $i \leq j$:

$$E\{\varphi_i \varphi_j\} = E\{\varphi_i \varphi_j|i > K\} \cdot P\{i > K\}$$

$$+ E\{\varphi_i \varphi_j|j \leq K\} \cdot P\{j \leq K\}$$

$$+ E\{\varphi_i \varphi_j|i \leq K < j\} \cdot P\{i \leq K < j\}$$

(i, j both to the right of the level-transition)

(i, j both to the left of the level-transition)

(the level-transition is between $i$ and $j$)

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

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

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 $i \leq j$:

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

Similar developments for $i \geq j$ show:

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

Hence, for any $i, j$:

$$E\{\varphi_i\varphi_j\} = \sigma_L^2 \cdot \left( 1 - \frac{|j-i|}{N} \right)$$
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

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

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):

\[
\bar{\varphi}_{est}^{opt} = \mathbf{R}_\varphi \left( \mathbf{R}_\varphi + \sigma_n^2 \mathbf{I} \right)^{-1} \bar{\varphi}_{data}
\]
Wiener Filtering for Signal Denoising: An Example

The corresponding signal estimate (denoised signal):
\[
\phi_{est}^{opt} = R_\phi(R_\phi + \sigma_n^2 I)^{-1} \phi_{data}
\]

where
\[
R_{\phi} = \sigma_L^2 \cdot \begin{bmatrix}
1 & \frac{N-1}{N} & \cdots & \frac{1}{N} \\
\frac{N-1}{N} & 1 & \ddots & \vdots \\
\vdots & \ddots & \ddots & \frac{N-1}{N} \\
\frac{1}{N} & \cdots & \frac{N-1}{N} & 1
\end{bmatrix}
\]

What is the unitary matrix \( U \) that transforms the denoising task to be componentwise?

Note that whereas \( R_\phi \) is Toeplitz, it is not circulant and therefore the DFT matrix does not diagonalize it.

Accordingly, we generally state that \( U \) is the PCA matrix of \( R_\phi \).
Wiener Filtering for Signal Denoising: An Example

\[
\overline{\phi}^{opt} = R_\phi (R_\phi + \sigma_n^2 I)^{-1} \overline{\phi}_{data}
\]

The transition into the PCA-domain component-based denoising is observed via:

\[
\overline{\phi}^{opt,(U)} = U^* \overline{\phi}^{opt} = U^* R_\phi (R_\phi + \sigma_n^2 I)^{-1} \overline{\phi}_{data}
\]

\[
= U^* R_\phi U U^* (R_\phi + \sigma_n^2 I)^{-1} U U^* \overline{\phi}_{data}
\]

\[
= U^* R_\phi U (U^* R_\phi U + \sigma_n^2 U^* U)^{-1} U^* \overline{\phi}_{data}
\]

\[
= \Lambda_\phi (\Lambda_\phi + \sigma_n^2 I)^{-1} \overline{\phi}^{(U)}_{data}
\]

\[
\text{(since } UU^* = I \text{ due to unitary)}
\]

\[
\text{(since } U \text{ provides the PCA of } R_\phi \text{ then } \Lambda_\phi = U^* R_\phi U \text{ is a diagonal matrix)}
\]

diagonal matrix --> component-based computation

The solution \(\overline{\phi}^{opt,(U)}\) should be transformed back to the signal domain via

\[
\overline{\phi}_{est}^{opt} = U \overline{\phi}_{est}^{opt,(U)}
\]
Constrained Deconvolution: 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:

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

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

The task is to estimate the unknown signal $\phi$. 

236200, CS Department, Technion, Winter 2017-18
Constrained Deconvolution: Discrete Problem Settings

The degradation model:

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

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

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

is small.

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

For example:

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

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

The degradation model:

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

It's important to note the difference from the Wiener filter settings presented in the lecture, where the signal $\varphi$ 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 $||A\bar{\varphi}||_2^2$. 
Constrained Deconvolution: The Signal Estimate

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

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

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

leading to

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

We notice that

$$\|\varphi_{data} - H\varphi_{est}\|_2^2 \approx \||\bar{n}\|_2^2$$
**Constrained Deconvolution:**

**The Signal Estimate**

\[
\|\Phi_{\text{data}} - H\Phi_{\text{est}}\|_2^2 \approx \|\hat{n}\|_2^2 = \hat{n}^T \hat{n} = \sum_{i=1}^{N} n_i^2 \approx N\sigma_n^2
\]

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 \(x_1, \ldots, x_M\), 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 $\Phi_{est}$, we pose the (constrained) optimization problem as

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

having the \textbf{unconstrained} Lagrangian form of

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

The Lagrangian cost function to be minimized is defined as

$$\Psi(\varphi_{est}) = \| A \varphi_{est} \|_2^2 + \lambda \| \varphi_{data} - H \varphi_{est} \|_2^2$$

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

$$\frac{\partial}{\partial \varphi_{est}} \Psi(\varphi_{est}) = 0$$

We notice that

$$\Psi(\varphi_{est}) = \varphi_{est}^T A^T A \varphi_{est} + \lambda (\varphi_{data} - H \varphi_{est})^T (\varphi_{data} - H \varphi_{est})$$
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:

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

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

Using the above formulas we get

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

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

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

Setting the last result in \( \frac{\partial}{\partial \Phi_{est}} \Psi(\Phi_{est}) = 0 \)
gives \( 2(A^T A + \lambda H^T H)\Phi_{est} - 2\lambda H^T \Phi_{data} = 0 \)

and the optimal estimate is

\[
\Phi_{est}^{opt} = (A^T A + \lambda H^T H)^{-1} \lambda H^T \Phi_{data}
\]

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

Hence, we consider the optimal estimate as a function of \( \lambda \): \( \Phi_{est}^{opt}(\lambda) \)

Accordingly, we should find a \( \lambda \) that satisfies (or at least approximates)

\[
\| \Phi_{data} - H\Phi_{est}^{opt}(\lambda) \|_2^2 \approx N\sigma_n^2
\]

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

An appropriate $\lambda$ should be determined with respect to $\sigma_n^2$ (this procedure involves the given $\bar{\varphi}_{data}$). The filter matrix is independent of the given realization $\bar{\varphi}_{data}$.  

236200, CS Department, Technion, Winter 2017-18