Optimal and Adaptive Filtering

Murat Üney

M.Uney@ed.ac.uk

Institute for Digital Communications (IDCOM)



THE UNIVERSITY of EDINBURGH School of Engineering



27/06/2016

Murat Üney (IDCOM)

27/06/2016 1 / 69

- This presentation aims to provide an introductory level tutorial to optimal and adaptive filtering.
- The structures involved in optimal filtering problems (e.g., prediction, interpolation etc.) and adaptive solutions are highlighted while technical details of the theory are kept in a minimum.
- For example, for the sake of simplicity in technical discussions, we assume real valued random processes, i.e., if *x*(*n*) is a random process

$$x^*(n) = x(n)$$

throughout the presentation.

- Therefore, complex conjugations are omitted where appropriate, and, simplified mathematical expressions valid for the case of real valued sequences are used, for example, in complex spectra representations (Note that a real valued sequence has a complex valued transfer function).
- This is a living document the latest version of which can be downloaded from the UDRC Summer School 2016 website.
- Your feedback is always welcome.

Table of Contents

- Optimal Filtering
 - Optimal filter design
 - Application examples
 - Optimal solution: Wiener-Hopf equations
 - Example: Wiener equaliser

2 Adaptive filtering

- Introduction
- Recursive Least Squares Adaptation
- Least Mean Square Algorithm
- Applications
- Optimal signal detection
 - Application examples and optimal hypothesis testing
 - Additive white and coloured noise
- 4 Summary

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 2 / 69

- This presentation starts by introducing the problem definition for optimal filtering. Application examples follow this introduction.
- Wiener-Hopf equations are derived which characterise the solution of the problem. Then, the Wiener filter is introduced for both infinite impulse response (IIR) and finite impulse response FIR settings. Wiener channel equalisation is explained with an example.
- Adaptive filtering is introduced as an online and iterative strategy to optimal filtering. We emphasise that this strategy is useful especially when the statistical moments relevant to solving the optimal filtering problem are unknown and should be estimated from the incoming data and a training sequence.
- We derive the recursive least squares (RLS) and the least mean square (LMS) algorithms, and, compare them in an example. We provide system configurations for various applications of adaptive (optimal) filters.
- Finally, we give an overview of known signal detection in noise and relate the "matched filtering" technique to optimal hypothesis testing in a Bayesian sense.





- *y*(*n*): Observation related to a **stationary** signal of interest *x*(*n*).
- h(n): The impulse response of an LTI estimator.
- $\hat{x}(n)$: Estimate of x(n) given by

$$\hat{x}(n) = h(n) * y(n) = \sum_{i=-\infty}^{\infty} h(i)y(n-i).$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 3 / 69

- We observe a **stationary** sequence *y*(*n*) which contains information on a desired signal *x*(*n*) based on these observations (Fig. 1).
- The estimator we want to use is a linear time invariant (LTI) filter *h* characterised by its impulse response *h*(*n*).
- The output of this estimator is given by the convolution of its input with the impulse response *h*(*n*):

$$\hat{x}(n) = h(n) * y(n) = \sum_{i=-\infty}^{\infty} h(i)y(n-i).$$
 (1)





• Find *h*(*n*) with the best error performance:

$$e(n) = x(n) - \hat{x}(n) = x(n) - h(n) * y(n)$$

• The error performance is measured by the mean squared error (MSE)

$$\xi = E\left[\left(e(n)\right)^2\right].$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 4 / 69

 We would like to find h(n) that would generate an output as close to the desired signal x(n) as possible when driven by the input y(n). Let us define the estimation error by

$$e(n) = x(n) - \hat{x}(n) \tag{2}$$

- $\hat{x}(n)$ is stationary owing to that the estimator is LTI, and, its input y(n) is stationary. Therefore, the error sequence e(n) is also stationary.
- Because *e*(*n*) is stationary, it can be characterised by the expectation of its square at any time step *n*, or, the mean squared error (MSE):

$$\xi \triangleq E\left[\left(e(n)\right)^2\right] = E\left[\left(x(n) - \sum_{i=-\infty}^{\infty} h(i)y(n-i)\right)^2\right].$$
 (3)



Figure 1: Optimal filtering scenario.

• The MSE is a function of h(n), i.e.,

$$\mathbf{h} = [\cdots, h(-2), h(-1), h(0), h(1), h(2), \cdots]$$

$$\xi(\mathbf{h}) = E\left[\left(e(n)\right)^2\right] = E\left[\left(x(n) - h(n) * y(n)\right)^2\right]$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 5 / 69

• Note that, the MSE is a function of the estimator impulse response. This point becomes more clear after the error term *e*(*n*) is fully expanded to its components.

$$\xi(h) = E\left[\left(e(n)\right)^2\right] = E\left[\left(x(n) - h(n) * y(n)\right)^2\right].$$

- It is useful to use a vector-matrix notation to cast the filter design problem as an optimisation problem.
- Consider the impulse response as a vector, i.e.,

$$\mathbf{h} = [\cdots, h(-2), h(-1), h(0), h(1), h(2), \cdots]^T$$



Figure 1: Optimal filtering scenario.

• The MSE is a function of *h*(*n*), i.e.,

$$\mathbf{h} = [\cdots, h(-2), h(-1), h(0), h(1), h(2), \cdots]$$

$$\xi(\mathbf{h}) = E\left[\left(e(n)\right)^2\right] = E\left[\left(x(n) - h(n) * y(n)\right)^2\right]$$

• Thus, optimal filtering problem is

$$\mathbf{h}_{opt} = \arg\min_{\mathbf{h}} \xi(\mathbf{h})$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 5 / 69

- Let **x** and **e** denote the desired signal vector and the error vector constructed in a similar fashion, respectively. Then, $\mathbf{e} = \mathbf{x} - \mathbf{Y}\mathbf{h}$ expanded as

[:]]	[:]		[:]	÷	÷	÷	÷	:]	[:]
<i>e</i> (0)		<i>x</i> (0)			<i>y</i> (0)	<i>y</i> (-1)	<i>y</i> (-2)	<i>y</i> (-3)		h(0)
<i>e</i> (1)		<i>x</i> (1)			<i>y</i> (1)	<i>y</i> (0)	<i>y</i> (-1)	<i>y</i> (-2)		<i>h</i> (1)
<i>e</i> (2)	-	x(2)	_		<i>y</i> (2)	<i>y</i> (1)	<i>y</i> (0)	<i>y</i> (-1)		h(2)
<i>e</i> (3)		<i>x</i> (3)			<i>y</i> (3)	y(2)	<i>y</i> (1)	<i>y</i> (0)		h(3)
[:.				L:	÷	÷	÷	÷	:]	
≜e		≜x		≜ Y :	Convolut	ion (or, data) matrix of y	which is Toe	eplitz.	≜h (4)

 Optimal filtering problem is the problem of finding h that leads to the minimum *ξ*. Equivalently, we want to solve the following optimisation problem:

$$\mathbf{h}_{opt} = \arg\min_{\mathbf{h}} \xi(\mathbf{h}). \tag{5}$$

1) Prediction, interpolation and smoothing of signals



Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 6 / 69

- A mile stone in optimal linear estimation was a World War II-time classified report by Norbert Wiener (1894–1964), a celebrated American mathematician and philosopher. This report was published in 1949 as a monograph.
- The picture is the cover of one of the (300) original copies of this report sold for \$7200 by an auction house.
- A review of this book by J.W. Tukey published in the *Journal of the American Statistical Association* in 1952 mentions that

...Wiener's report... was followed by a host (at least a dozen to my knowledge) of similarly classified "simplifications" or "explanations" of the procedure...



1) Prediction, interpolation and smoothing of signals



Prediction for anti-aircraft fire control.

Murat Üney (IDCOM) Optimal and Adaptive Filtering 27/06/2016 7 / 69

- Here, y(n) is the noisy measurements of a shifted version of the desired signal x'(n). The noise ω(n) is also stationary.
- If the shift *d* is a positive integer, then, the optimal filter h_{opt} is the best linear *d*-step predictor of *x*(*n*). For example, if *d* = 1, then, h_{opt} is a one-step predictor.
- If the shift is a negative integer, then, the optimal filter performs smoothing. For example, if *d* = −1, **h**_{opt} is the best linear one-lag smoother.
- For a rational *d*, the optimal filter is an interpolator aiming to estimate the (missing) sample between two consecutive data points. For example, for d = -1/2, the optimal filter is an interpolator trying to estimate the (missing) sample between x(n) and x(n-1).
- Wiener's work was (partly) motivated by the prediction problem for anti-aircraft fire control. The aircraft's bearing (and altitude) is tracked manually, i.e., x(n) and $y(n) = x(n) + \omega(n)$ are collected. *d* is selected in accordance with the flight time of anti-aircraft shells (which could be > 20s) and the guns are pointed towards the predicted location.

1) Prediction, interpolation and smoothing of signals



Signal denoising applications, estimation of missing data points.

```
Murat Üney (IDCOM)
```

Optimal and Adaptive Filtering

27/06/2016 8 / 69

- Linear predictive coding (LPC) of waveforms is another popular application of optimal predictors. Here, **h**_{opt} is used as an encoding of *x*(*n*). Non-zero *h*_{opt}(*n*) values provide a lossy compression of *x*(*n*), in this respect.
- Signal denoising applications benefit from **h**_{opt} designed for smoothing.
- Interpolation is used when estimating missing data points.

2) System identification



Figure 2: System identification using a training sequence t(n) from an ergodic and stationary ensemble.

Optimal and Adaptive Filtering

27/06/2016 9 / 69

- The optimal filtering framework can be used to solve system identification problems.
- Here, the system to be identified is *g*(*n*). First, a training sequence *t*(*n*) is generated to drive the system. *t*(*n*) is an instance from an independent and identically distributed (i.i.d) process, e.g., a white noise sequence. Thus, its time averages matches its ensemble averages (first and second order moments).
- The output of the system to this input is used as the desired signal in the optimal filtering problem.
- The optimal filter h(n) that produces an output $\hat{x}(n)$ which is closest to x(n) when driven by t(n) will be the best linear time invariant approximation of g(n).
- One application of this design setting is echo cancellation in full duplex data transmission. For example, line modems including v.32 (ITU-T recommendation v.32 –https://www.itu.int/rec/T-REC-V.32-199303-I/en) identify the "echo path" during the modem hand-shake protocol. Echo cancellation involves synthesising the echo signal and subtraction from the receiver front-end signal, thereby isolating the signal transmitted from the remote modem.

2) System identification



Figure 2: System identification using a training sequence t(n) from an ergodic and stationary ensemble.



- The optimal filtering framework can be used to solve system identification problems.
- Here, the system to be identified is g(n). First, a training sequence t(n) is generated to drive the system. t(n) is an instance from an independent and identically distributed (i.i.d) process, e.g., a white noise sequence. Thus, its time averages matches its ensemble averages (first and second order moments).
- The output of the system to this input is used as the desired signal in the optimal filtering problem.
- The optimal filter h(n) that produces an output $\hat{x}(n)$ which is closest to x(n) when driven by t(n) will be the best linear time invariant approximation of g(n).
- One application of this design setting is echo cancellation in full duplex data transmission. For example, line modems including v.32 (ITU-T recommendation v.32 –https://www.itu.int/rec/T-REC-V.32-199303-I/en) identify the "echo path" during the modem hand-shake protocol. Echo cancellation involves synthesising the echo signal and subtraction from the receiver front-end signal, thereby isolating the signal transmitted from the remote modem.

3) Inverse System identification



Figure 3: Inverse system identification using x(n) as a training sequence.

Murat Line	
marat one	

Optimal and Adaptive Filtering

27/06/2016 10 / 69

- The optimal filtering framework can be used to find the inverse of a system: In the block diagram, the system to be inverted is *g*(*n*) which models, for example, a channel that distorts the desired signal *x*(*n*). The receiver side observes *y*(*n*), which is a noisy version of the distorted signal.
- We would like to design a filter h(n) which mitigates effects of g(n) and rejects the noise $\omega(n)$ optimally, in order to restore x(n).
- To do that, a training sequence t(n) which is known at both the transmitter and receiver ends drives the channel. This sequence is an instance from an ergodic and stationary ensemble, i.e., t(n) is randomly generated such that its time statistics match the ensemble averages of x(n).
- Thus, the receiver side can find the estimation error *e*(*n*) corresponding to any **h** by

$$e(n) = x(n) - \hat{x}(n) = x(n) - t(n)$$

• The *h*(*n*) that minimises this error is the best linear inverse system. Therefore, the optimal filter design framework can be used for finding inverse systems.

3) Inverse System identification



Figure 3: Inverse system identification using x(n) as a training sequence.

Channel equalisation in digital communication systems.



- In a telecommunications context, the channel inverse is known as the "equaliser."
- In the figure, the time-frequency plot of the transmission line signal during the hand-shake of two v.32 modems is seen (source: http://www.windytan.com).
- At the right hand side of the figure, both modems send scrambled data to each other (the time window in the blue and red boxes, respectively) allowing the receiver side to find the channel inverse. The data sequence is known at both sides and described in the International Telecommunications Union (ITU) standard recommendation v.32.

Optimal solution: Normal equations • Consider the MSE $\xi(\mathbf{h}) = E\left[(e(n))^2\right]$

• The optimal filter satisfies $\nabla \xi(\mathbf{h})|_{h_{opt}} = \mathbf{0}$. Equivalently, for all $j = \dots, -2, -1, 0, 1, 2, \dots$

$$\begin{aligned} \frac{\partial \xi}{\partial h(j)} &= E\left[2e(n)\frac{\partial e(n)}{\partial h(j)}\right] \\ &= E\left[2e(n)\frac{\partial \left(x(n) - \sum_{i=-\infty}^{\infty} h(i)y(n-i)\right)}{\partial h(j)}\right] \\ &= E\left[2e(n)\frac{\partial \left(-h(j)y(n-j)\right)}{\partial h(j)}\right] \\ &= -2E\left[e(n)y(n-j)\right] \end{aligned}$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 11 / 69

- The optimisation problem for finding the optimal filter has an objective function the MSE of estimation–, which is quadratic in the unknowns.
- Hence, a unique solution exists which can be characterised by the gradient of the objective the vector of partial derivatives of the objective with respect to the unknowns. At the optimal point, the gradient equals to the zero vector.
- In the first step, the differentiation is moved into the expectation since expectation is a linear operator. In the following steps, well known rules of differentiation are used.
- Note that, we can evaluate the gradient for any given **h**, and the error e(n) inside the expectation corresponds to the chosen filter **h**.

Optimal solution: Normal equations

- Consider the MSE $\xi(\mathbf{h}) = E\left[\left(e(n)\right)^2\right]$
- The optimal filter satisfies $\nabla \bar{\xi}(\mathbf{h})|_{h_{opt}} = \mathbf{0}$. Equivalently, for all $j = \dots, -2, -1, 0, 1, 2, \dots$

$$\begin{aligned} \frac{\partial \xi}{\partial h(j)} &= E\left[2e(n)\frac{\partial e(n)}{\partial h(j)}\right] \\ &= E\left[2e(n)\frac{\partial \left(x(n) - \sum_{i=-\infty}^{\infty} h(i)y(n-i)\right)}{\partial h(j)}\right] \\ &= E\left[2e(n)\frac{\partial \left(-h(j)y(n-j)\right)}{\partial h(j)}\right] \\ &= -2E\left[e(n)y(n-j)\right] \end{aligned}$$

• Hence, the optimal filter solves the "normal equations"

$$E[e(n)y(n-j)] = 0, j = \dots, -2, -1, 0, 1, 2, \dots$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 11 / 69

 In order to find the optimal filter, we use its characterisation through its gradient, i.e., the optimal filter solves the set of equations

$$E[e(n)y(n-j)] = 0, j = \dots, -2, -1, 0, 1, 2, \dots$$

which are known as "the normal equations."

Optimal solution: Wiener-Hopf equations

• The error of h_{opt} is orthogonal to its observations, i.e., for all $j \in \mathbb{Z}$

 $E\left[e_{opt}(n)y(n-j)\right]=0$

which is known as "the principle of orthogonality".

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 12 / 69

- Because the optimal filter solves the normal equations, its error $e_{opt}(n)$ satisfies the statistical orthogonality condition with the input variables y(n-j) for j = ..., -2, -1, 0, 1, 2, ...
- The geometric interpretion of the normal equations follows the statistical norm of the desired signal expressed in terms of the optimal estimate and the associated error:

$$\langle \boldsymbol{x}(n), \boldsymbol{x}(n) \rangle \triangleq E\left[(\hat{\boldsymbol{x}}(n))^2 \right]$$

$$= E\left[(\hat{\boldsymbol{x}}_{opt}(n) + \boldsymbol{e}_{opt}(n))^2 \right]$$

$$= E\left[(\hat{\boldsymbol{x}}_{opt}(n))^2 + 2\hat{\boldsymbol{x}}_{opt}(n)\boldsymbol{e}_{opt}(n) + (\boldsymbol{e}_{opt}(n))^2 \right]$$

$$= E\left[(\hat{\boldsymbol{x}}_{opt}(n))^2 \right] + E\left[(\boldsymbol{e}_{opt}(n))^2 \right]$$

$$+ 2\sum_{i=-\infty}^{\infty} h_{opt}(i) \underbrace{E\left[\boldsymbol{e}_{opt}(n)\boldsymbol{y}(n-i) \right]}_{=0 \text{ by the principle of orthogonality}}$$

$$= E\left[(\hat{\boldsymbol{x}}_{opt}(n))^2 \right] + E\left[(\boldsymbol{e}_{opt}(n))^2 \right]$$

$$= \langle \hat{\boldsymbol{x}}_{opt}(n), \hat{\boldsymbol{x}}_{opt}(n) \rangle + \langle \boldsymbol{e}_{opt}(n), \boldsymbol{e}_{opt}(n) \rangle.$$

$$(6)$$

• Thus, the optimal estimate and the associated error are orthogonal and follow a Pythagorean relation with the desired signal *x*(*n*).

Optimal solution: Wiener-Hopf equations

• The error of h_{opt} is orthogonal to its observations, i.e., for all $j \in \mathbb{Z}$

$$E\left[e_{opt}(n)y(n-j)\right]=0$$

which is known as "the principle of orthogonality".

• Furthermore,

$$E\left[e_{opt}(n)y(n-j)\right] = E\left[\left(x(n) - \sum_{i=-\infty}^{\infty} h_{opt}(i)y(n-i)\right)y(n-j)\right]$$
$$= E\left[x(n)y(n-j)\right] - \sum_{i=-\infty}^{\infty} h_{opt}(i)E\left[y(n-i)y(n-j)\right] = 0$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 12 / 69

- We expand the optimal error term $e_{opt}(n)$ inside the expection.
- After distributing y(n − j) over the summation, and, using the linearity of the expectation operator, we obtain the last line which equals to zero by the principle of orthogonality.

Optimal solution: Wiener-Hopf equations

• The error of h_{opt} is orthogonal to its observations, i.e., for all $j \in \mathbb{Z}$

$$E\left[e_{opt}(n)y(n-j)\right]=0$$

which is known as "the principle of orthogonality".

Furthermore,

$$E[e_{opt}(n)y(n-j)] = E\left[\left(x(n) - \sum_{i=-\infty}^{\infty} h_{opt}(i)y(n-i)\right)y(n-j)\right]$$
$$= E[x(n)y(n-j)] - \sum_{i=-\infty}^{\infty} h_{opt}(i)E[y(n-i)y(n-j)] = 0$$

Result (Wiener-Hopf equations)

$$\sum_{i=-\infty}^{\infty} h_{opt}(i) r_{yy}(i-j) = r_{xy}(j)$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 12 / 69

 We obtain the Wiener-Hopf equations after carrying the summation term to the right hand side of the equation above and realising that

$$r_{xy}(j) = E\left[x(n)y(n-j)\right]$$

and

$$r_{yy}(i-j) = E\left[y(n-i)y(n-j)\right]$$

- Note that we consider real valued sequences throughout this presentation and omit complex conjugations above.
- When the equalities above are used together with the symmetricity of auto-correlation $r_{yy}(n) = r_{yy}(-n)$, the Weiner-Hopf equations can be written simply as

$$h_{opt}(n) * r_{yy}(n) = r_{xy}(n)$$

where we have the convolution of the optimal filter with the auto-correlation function of the observations on the left hand side, and, the cross-correlation sequence on the right hand side.

The Wiener filter

• Wiener-Hopf equations can be solved indirectly, in the complex spectral domain:

$$h_{opt}(n) * r_{yy}(n) = r_{xy}(n) \leftrightarrow H_{opt}(z)P_{yy}(z) = P_{xy}(z)$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 13 / 69

- We have not placed any constraints on the optimal filter so as to guarantee an impulse repsonse which can be finitely parameterised.
- Therefore, it is more convenient to consider an indirect characterisation of the optimal impulse response provided by the complex spectral domain (or, the z-transform domain).
- Let us consider the z-transform domain representation of the Wiener-Hopf equations.
- The multiplication of *H*_{opt}(*z*) with the power spectral density (PSD) of the input equals to the complex spectra of the cross-correlation sequence.

The Wiener filter

• Wiener-Hopf equations can be solved indirectly, in the complex spectral domain:

$$h_{opt}(n) * r_{yy}(n) = r_{xy}(n) \leftrightarrow H_{opt}(z)P_{yy}(z) = P_{xy}(z)$$

Result (The Wiener filter)

$$H_{opt}(z) = rac{P_{xy}(z)}{P_{yy}(z)}$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 13 / 69

- The optimal filter is obtained in the complex spectral domain by the division of the cross-correlation complex spectra with the PSD of the input.
- The impulse response *h*_{opt}(*n*) can be found, in principle, using the inverse z-transform:

$$h_{opt}(n) = \frac{1}{2\pi j} \oint_C H_{opt}(z) z^{n-1} \mathrm{d}z.$$
 (6)

The Wiener filter

 Wiener-Hopf equations can be solved indirectly, in the complex spectral domain:

$$h_{opt}(n) * r_{yy}(n) = r_{xy}(n) \leftrightarrow H_{opt}(z)P_{yy}(z) = P_{xy}(z)$$

Result (The Wiener filter)

$$H_{opt}(z) = rac{P_{xy}(z)}{P_{yy}(z)}$$

 The optimal filter has an infinite impulse response (IIR), and, is non-causal, in general.

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 13 / 69

- The region of convergence (ROC) of H_{opt}(z) is not necessarily the outer region of a circle centered at the origin. Correspondingly, h_{opt}(n) is not necessarily a right sided (causal) sequence.
- We assume that the processes we consider are regular, hence, the ROC of $H_{opt}(z)$ contain the unit circle on the z-plane. Correspondingly, $h_{opt}(n)$ is a stable sequence.
- For a process to be regular, its PSD $P(z = e^{j\omega})$ should not have extended regions along ω where it is zero.
- For the case, it can be shown that P(z) can be factorised as

$$P(z) = \sigma^2 Q(z) Q^*(1/z^*)$$

where Q(z) is a minimum-phase (causal) sequence and $Q^*(1/z^*)$ is its anti-causal counterpart.

• The task of identification of Q(z) given P(z) (and σ^2) is referred to as "spectral factorisation".

Causal Wiener filter

- We project the unconstrained solution $H_{opt}(z)$ onto the set of causal and stable IIR filters by a two step procedure:
- First, factorise $P_{yy}(z)$ into causal (right sided) $Q_{yy}(z)$, and anti-causal (left sided) parts $Q_{yy}^*(1/z^*)$, i.e., $P_{yy}(z) = \sigma_y^2 Q_{yy}(z) Q_{yy}^*(1/z^*)$.
- Select the causal (right sided) part of $P_{xy}(z)/Q_{yy}^*(1/z^*)$.

Result (Causal Wiener filter)

$$\mathcal{H}_{opt}^{+}(z) = \frac{1}{\sigma_y^2 Q_{yy}(z)} \left[\frac{P_{xy}(z)}{Q_{yy}^*(1/z^*)} \right]_+$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 14 / 69

- The "unconstrained" solution of the optimal filtering problem can be projected onto the space of causal IIR filters in a two step procedure.
- First, the causal and anti-causal factors of $P_{yy}(z)$ are identified. This factorisation splits the optimal filter as a cascade of two filters; one causal system followed by a non-causal one.
- The causal filter is then characterised by $H_{opt,1}(z) = \frac{1}{\sigma_y^2 Q_{yy}(z)}$.
- The second system has a non-causal impulse response (both left and right sided) with comlex spectra $H_{opt,2}(z) = \frac{P_{xy}(z)}{Q_{yy}^*(1/z^*)}$.
- Let $h_{opt,2}(n)$ denote the corresponding sequence.
- In order to find the projection of the optimal filter onto the space of causal IIR filters, the second filter is selected as the right sided part of $h_{opt,2}(n)$, i.e., $h'_{opt,2}(n) = h_{opt,2}(n)$, for n = 0, 1, 2, ... and $h'_{opt,2}(n) = 0$, otherwise.
- This is often carried out in the spectral domain using partial fraction expansion.

FIR Wiener-Hopf equations



Figure 4: A finite impulse response (FIR) estimator.

• Wiener-Hopf equations for the FIR optimal filter of N taps:



- Finite impulse response (FIR) filters are stable.
- FIR filters are causal without loss of generality, in that, they can always be cascaded to a delay line z^{-d} to have a causal overall response, where *d* is the length of the left sided part of the FIR impulse response.
- It is helpful for the designer to restrict the optimisation problem such that the space of LTI systems is constrained to the space of FIR filters as they naturally admit a finite parameterisation – N unknowns for an N-tap FIR filter.

FIR Wiener Filter

• FIR Wiener-Hopf equations in vector-matrix form.

 $\begin{bmatrix} r_{yy}(0) & r_{yy}(1) & \dots & r_{yy}(N-1) \\ r_{yy}(1) & r_{yy}(0) & \dots & r_{yy}(N-2) \\ \vdots & \vdots & \vdots & \vdots \\ r_{yy}(N-1) & r_{yy}(N-2) & \dots & r_{yy}(0) \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(N-1) \end{bmatrix} = \begin{bmatrix} r_{xy}(0) \\ r_{xy}(1) \\ \vdots \\ r_{xy}(N-1) \end{bmatrix}$



- FIR Wiener-Hopf equations specify a system of *N* equations in *N* unknowns.
- In order to solve this system, it is useful to consider the corresponding algebraic form.
- In this presentation, we assume real valued stationary processes. In the case of complex valued stationary processes, r_{xy} has r^{*}_{xy}(*I*) for *I* = 0, 1, 2, ..., *N* 1 in its fields.
- Similarly \mathbf{R}_{yy} is conjugate transpose symmetric (or, Hermitian symmetric). For example, the first column of \mathbf{R}_{yy} has $r_{yy}(0), r_{yy}^*(1), \ldots, r_{yy}^*(N-1)$ in its fields.

FIR Wiener Filter

• FIR Wiener-Hopf equations in vector-matrix form.

 $\begin{bmatrix} r_{yy}(0) & r_{yy}(1) & \dots & r_{yy}(N-1) \\ r_{yy}(1) & r_{yy}(0) & \dots & r_{yy}(N-2) \\ \vdots & \vdots & \vdots & \vdots \\ r_{yy}(N-1) & r_{yy}(N-2) & \dots & r_{yy}(0) \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(N-1) \end{bmatrix} = \begin{bmatrix} r_{xy}(0) \\ r_{xy}(1) \\ \vdots \\ r_{xy}(N-1) \end{bmatrix}$ ≜h_{opt}

 $\triangleq \mathbf{R}_{VV}$: Autocorrelation matrix of y(n) which is Toeplitz.

Result (FIR Wiener filter)

 $\mathbf{h}_{opt} = \mathbf{R}_{vv}^{-1} \mathbf{r}_{xv}$

Murat Üney (IDCOM)

27/06/2016 16/69

MSE surface

• MSE is a quadratic function of **h**



Figure 5: For a 2-tap Wiener filtering example: (a) the MSE surface, (b) gradient vectors.

• The MSE can be found as

$$\xi(\mathbf{h}) = \mathbf{h}^{T} \mathbf{R}_{yy} \mathbf{h} - 2\mathbf{h}^{T} \mathbf{r}_{xy} + E\left[(x(n))^{2}\right]$$
(7)

which can be written in the following quadratic form:

$$\xi(\mathbf{h}) = (\mathbf{h} - \mathbf{h}_{opt})^T \mathbf{R}_{yy} (\mathbf{h} - \mathbf{h}_{opt}) + \xi(\mathbf{h}_{opt})$$

$$\xi(\mathbf{h}_{opt}) = E\left[(x(n))^2\right] - \mathbf{h}_{opt}^T \mathbf{r}_{xy}$$
(8)

- As ξ(h) is a quadratic function of h, it is smooth (its gradient is defined for all values of h) and has a unique minimum.
- In the case of 2-dimensional h, equal MSE lines are ellipses whose centre is the optimal filter vector and axes are along the eigenvectors of R_{yy}. The major and minor semi-axis lengths are specified by the eigenvalues.

Example: Wiener equaliser



Figure 6: (a) The Wiener equaliser. (b) Alternative formulation.

```
Murat Üney (IDCOM) Optimal and Adaptive Filtering 27/06/2016 18 / 69
```

- In this example, we consider optimal filtering for inverse system identification.
- A white random signal x(n) is transmitted through a communication channel which distorts the signal with the transfer function C(z).
- The receiver front-end receives noisy versions of the distorted signal.
- The goal of the equaliser is to optimally denoise the received signal y(n) and mitigate the effects of distortion.



Figure 7: Channel equalisation scenario.

• For notational convenience define:

$$x'(n) = x(n-d)$$

 $e'(n) = x(n-d) - \hat{x}(n-d)$ (9)

• Label the output of the channel filter as y'(n) where

$$y(n) = y'(n) + \eta(n)$$

|--|

- Let us define the desired signal as a delayed version of *x*(*n*). Hence, the estimation error at time *n* will be *e*'(*n*) as defined above.
- The channel output without noise is also a sequence with distinct properties, so, we will label it as y'(n).



Figure 7: Channel equalisation scenario.



Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 20 / 69

• Let us find the fields of the input autocorrelation matrix \mathbf{R}_{yy} and the cross correlation vector $\mathbf{r}_{x'y}$.



Figure 7: Channel equalisation scenario.

• Wiener filter

$$\mathbf{h}_{opt} = \mathbf{R}_{yy}^{-1} \mathbf{r}_{x'y} \tag{10}$$

• The
$$(i, j)$$
th entry in \mathbf{R}_{yy} is

$$\begin{aligned} r_{yy}(j-i) &= E\left[y(j)y(i)\right] \\ &= E\left[(y'(j) + \eta(j))(y'(i) + \eta(i))\right] \\ &= r_{y'y'}(j-i) + \sigma_{\eta}^2 \delta(j-i) \\ &\leftrightarrow P_{yy}(z) = P_{y'y'}(z) + \sigma_{\eta}^2 \end{aligned}$$
Murat Üney (IDCOM) Optimal and Adaptive Filtering 27/06/2016 20/69

- The input autocorrelation of the equaliser is the sum of the autocorrelation of the channel output and that of the noise sequence.
- Since the noise sequence is white, its autcorrelation is Dirac's delta function weighted by the variance of the noise. The corresponding complex spectra equals to this variance for all *z*.



Figure 7: Channel equalisation scenario.

• Remember y'(n) = c(n) * x(n)

$$\leftrightarrow r_{y'y'} = c(n) * c(-n) * r_{xx}(n) \leftrightarrow P_{y'y'}(z) = c(z)c(z^{-1})P_{xx}(z)$$

• Consider a white data sequence *x*(*n*), i.e.,

$$\sigma_{xx}(n) = \sigma_x^2 \delta(n) \leftrightarrow \mathcal{P}_{xx}(z) = \sigma_x^2.$$

Then, the complex spectra of the autocorrelation sequence of interest is

$$P_{yy}(z) = P_{y'y'}(z) + \sigma_x^2 = C(z)C(z^{-1})\sigma_x^2 + \sigma_\eta^2$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 21 / 69

• Let us find the autocorrelation of the channel output in terms of that of the channel input *x*(*n*) and the channel transfer function.



Figure 7: Channel equalisation scenario.

• Wiener filter

$$\mathbf{h}_{opt} = \mathbf{R}_{yy}^{-1} \mathbf{r}_{x'y} \tag{11}$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 22 / 69

- We have found the complex spectra of the sequence that specifies **R**_{yy}.
- Now, let us consider $\mathbf{r}_{x'y}$.



Figure 7: Channel equalisation scenario.

• Wiener filter

$$\mathbf{h}_{opt} = \mathbf{R}_{yy}^{-1} \mathbf{r}_{x'y} \tag{11}$$

• The (j)th entry in
$$\mathbf{r}_{x'y}$$
 is

$$\begin{aligned} r_{x'y}(j) &= E\left[x'(n)y(n-j)\right] \\ &= E\left[x(n-d)(y'(n-j)+\eta(n-j))\right] \\ &= r_{xy'}(j-d) \\ &\leftrightarrow P_{x'y}(z) = P_{xy'}(z)z^{-d} \end{aligned}$$
(12)
Mutat Unev. (IDCOM) Optimal and Adaptive Filtering 27/06/2016 22/69

- We have found the complex spectra of the sequence that specifies **r**_{x'y} in terms of *P*_{xy'}.
- Next, we specify this cross correlation sequence.



Figure 7: Channel equalisation scenario.

• Remember
$$y'(n) = c(n) * x(n)$$

$$\leftrightarrow r_{xy'} = c(-n) * r_{xx}(n) \leftrightarrow P_{xy'}(z) = C(z^{-1})P_{xx}(z)$$

• Then, the complex spectra of the cross correlation sequence of interest is

$$P_{x'y}(z) = P_{xy'}(z)z^{-d} = \sigma_x^2 C(z^{-1})z^{-d}$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 23 / 69

• We have found the complex spectra of the sequence of concern, in terms of the input auto-correlation and the channel transfer function.

- Suppose that $\mathbf{c} = [c(0) = 0.5, c(1) = 1]^T \leftrightarrow C(z) = (0.5 + z^{-1})$
- Then,

$$P_{yy}(z) = C(z)C(z^{-1})\sigma_x^2 + \sigma_\eta^2 = (0.5 + z^{-1})(0.5 + z)\sigma_x^2 + \sigma_\eta^2$$

$$P_{x'y}(z) = \sigma_x^2 C(z^{-1})z^{-d} = (0.5z^{-d} + z^{-d+1})\sigma_x^2$$

• Suppose that d = 1, $\sigma_x^2 = 1$, and, $\sigma_\eta^2 = 0.1$

$$r_{yy}(0) = 1.35, r_{yy}(1) = 0.5, \text{ and } r_{yy}(2) = 0$$

 $r_{x'y}(0) = 1, r_{x'y}(1) = 0.5, \text{ and } r_{x'y}(2) = 0$

The Wiener filter is obtained as

$$\mathbf{h}_{opt} = \left(\begin{bmatrix} 1.35 & 0.5 & 0\\ 0.5 & 1.35 & 0.5\\ 0 & 0.5 & 1.35 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1\\ 0.5\\ 0 \end{bmatrix} = \begin{bmatrix} 0.69\\ 0.13\\ -0.05 \end{bmatrix}$$

• The MSE is found as $\xi(\mathbf{h}_{opt}) = \sigma_x^2 - \mathbf{h}_{opt}^T \mathbf{r}_{x'y} = 0.24$.

```
Murat Üney (IDCOM)
```

Optimal and Adaptive Filtering

27/06/2016 24 / 69

- Since we have found all the required quantities that specify Wiener FIR filter in terms of the complex spectra of the input autocorrelation and the transfer function, we can solve the optimal filtering problem for any selection of these functions.
- An example channel response is given in the slide.
- The solution follows trivially from our previous derivations.

Adaptive filtering - Introduction



Figure 8: Adaptive filtering configuration.

• For notational convenience, define

 $\mathbf{y}(n) \triangleq [y(n), y(n-1), \dots, y(n-N+1)]^T$, $\mathbf{h}(n) \triangleq [h_0, h_1, \dots, h_{N-1}]^T$

• The output of the adaptive filter is

$$\hat{x}(n) = \mathbf{h}^T(n)\mathbf{y}(n)$$

Optimum solution

$$\mathbf{h}_{opt} = \mathbf{R}_{yy}^{-1} \mathbf{r}_{xy}$$

Murat Üney (IDCOM) Optimal and Adaptive Filtering 27/06/2016 25 / 69

- In our previous treatment, the optimal filter design was an offline procedure. The filter works in an open-loop fashion, without any mechanism to compansate for any changes in the second order statistics used for design, during its operation.
- Adaptive filters provide a feedback mechanism to adjust the filter to the actual working conditions. Hence, they are closed-loop systems.
- They can be viewed as strategies to find the optimal filter online, in an iterative fashion, during the operation of the filter.
Recursive least squares

Minimise cost function

$$\xi(n) = \sum_{k=0}^{n} (x(k) - \hat{x}(k))^2$$
(13)

Solution

$$\mathbf{R}_{yy}(n)\mathbf{h}(n)=\mathbf{r}_{xy}(n)$$

LS "autocorrelation" matrix

$$\mathbf{R}_{yy}(n) = \sum_{k=0}^{n} \mathbf{y}(k) \mathbf{y}^{\mathsf{T}}(k)$$

LS "cross-correlation" vector

$$\mathbf{r}_{xy}(n) = \sum_{k=0}^{n} \mathbf{y}(k) x(k)$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 26 / 69

- In optimal filtering, we considered MSE as the cost function to be minimised.
- Let us choose a cost function which does not involve expectations and can be computed using the desired signal -or, the training sequencex(n) and its estimates x̂(n).
- The sum of squared error terms over time (13) is such a cost function.
- Let us use the notation $\mathbf{e}(n) = [e(0), e(1), \dots, e(n)]^T$. It can easily be seen that $\xi(n) = \|\mathbf{e}(n)\|^2 = \mathbf{e}(n)^T \mathbf{e}(n)$.
- The LS "autocorrelation" matrix and "cross-correlation" vector can be found after expanding the error vector in the form given in Eq.(4), and, taking the gradient of this expression with respect to **h**(*n*):

$$\mathbf{e} = \mathbf{x} - \mathbf{Y}\mathbf{h}$$

 $abla_{\mathbf{h}}\mathbf{e}^{\mathsf{T}}\mathbf{e} = -2\mathbf{x}^{\mathsf{T}}\mathbf{Y} + 2\mathbf{Y}^{\mathsf{T}}\mathbf{Y}\mathbf{h}$

• Moreover, if the signals involved are ergodic (i.e., if the time averages and the ensemble averages are the same), then

$$\xi(\mathbf{h}) = E\left[(e(n))^2\right] = \lim_{N \to \infty} \frac{1}{N+1} \sum_{k=0}^N (x(k) - \hat{x}(k))^2.$$

Recursive least squares

• Recursive relationships

$$\mathbf{R}_{yy}(n) = \mathbf{R}_{yy}(n-1) + \mathbf{y}(n)\mathbf{y}^{T}(n)$$

$$\mathbf{r}_{xy}(n) = \mathbf{r}_{xy}(n-1) + \mathbf{y}(n)\mathbf{x}(n)$$

• Substitute for **r**_{xy}

$$\mathbf{R}_{yy}(n)\mathbf{h}(n) = \mathbf{R}_{yy}(n-1)\mathbf{h}(n-1) + \mathbf{y}(n)x(n)$$

• Replace $\mathbf{R}_{yy}(n-1)$

$$\mathbf{R}_{yy}(n)\mathbf{h}(n) = \left(\mathbf{R}_{yy}(n) - \mathbf{y}(n)\mathbf{y}^{T}(n)\right)\mathbf{h}(n-1) + \mathbf{y}(n)x(n)$$

• Multiple both sides by $\mathbf{R}_{yy}^{-1}(n)$

$$\mathbf{h}(n) = \mathbf{h}(n-1) + \mathbf{R}_{yy}^{-1}(n)\mathbf{y}(n)\mathbf{e}(n)$$
$$\mathbf{e}(n) = x(n) - \mathbf{h}^{T}(n-1)\mathbf{y}(n)$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 27 / 69

• Let us consider the recursive relationships that will allow us to derive online update rules for the filter coefficients.

Recursive least squares

Recursive relationships

$$\mathbf{R}_{yy}(n) = \mathbf{R}_{yy}(n-1) + \mathbf{y}(n)\mathbf{y}^{\mathsf{T}}(n)$$

• Apply Sherman-Morrison identity

$$\mathbf{R}_{yy}^{-1}(n) = \mathbf{R}_{yy}^{-1}(n-1) - \frac{\mathbf{R}_{yy}^{-1}(n-1)\mathbf{y}(n)\mathbf{y}^{T}(n)\mathbf{R}_{yy}^{-1}(n-1)}{1 + \mathbf{y}^{T}(n)\mathbf{R}_{yy}^{-1}(n-1)\mathbf{y}(n)}$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 28 / 69

- The inverse of the autocorrelation matrix at time *n* can further be found in terms of the previous inverse, and the current observation vector.
- Sherman-Morrison identity gives the inverse of a matrix which can be written as the sum of a matrix with a vector outer product.

Summary



Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 29 / 69

• The steps of the resulting algorithm are given above.

Stochastic gradient algorithms

• MSE contour - 2-tap example:



Figure 9: Method of steepest descent.

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 30 / 69

- Another approach that would iteratively converge to the optimal filter would draw from the descent directions approaches in the optimisation literature.
- A well known procedure steepest descent starts with an initial point and moves along the inverse direction of the gradient at that point in order to find the minimiser of a function.

Steepest descent

MSE contour - 2-tap example:



- Let us assume that we can evaluate the gradient of the MSE for any FIR filter $\boldsymbol{h}.$
- In this case, we can find the optimal filter coefficients, nevertheless, let us try to consider the steepest descent procedure in order to start with an initial filter and iteratively converge to the optimal one.
- In this respect, let us consider *n* as the iteration counter –not as the time index for now.

Steepest descent

• MSE contour - 2-tap example:



 $\mathbf{h}(n+1) = \mathbf{h}(n) - \mu \nabla_{\mathbf{h}}(n)$

• Step-size μ .

Murat Üney (IDCOM) C

27/06/2016 32 / 69

- At each iteration *n*, we move along the inverse direction of the MSE gradient.
- One way to perform this move is to make a line search along that direction using, for example, the golden section search, in order to solve a 1-D optimisation problem.

 $\lambda^* = \arg\min_{\lambda} \xi(\mathbf{h}(n) - \lambda \nabla_{\mathbf{h}}(n))$

 However, this would bring an additional computational cost. Instead, let us select a fixed step size μ and use it as the optimal step size λ* = μ. We will show that, under certain conditions, this selection still results with convergence to the optimal point.

Steepest descent

• MSE contour - 2-tap example:



• Gradient at new guess.

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 33 / 69

Convergence of steepest descent

MSE contour - 2-tap example:



- Because the MSE is a quadratic function of the filter coefficients, equal MSE contours are multidimensional elliptic structures (for example, for the 3-dimensional case, ellipsoids). The eigenvectors of the autocorrelation matrix specify the principle axes, and the eigenvalues specify how stretched these surfaces are.
- With a fixed step size μ , the distance from the optimal point decreases, provided that μ is smaller than the inverse of the largest eigenvalue.

Stochastic gradient algorithms

• A time recursion:

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \mu \hat{\nabla}_{\mathbf{h}}(n)$$

• The exact gradient:

$$\nabla_{\mathbf{h}}(n) = -2E\left[\mathbf{y}(n)(x(n) - \mathbf{y}(n)^{T}\mathbf{h}(n))\right]$$
$$= -2E\left[\mathbf{y}(n)e(n)\right]$$

• A simple estimate of the gradient

$$\hat{\nabla}_{\mathbf{h}}(n) = -2\mathbf{y}(n+1)\mathbf{e}(n+1)$$

The error

$$e(n+1) = x(n+1) - \mathbf{h}(n)^T \mathbf{y}(n+1)$$
 (15)

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 35 / 69

- Stochastic gradient algorithms replace the gradient in gradient descent procedures with a "noisy" estimate.
- The simplest guess of the gradient would ignore the expectation and use the instantenous values of the variables involved.

The Least-mean-squares (LMS) algorithm:



▷ Initialisation ▶ Iterations \triangleright Estimate x(n)▷ Find the error

Murat Üney (IDCOM)

27/06/2016 36 / 69

- The resulting algorithm is known as the least mean square (LMS) algorithm.
- The steps of the algorithm are as given.

LMS block diagram



Figure 10: Least mean-square adaptive filtering.

Murat Üney (IDCOM)	Optimal and Adaptive Filtering	27/06/2016	37 / 69

• LMS algorithm admits a computational structure convenient for hardware implementations.

Convergence of the LMS

• MSE contour - 2-tap example:



• Because the gradient in the steepest descent is replaced with its noisy estimate in the LMS algorithm, the study of its convergence behaviour slightly different from that for the steepest descent algorithm.

Eigenvalue ratio (EVR)



Figure 11: Eigenvectors, eigenvalues and convergence: (a) the relationship between eigenvectors, eigenvalues and the contours of constant MSE; (b) steepest descent for EVR of 2; (c) EVR of 4.

Murat Üney (IDCOM)	Optimal and Adaptive Filtering	27/06/2016	39 / 69

Comparison of RLS and LMS



Figure 12: Adaptive system identification configuration.

Error vector norm

$$\rho(n) = E\left[(\mathbf{h}(n) - \mathbf{h}_{opt})^{T} (\mathbf{h}(n) - \mathbf{h}_{opt}) \right]$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 40 / 69

- In this example, we would like to identify the unknown system. Its impulse response is, hence, the optimal solution.
- The noise shaping filter allows us to change the EVR by colorating the white noise at its input.
- It is often than not the case that we can only have noisy measurements from the system to be identified. The additive white noise in the block diagram is used to model this aspect.

Comparison: Performance



Figure 13: Covergence plots for N = 16 taps adaptive filtering in the system identification configuration: EVR = 1 (i.e., the impulse response of the noise shaping filter is $\delta(n)$).

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 41 / 69

Comparison: Performance



Figure 14: Covergence plots for N = 16 taps adaptive filtering in the system identification configuration: EVR = 11.

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 42 / 69

Comparison: Performance



Figure 15: Covergence plots for N = 16 taps adaptive filtering in the system identification configuration: EVR (and, correspondingly the spectral coloration of the input signal) progressively increases to 68.

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 43 / 69

Comparison: Complexity

Table: Complexity comparison of *N*-point FIR filter algorithms.

Algorithm	Implementation	Computational load		
class				
		multiplications	adds/subtractions	divisions
RLS	fast Kalman	10 <i>N</i> +1	9 <i>N</i> +1	2
SG	LMS	2 <i>N</i>	2 <i>N</i>	_
	BLMS (via FFT)	10log(N)+8	$15\log(N) + 30$	_

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 44 / 69

Applications

- Adaptive filtering algorithms can be used in all application areas of optimal filtering.
- Some examples:
 - Adaptive line enhancement
 - Adaptive tone suppression
 - Echo cancellation
 - Channel equalisation

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 45 / 69



Figure 16: Adaptive filtering configurations: (a) direct system modelling; (b) inverse system modelling; (c) linear prediction.

Murat Üney (IDCOM) Optimal and Adaptive Filtering 27/06/2016 46 / 69

Adaptive line enhancement



Adaptive filtering

Adaptive predictor



Prediction filter: a₀ + a₁z⁻¹ + a₂z⁻²
Prediction error filter: 1 - a₀z⁻¹ - a₁z^{z-2} - a₂z⁻³

Murat Üney (IDCOM)

otimal and Adaptive Filtering

27/06/2016 48 / 69

Adaptive tone suppression



Adaptive noise whitening



Echo cancellation

• A typical telephone connection



• Hybrid transformers to route signal paths.

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 51 / 69

Echo cancellation (contd)

• Echo paths in a telephone system



• Near and far echo paths.

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 52 / 69

Echo cancellation (contd)



• Fixed filter?

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 53 / 69

Echo cancellation (contd)



Figure 20: Application of adaptive echo cancellation in a telephone handset.

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 54 / 69

Channel equalisation





Simple channel

$$y(n) = \pm h_0 + \text{noise}$$

Decision circuit

if $y(n) \ge 0$ then x(n) = +1 else x(n) = -1

• Channel with intersymbol interference (ISI)

$$y(n) = \sum_{i=0}^{2} h_i x(n-i) + \text{noise}$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 55 / 69

Channel equalisation



Figure 22: Decision directed equaliser.

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 56 / 69

Optimal signal detection - Introduction

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 57 / 69

- Our discussion so far, has been on optimal estimation of signals.
- In some applications, one needs to test the hypothesis that a (known) signal x(n) exists in the measured signal y(n).
- *y*(*n*) consists of either only a stochastic component in the absence of *x*(*n*), or, a stochastic component together with *x*(*n*).

Optimal signal detection - Introduction

• Example: Detection of gravitational waves



Figure 23: Gravitational waves from a binary black hole merger (left, Uni. of Birmingham, Gravitational Wave Group), LIGO block diagram (middle), expected signal (right) Abbot et. al., "Observation of gravitational waves from a binary black hole merger", Phys. Rev. Let., Feb. 2016.

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 57 / 69

- An example is the detection of gravitational waves. This has been
 recently succeeded by an international consortium of scientists affiliated
 with LIGO (Laser Interferometer Gravitational-Wave Observatory) an
 achievement which is likely to be recognised with a Nobel Prize in
 Physics– is a problem that involves optimal signal detection.
- Briefly, Einstein's general theory of relativity predicts the existence of graviational waves: Objects with large masses bend the gravitational field around them and if they accelerate – as in the case of a binary black hole merger – waves will be propagated (Fig.23, left pane).
- These waves will cause contractions/expansions in the space and will induce a particular signal at the photo-detector in the setup in Fig.23. This signal captures the difference in the distance two identical light beams (originated from the same source) travel in different directions.
- The expected signal in the case of a binary black hole merger is on the right pane in Fig.23.

Optimal signal detection

• Signal detection as 2-ary (binary) hypothesis testing:

$$H_0: y(n) = \eta(n) H_1: y(n) = x(n) + \eta(n)$$
(16)

- In a sense, decide which of the two possible ensembles y(n) is generated from.
- Finite length signals, i.e.,

Vector notation

$$egin{array}{l} eta_0: \mathbf{y} = oldsymbol{\eta}\ eta_1: \mathbf{y} = \mathbf{x} + oldsymbol{\eta} \end{array}$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 58 / 69

- Let us introduce a mathematical statement of the optimal detection problem as statistical hypothesis testing:
- When the variable we would like to estimate takes values from a countable and finite set, the problem setting is referred to as a hypothesis testing problem. For *M* possible values of the variable, we have a *M*-ary hypothesis testing problem. A binary hypothesis testing problem in which case M = 2 is referred to as a detection problem.
- Let us consider the detection problem in (16). In a sense, we are asked, of two possible ensembles, which one generated the observations y(n).

Optimal signal detection

 Example (radar): In active sensing, x(t) is the probing waveform subject to design.



Figure 24: Probing waveform x(n) and returns from the surveillance region constituting y(n).

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 59 / 69

- In active sensing applications such as radar and sonar surveillance, x(n) is a waveform transmitted by the sensor for probing the surveillance region.
- A typical choice of *x*(*n*) is the chirp waveform (which has a very similar form with the expected waveform in the gravitational-wave detection example).
- For different selection of *x*(*n*), the performance of the system in terms of, for example, accuracy, bandwidth and processing requirements differ. "Waveform design" has been one of the hot research topics in signal processing for active sensing (e.g., radar/sonar).
- The reflections from the objects in the surveillance region results with a superposition of scaled and time shifted versions of x(n) with additive noise at the receiver front-end. The scaling factors are related to the properties of the reflectors. Time shifts of the pulse returns encode the distance of the reflectors.
- If x(n) is transmitted multiple times with a silent period in between, there will also be a phase shift between consecutive pulse returns from the same reflector the magnitude of which is related to the (approach) speed (and equals to the doppler angular frequency).

Optimal signal detection

 Example (radar): In active sensing, x(t) is the probing waveform subject to design.

$$y(n) = a_0 x(n - n_0) + a_1 x(n - n_1) + \dots + \eta(n)$$
(17)

Figure 24: Probing waveform x(n) and returns from the surveillance region constituting y(n).

• A similar problem also arises in digital communications.

A similar problem also arises in digital communications, where the waveform *x*(*n*) is used to encode a binary stream ··· , *a*₋₁, *a*₀, *a*₁, ··· where *a_k* ∈ {0, 1}. The received signal is given by

$$y(n) = \sum_{k} a_k x(n - kN) + \eta(n)$$
(18)

- This signal is divided into consequtive time windows of length *N* and the binary hypothesis test is applied to each window to decide on whether $a_k = 0$ or $a_k = 1$.
- In other words, for recovering a_k , $\mathbf{y} = [y(kN), y(kN+1), ..., y(kN+N-1)]^T$ is used in the following test:

$$egin{aligned} & H_0(a_k=0): \mathbf{y}=\boldsymbol{\eta} \ & H_1(a_k=1): \mathbf{y}=\mathbf{x}+\boldsymbol{\eta} \end{aligned}$$
Bayesian hypothesis testing

- Consider a random variable **H** with $H = \mathbf{H}(\zeta)$ and $H \in \{H_0, H_1\}$.
- The measurement $\bar{y} = \mathbf{y}(\zeta)$ is a realisation of \mathbf{y} .
- The measurement model specifies the likelihood $p_{\mathbf{Y}|H}(\bar{y}|H)$
- Find the probabilities of $H = H_1$ and $H = H_0$ based on the measurement vector \bar{y} .
- Decide on the hypothesis with the maximum a posteriori probability:

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 60 / 69

- Let us introduce a mathematical statement of hypothesis testing, by first introducing a decision variable *H* which takes values from the set of possible hypothesis {*H*₀, *H*₁}.
- Next, we model *H* as a random variable **H**. For the case, given an experiment outcome *ζ* from the sample space *S* (see, Chp.3 in the Statistical Signal Processing lecture notes by Dr. James Hopgood) a measurement and a hypothesis is realised as *y* = **y**(*ζ*) and *H* = **H**(*ζ*).
- The measurement model specifies the likelihood $p_{\mathbf{Y}|H}(\bar{y}|H)$ which is defined even if *H* is non-random.
- Equivalently, we specify a joint probability model with the density function $p_{\mathbf{y},\mathbf{H}}(\bar{y},H)$ where \bar{y} and H are realisations of the random variables \mathbf{y} and \mathbf{H} , respectively.
- Because $p_{\mathbf{Y},H}(\bar{y},H) = p_{\mathbf{Y}|H}(\bar{y}|H)P_H(H)$, this model implies that the events that $H = H_0$ and $H = H_1$ have probabilities associated with them, a priori to the observation of the measurements. We can select these probabilities, i.e., $P(H = H_0)$ and $P(H = H_1)$, respectively, as design parameters.
- Hypothesis testing in a Bayesian framework then involves finding a posteriori probabilities of the hypothesis.

Bayesian hypothesis testing

- Consider a random variable **H** with $H = \mathbf{H}(\zeta)$ and $H \in \{H_0, H_1\}$.
- The measurement $\bar{y} = \mathbf{y}(\zeta)$ is a realisation of \mathbf{y} .
- The measurement model specifies the likelihood $p_{\mathbf{Y}|H}(\bar{y}|H)$
- Find the probabilities of H = H₁ and H = H₀ based on the measurement vector y
 .
- Decide on the hypothesis with the maximum a posteriori probability:
- Find the maximum a-posteriori (MAP) estimate of H.

$$\hat{H} = \arg \max_{H \in \{H_0, H_1\}} P_{\mathbf{H}|\mathbf{y}}(H|\bar{\mathbf{y}})$$
(19)

where the a posterior probability of \mathbf{H} is given by

$$P_{\mathbf{H}|\mathbf{y}}(H_i|\bar{\mathbf{y}}) = \frac{p_{\mathbf{Y}|H}(\bar{\mathbf{y}}|H_i)P_{\mathbf{H}}(H_i)}{p_{\mathbf{Y}|H}(\bar{\mathbf{y}}|H_0)P_{\mathbf{H}}(H_0) + p_{\mathbf{Y}|H}(\bar{\mathbf{y}}|H_1)P_{\mathbf{H}}(H_1)}$$
(20)

for *i* = 0, 1.

Murat Üney (IDCOM)	Optimal and Adaptive Filtering	27/06/2016	60 / 69

- The decision \hat{H} is an estimate of H which is the (a posteriori) most probable hypothesis.
- The numerator in the right hand side (RHS) of Eq.(20) is the multiplication of the measurement likelihood with the a priori probability of the hypothesis *H_i*. The denominator is a scaling factor which is the same for all *H_i*. It can be found using the total probability rule.
- It can be shown that the MAP detection rule in Eq.(19) minimises the total probability of error given by

$$P_e = P(\hat{H} = H_1 | H = H_0) + P(\hat{H} = H_0 | H = H_1)$$

where the first term on the RHS is known as the probability of false alarms (or, false positives) and the second term as the probability of missed detections.

• Different detection rules can be found by using different models and performance criteria. In any case, exactly the same measurement likelihoods would be used. For example, a non-random model for *H* leads to the following maximum likelihood detector:

$$\hat{H} = rg\max_{H \in \{H_0, H_1\}} p_{\mathbf{y}|\mathbf{H}}(ar{\mathbf{y}}|H)$$

Bayesian hypothesis testing: The likelihood ratio test

MAP decision rule:

$$\hat{H} = \arg \max_{H \in \{H_0, H_1\}} P(H|\bar{y})$$

• MAP decision as a likelihood ratio test:

$$\begin{array}{c} p(H_{1}|\bar{y}) \stackrel{\hat{H}=H_{1}}{\underset{\hat{H}=H_{0}}{\overset{P}{=}}} p(H_{0}|\bar{y}) \\ p(\bar{y}|H_{1})P(H_{1}) \stackrel{H_{1}}{\underset{H_{0}}{\overset{P}{=}}} p(\bar{y}|H_{0})P(H_{0}) \\ \frac{p(\bar{y}|H_{1})}{p(\bar{y}|H_{0})} \stackrel{H_{1}}{\underset{H_{0}}{\overset{P}{=}}} \frac{P(H_{0})}{P(H_{1})} \end{array}$$

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 61 / 69

- Finding the maximum of the a posteriori probabilities can be expressed as a likelihood ratio test.
- Note that, these likelihoods are specified by the uncertainties in relating *H* to the measurement values \bar{y} , and, hence the same likelihoods will be used for decision rules under different criteria.
- For example, the aforementioned maximum likelihood detector can equivalently be realised by testing the likelihood against 1.
- This is equivalent to using non-informative a priori probabilities in a Bayesian test, i.e., $P(H_0) = P(H_1) = 0.5$.

Bayesian hypothesis testing - AWGN Example

• Example: Detection of deterministic signals in additive white Gaussian noise:

$$egin{array}{l} H_0: \mathbf{y} = oldsymbol{\eta} \ H_1: \mathbf{y} = \mathbf{x} + oldsymbol{\eta} \end{array}$$

where **x** is a known vector, $\boldsymbol{\eta} \sim \mathcal{N}(.; \boldsymbol{0}, \sigma^2 \boldsymbol{I})$.

• The likelihoods are specified by the noise distribution:

$$\frac{p(\bar{y}|H_1)}{p(\bar{y}|H_0)} \stackrel{H_1}{\underset{H_0}{\overset{>}{\leftarrow}}} \frac{P(H_0)}{P(H_1)}$$
$$\frac{\mathcal{N}(\bar{y}; \mathbf{x}, \sigma^2 \mathbf{I})}{\mathcal{N}(\bar{y}; \mathbf{0}, \sigma^2 \mathbf{I})} \stackrel{H_1}{\underset{H_0}{\overset{>}{\leftarrow}}} \frac{P(H_0)}{P(H_1)}$$

Optimal and Adaptive Filtering

27/06/2016 62 / 69

- In the case of additive noise, the likelihoods involved in the decision rule are specified by the noise distribution.
- For the Gaussian noise example, the hypothesis testing problem is equivalent to deciding on whether \bar{y} is distributed by a Gaussian distribution with mean and covariance that equals to the noise, or, a Gaussian distribution with mean that equals to the signal to be detected.

Detection of deterministic signals - AWGN (contd)
The numerator and denominator of the likelihood ratio are

$$p(\bar{y}|H_1) = \mathcal{N}(\bar{y} - \mathbf{x}; \mathbf{0}, \sigma^2 \mathbf{I})$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \prod_{n=0}^{N-1} \exp\left\{-\frac{(\bar{y}(n) - x(n))^2}{2\sigma^2}\right\}$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} (\bar{y}(n) - x(n))^2\right)\right\}$$
(21)

Similarly

$$p(\bar{\mathbf{y}}|H_0) = \mathcal{N}(\bar{\mathbf{y}}; \mathbf{0}, \sigma^2 \mathbf{I})$$
$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} (\bar{\mathbf{y}}(n))^2\right)\right\}$$
(22)

• Therefore

$$\frac{p(\bar{y}|H_1)}{p(\bar{y}|H_0)} = \exp\left\{\frac{1}{\sigma^2} \left(\sum_{n=0}^{N-1} (\bar{y}(n)x(n) - \frac{1}{2}x(n)^2)\right)\right\}$$
(23)

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 63 / 69

Detection of deterministic signals - AWGN (contd)

Take the logarithm of both sides of the likelihood ratio test

$$\log \exp\left\{\frac{1}{\sigma^2}\left(\sum_{n=0}^{N-1}(\bar{y}(n)x(n)-\frac{1}{2}x(n)^2)\right)\right\}\underset{H_0}{\overset{R_1}{\underset{H_0}{\overset{N}{\rightarrow}}}\log\frac{P(H_0)}{P(H_1)}$$

Now, we have a linear statistical test



 $\triangleq \tau$:Decision threshold



```
Murat Üney (IDCOM)
```

Optimal and Adaptive Filtering

27/06/2016 64 / 69

- This linear operation finds the evaluation of the cross-correlation of the $\bar{y}(n)$ with the signal of interest x(n) for zero-lag.
- An equivalent operation is filtering the input stream $\bar{y}(n)$ with a linear time-invariant system that has a time inversed version of x(n) as its impulse response. When the output of this filter is sampled at the length of x(n), the output will be the optimal decision variable we aim to compute.
- This filter matches *x*(*n*) in that it produces the optimal decision variable for detection *x*(*n*) under AWGN, and, is known as the matched filter.
- In other words, the matched filter is the optimal filter for detecting a known signal under AWGN.

Detection of deterministic signals - AWGN (contd)

• Matched filtering for optimal detection:



- In active sensing applications, the existence of time shifted versions of the waveform *x*(*n*) is often found by matched filtering followed by sampling with a pulse-width period.
- The decision variable sequence is then thresholded to test the existence of reflectors at different ranges (i.e., distances from the receiver).

Detection of deterministic signals under coloured noise

• For the case, the noise sequence $\rho(n)$ has an auto-correlation function $r_{\nu}[I]$ different than $r_{\nu}[I] = \sigma_{\eta}^2 \times \delta[I]$, and,

 $H_{0}: \mathbf{y} = \boldsymbol{\nu} \\ H_{1}: \mathbf{y} = \mathbf{x} + \boldsymbol{\nu} \qquad \qquad \boldsymbol{\nu} \sim \mathcal{N} \left(:: \mathbf{0}, \mathbf{C}_{\nu} = \begin{bmatrix} r_{\nu}(0) & r_{\nu}(-1) & \dots & r_{\nu}(-N+1) \\ r_{\nu}(1) & r_{\nu}(0) & \dots & r_{\nu}(-N+2) \\ \vdots & \vdots & \vdots & \vdots \\ r_{\nu}(N-1) & r_{\nu}(N-2) & \dots & r_{\nu}(0) \end{bmatrix} \right)$



Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 66 / 69

- In the case that the noise is non-white, i.e., its autocorrelation sequence is not a scaled version of Dirac's delta function, the previous results do not apply immediately.
- Nevertheless, the linearity and commutativity of LTI systems can be used to show that, we can still use the results for detection under white noise if we use a "whitening filter" as a pre-processing stage.
- Note that, in this case, the signal to be detected will be the convolution of the whitening filter with the original signal *x*(*n*).
- Design of a whitening filter is an optimal filtering problem and was discussed in this presentation in both offline and adaptive settings.
- Thus, the solution to this problem draws from all the methods we have presented throughout.

Detection of deterministic signals - coloured noise ex.



(upper left) Noise (amplitude) spectral density. (upper right) Abstract, Abbot et. al., Phys. Rev. Let., Feb. 2016.. (lower left) Matched filter outputs: Best MF (blue) and the expected MF (purple). (lower right) Measurement, reconstructed and noise signals around the detection.

Murat Üney (IDCOM) Optimal and Adaptive Filtering 27/06/2016 67 / 69

- An example to coloured noise is the background in the detector signal used in the LIGO (Abbot et. al., Phys. Rev. Let., Feb. 2016).
- Them aplitude (square root of the power) spectral density of the noise is given in the upper-left pane. Note that, there is a countable number of "line spectra" which are "predictable" components of the background. The cup-shaped spectra is the stochastic component of the background.
- More details on processing can be found in Section IV of the following document:

GW150914: First results from the search for binary black hole coalescence with Advanced LIGO

(Dated: March 4, 2016)

On September 14, 2015 at 09:5045 UTC the two detectors of the Later Interformative Gravitational-wave Observator 11,005 variantimosously observed the Istany black blac marger (VM 15914; at most of a matched Ther scarb using relativistic models of compact-object binaries that resources (FM 15914; at the Section 14, 2014) and the Section 14, 2014 (Section 14, 2014) and the Section 14, 2014 the Section 14, 2014 (Section 14, 2014) and the Section 14, 20

> The Gut AL [88] analysis implements a time-domain miched I for search (6) using techniques that were devioped to perform the near real-time compact-object binary searches 17.8, 17 occumplich this, the data (1) and templates b(d) are each whiteed in the frequency domain by dividing them by an estimate of the power spectra density of the despectrum is obtained with a combined median-geometricneam modif cation O Welch's method [8]. This proceedure is applied piece-wise on overlapping Hann-windowed timedomain blacks that are subsequently numel together to yield a cominnous whitened time series $s_i(0)$. The time-domain data $s_i(0)$ to obtain the matched firm SNN time series p(0) for each template. By the convolution theorem, p(1) obtained in this numeri to keam ea sub req (1)

Summary

- Optimal filtering: Problem statement
- General solution via Wiener-Hopf equations
- FIR Wiener filter
- Adaptive filtering as an online optimal filtering strategy
- Recursive least-squares (RLS) algorithm
- Least mean-square (LMS) algorithm
- Application examples
- Optimal signal detection via matched filtering

Murat Üney (IDCOM)

Optimal and Adaptive Filtering

27/06/2016 68 / 69

Further reading

- C. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall, 1992.
- S. Haykin, Adaptive Filter Theory, 5th ed., Prentice-Hall, 2013.
- B. Mulgrew, P. Grant, J. Thompson, *Digital Signal Processing: Concepts and Applications*, 2nd ed., Palgrave Macmillan, 2003.
- D. Manolakis, V. Ingle, S. Kogon, *Statistical and Adaptive Signal Processing*, McGraw Hill, 2000.

Optimal and Adaptive Filtering

27/06/2016 69 / 69