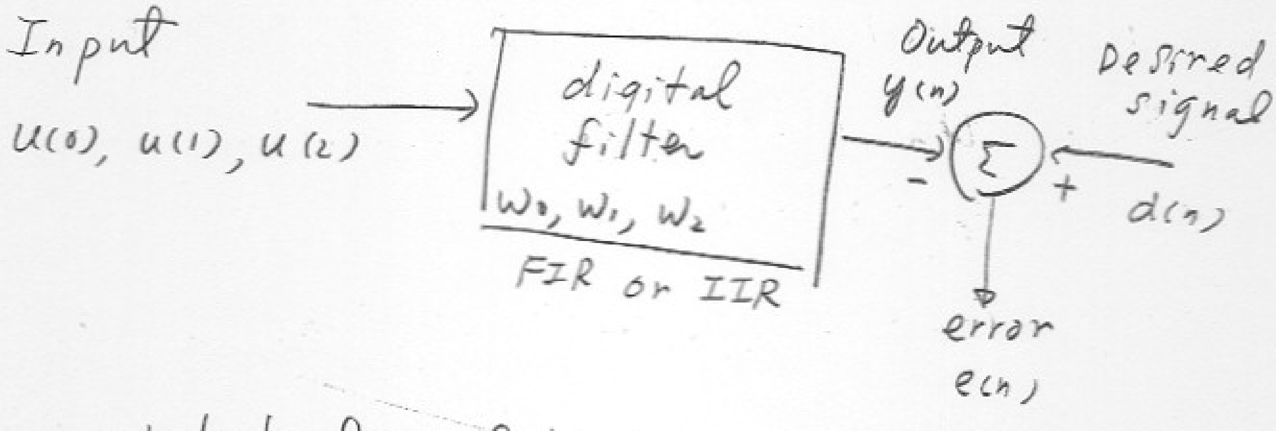
Chapter 2 Wiener Filters



statistical filtering problem

$$y(n) = \sum_{k=0}^{\infty} w_k^* u(n-k), \quad n = 0, 1, 2, \cdots$$

The above equation describes the filter output in the diagram.

The error sequence is defined as

$$e(n) = d(n) - y(n).$$

The estimation error $e(n)$ is also a random process. To optimize the filter design, we choose to minimize the mean-square value of $e(n)$. Thus a cost function of the mean-square error

$$J = E[e(n)e^*(n)] = E[|e(n)|^2]$$

is defined.

In general, the sample values of $u(n)$, $d(n)$ and the filter coefficients w_k are all complex-valued. If the filter coefficients can be denoted as

$$w_k = a_k + j b_k, \quad k = 0, 1, 2, \cdots,$$

then a gradient operator, the k^{th} element of which is written in terms of first-order partial derivatives with respect to the real part a_k and the imaginary part b_k, as

$$\nabla_k = \frac{\partial}{\partial a_k} + j \frac{\partial}{\partial b_k}, \quad k = 0, 1, 2, \cdots$$

can be defined.

Therefore, applying the operator ∇ to the cost function J, we obtain a multidimensional complex gradient vector ∇J, the k^{th} element of which is

$$\nabla_k J = \frac{\partial J}{\partial a_k} + j\frac{\partial J}{\partial b_k}, \quad k = 0, 1, 2, \cdots$$

For the cost function J to attain its extreme value (maximum or minimum), all the elements of the gradient vector ∇J must be simultaneously equal to zero, that is,

$$\nabla_k J = 0, \quad k = 0, 1, 2, \cdots,$$

and such a condition means the system is optimized in the J sence.

Since the stationary random process is discussed here, J is independent of the time variable n.

$$\nabla_k J = E\left[\frac{\partial e(n)}{\partial a_k} e^*(n) + \frac{\partial e^*(n)}{\partial a_k} e(n) + \frac{\partial e(n)}{\partial b_k} j e^*(n) + \frac{\partial e^*(n)}{\partial b_k} j e(n)\right]$$ in the mean-square error sense.

$$\frac{\partial e(n)}{\partial a_k} = -u(n-k)$$

$$\frac{\partial e(n)}{\partial b_k} = j u(n-k)$$

$$\frac{\partial e^*(n)}{\partial a_k} = -u^*(n-k)$$

$$\frac{\partial e^*(n)}{\partial b_k} = -j u^*(n-k)$$

Hence $\nabla_k J = -2E[u(n-k)e^*(n)]$. The operating conditions required for minimizing the cost function $J = E[|e(n)|^2]$ is $\nabla_k J = 0$, or $E[u(n-k)e_o^*(n)] = 0$, $k = 0, 1, 2, \ldots$, where $e_o(n)$ denote the special sequence of the estimation error. In other words, the necessary and sufficient condition for the cost function J to attain its extreme value is that the estimation error is orthogonal to the input samples.

※ Principle of Orthogonality:

Using Equation (2.1), we may express the correlation of the output samples and the error sequence as

$$E\left[y_o(n)\, e_o^*(n) \right] = 0 ,$$

where $y_o(n)$ denote the output produced by the optimal filter in the mean-square-error sense, with $e_o(n)$ denoting the corresponding estimation error.

In other words, the estimate of the desired response will be $\hat{d}(n|\vec{u}_n)$, where $\vec{u}_n = [u_1, u_2, \cdots, u_n]^T$. The optimal estimate $\hat{d}(n|\vec{u}_n) = y_o(n)$ because the conditional mean estimator is the minimum mean-square error estimator.

Theorem: Conditional mean estimator is equivalent to the minimum mean-square-error estimator.

Proof: Consider an observation, random variable y that depends on x and the requirement is to estimate y. Let $\hat{y}(x)$ denote an estimate of the parameter y. $\hat{y}(x)$ is a function of observation x. Let $C(y, \hat{y}(x))$ denote a cost function that depends on both y any $\hat{y}(x)$. According to the Bayes' estimation theory, the risk can be defined as:

$$R = E[C(y, \hat{y}(x))]$$
$$= \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} C(y, \hat{y}(x))\, f_{X,Y}(x,y)\, dx$$

In the minimum-mean-square error sense, such a cost function can be defined as

$$C(y, \hat{y}(x)) = [y - \hat{y}(x)]^2$$

Therefore,

$$R_{ms} = \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} [y - \hat{y}(x)]^2 f_{X,Y}(x,y)\,dx$$

$$= \int_{-\infty}^{\infty} dx \left[\underbrace{f_X(x)}_{\geq 0} \underbrace{\int_{-\infty}^{\infty} [y - \hat{y}(x)]^2 f_{Y|X}(y|x)\,dy}_{\geq 0} \right]$$

R_{ms} will reach its extrema when

when $I \equiv \int_{-\infty}^{\infty} [y - \hat{y}(x)]^2 f_{Y|X}(y|x)\,dy \Big|_{\hat{y}(x) = \hat{y}_{ms}(x)} = 0$

$$\left.\frac{\partial I}{\partial \hat{y}(x)}\right|_{\hat{y}(x) = \hat{y}_{ms}(x)} = 0 \Rightarrow -2 \int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\,dy + 2\hat{y}(x) \underbrace{\int_{-\infty}^{\infty} f_{Y|X}(y|x)\,dy}_{=1}$$

$$\Rightarrow \hat{y}_{ms}(x) = \int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\,dy$$

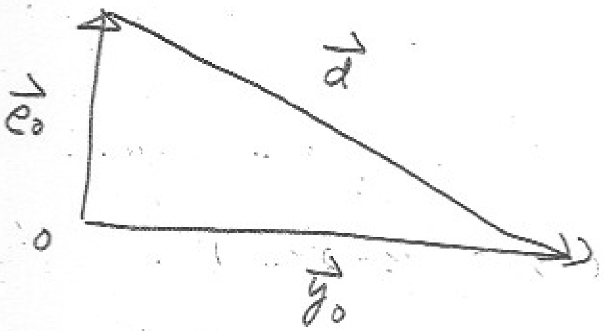
$$= E[y|x]$$

In this example, $d(n)$ depends on $u(n), u(n-1), \ldots$

$$\therefore y_o(n) \equiv \hat{d}_{ms}(\vec{u}_n) = E[d(n)|\vec{u}_n]$$

$$= \hat{d}(n|\vec{u}_n)$$

Geometric Interpretation of the Corollary to the principle of orthogonality.



2.3 Minimum Mean-square Error

When a linear discrete-time filter operates in its optimal condition,

$$e_o(n) = d(n) - y_o(n) = d(n) - \hat{d}(n|\vec{u}_n).$$

The minimal mean-square-error can be denoted as

$$J_{min} = E\{|e_o(n)|^2\}$$

$$E[|d(n)|^2] = E[|\hat{d}(n|\vec{u}_n)|^2] + E[|e_o(n)|^2] + 2\,\mathrm{Re}\{\overset{0}{\cancel{E[\hat{d}(n|\vec{u}_n)\, e_o^*(n)]}}\}$$

$$= \underbrace{E[|\hat{d}(n|\vec{u}_n)|^2]}_{\sigma_d^2} + \underbrace{E[|e_o(n)|^2]}_{\sigma_{\hat{d}}^2}$$

Hence, $J_{min} = \sigma_d^2 - \sigma_{\hat{d}}^2$

If we normalize it, we will have

$$\varepsilon \equiv \frac{J_{min}}{\sigma_d^2} = 1 - \frac{\sigma_{\hat{d}}^2}{\sigma_d^2}$$, where ε is

called the normalized mean-square-error.

According to the previous geometric interpretation,

$$\sigma_{\hat{d}}^2 \le \sigma_d^2 \Rightarrow 0 \le \frac{\sigma_{\hat{d}}^2}{\sigma_d^2} \le 1 \Rightarrow 0 \le \varepsilon \le 1.$$

If ε is zero, the optimum filter operates perfectly, in the sense that there is complete agreement between the estimate $\hat{d}(n|\vec{u}_n)$ at the filter output and the desired response $d(n)$.

X. Wiener-Hopf Equations

We start from the principle of the orthogonality,

i.e., $E[u(n-k)\, e_o^*(n)] = 0, \quad k = 0, 1, 2, \cdots$

$$\Rightarrow E\left[u(n-k)\left(d^*(n) - \sum_{i=0}^{\infty} w_{oi}\, u^*(n-i)\right)\right] = 0, \quad k=0,1,2,\cdots$$

$$\Rightarrow \sum_{i=0}^{\infty} w_{oi}\, E[u(n-k)\, u^*(n-i)] = E[u(n-k)\, d^*(n)],$$

$$k = 0, 1, 2, \cdots$$

$$\Rightarrow \sum_{i=0}^{\infty} w_{oi}\, r(i-k) = p(-k), \quad k = 0, 1, 2, \ldots$$

where $r(i-k) = E[u(n-k)\, u^*(n-i)]$

and $p(-k) = E[u(n-k)\, d^*(n)]$

If $w_0, w_1, \ldots, w_{M-1}$ is a set of FIR filter coefficients, the Wiener-Hopf equations reduce to the M simultaneous equations:

$$\sum_{i=0}^{M-1} w_{oi}\, r(i-k) = p(-k), \quad k = 0, 1, \ldots, M-1$$

where $w_{o0}, w_{o1}, \ldots, w_{o,M-1}$ are the optimum values of the tap weights of the FIR filter.

Usually, we can write the Wiener-Hopf equations in a matrix formulation.

※ Matrix Formulation of the Wiener-Hopf Equations:

Let $\tilde{R}$ denote the M-by-M correlation matrix of

$u(n), u(n-1), \ldots, u(n-M+1)$, i.e.,

$$\tilde{R} = E[\vec{u}(n)\vec{u}^H(n)]$$

where $\vec{u}(n) = [u(n), u(n-1), \cdots, u(n-M+1)]^T$.

$$\tilde{R} = \begin{bmatrix} r(0) & r(1) & \cdots & r(M-1) \\ r^*(1) & r(0) & \cdots & r(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(M-1) & r^*(M-2) & \cdots & r(0) \end{bmatrix}$$

$\vec{p}$ denotes a cross-correlation vector such that

$$\vec{p} = E[\vec{u}(n)\, d^*(n)] = [p(0), p(-1), \cdots, p(1-M)]^T$$

The Wiener-Hopf equations can be re-written in a compact matrix form as

$$\tilde{R}\,\vec{w}_o = \vec{p}$$

where $\vec{w}_o = [w_{o\emptyset}, w_{o1}, \cdots w_{o,M-1}]^T$

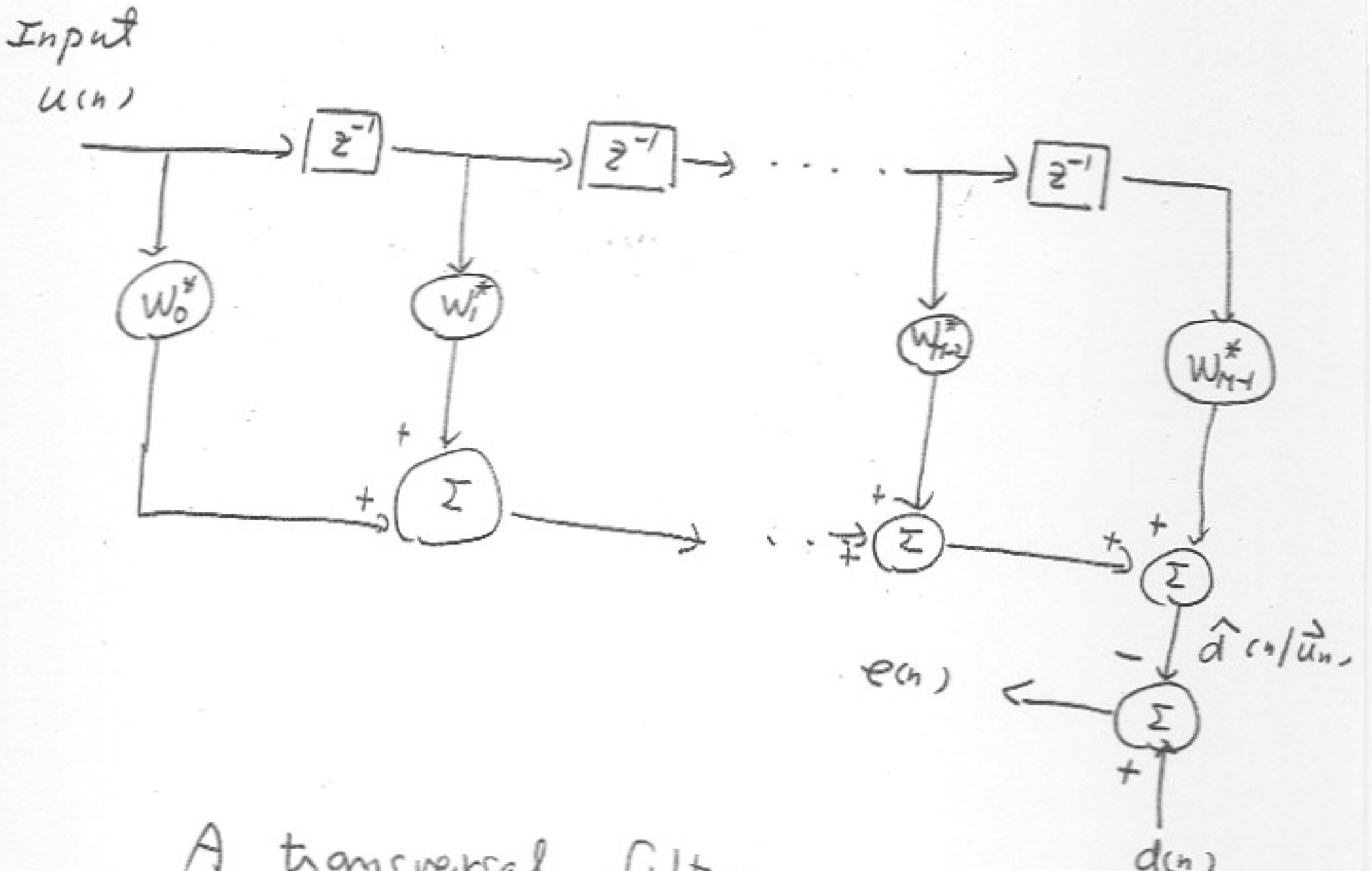
or $\vec{w}_o = \tilde{R}^{-1}\vec{p}$

2.5 Error-Performance Surface

For an arbitrary set of weight coefficients, w_k, $k=0,1,2,\cdots,M-1$, the estimation error can be stated as

$$e(n) = d(n) - \sum_{k=0}^{M-1} w_k^* u(n-k).$$

A transversal filter structure can be depicted in the following figure.



A transversal filter

Accordingly, we may define the cost function as

$$J = E[e(n)e^*(n)]$$

$$= E[|d(n)|^2] - \sum_{k=0}^{M-1} w_k^* E[u(n-k)d^*(n)]$$

$$- \sum_{k=0}^{M-1} w_k E[u^*(n-k)d(n)]$$

$$+ \sum_{k=0}^{M-1}\sum_{i=0}^{M-1} w_k^* w_i E[u(n-k)u^*(n-i)]$$

If we define:

$$\sigma_d^2 = E[|d(n)|^2]$$

$$p(-k) = E[u(n-k)d^*(n)]$$

$$\Rightarrow p^*(-k) = E[u^*(n-k)d(n)]$$

$$r(i-k) = E[u(n-k)u^*(n-i)]$$

Therefore,

$$J = \sigma_d^2 - \sum_{k=0}^{M-1} w_k^* p(-k) - \sum_{k=0}^{M-1} w_k p^*(-k) + \sum_{k=0}^{M-1}\sum_{i=0}^{M-1} w_k^* w_i r(i-k)$$

J is a quadratic function of the tap weights in the filter. Consequently we can visualize the dependence of J on the tap weights $w_0, w_1, w_2, \cdots w_{M-1}$ as a bowl-shaped $2(M+1)$- dimensional surface with $2M$ degrees of freedom represented by the tap weights of the filter. Since J is a quadratic function, this surface is characterized by a unique minimum and it is called the error-performance surface.

The minimum of J occurs when $\nabla_k J = 0$ such that $\nabla_k J = 0 \Rightarrow -2p(-k) = 2\sum_{i=0}^{M-1} w_i r(i-k)$, $\forall k$

$$\Rightarrow \sum_{i=0}^{M-1} w_{oi}\, r(i-k) = p(-k), \quad k = 0, 1, 2, \cdots, M-$$

(Wiener-Hopf Equations)

X. Minimum Mean-Square Error :

The output of the adaptive filter is

$$\hat{d}(n|\vec{u}_n) = \sum_{k=0}^{M-1} w_{ok}^{*}\, u(n-k) = \vec{w}_o^{H}\, \vec{u}(n)$$

where $\vec{w}_o$ is the tap-weight vector of the optimum filter such that

$$\vec{w}_o = [w_{o0}, w_{o1}, \ldots, w_{o,M-1}]^T.$$

and $\vec{u}(n)$ has zero mean, which makes the estimate $\hat{d}(n|\vec{u}_n)$ have zero mean too.

Therefore,

$$\sigma_{\hat{d}}^2 = E[\vec{w}_o^H \vec{u}(n) \vec{u}^H(n) \vec{w}_o]$$
$$= \vec{w}_o^H E[\vec{u}(n)\vec{u}^H(n)] \vec{w}_o$$
$$= \vec{w}_o^H \tilde{R} \vec{w}_o$$
$$= \vec{p}^H \vec{w}_o = \vec{p}^H \tilde{R}^{-1} \vec{p}$$

$$J_{min} = \sigma_d^2 - \vec{w}_o^H \tilde{R} \vec{w}_o = \sigma_d^2 - \vec{p}^H \vec{w}_o$$
$$= \sigma_d^2 - \vec{p}^H \tilde{R}^{-1} \vec{p}$$

-X- Canonical Form of the Error Performance Surface:

The mean-square error J can be written as a function of the tap weights such as

$$J(\vec{w}) = \sigma_d^2 - \vec{w}^H \vec{p} - \vec{p}^H \vec{w} + \vec{w}^H \tilde{R} \vec{w}$$

$$= \sigma_d^2 - \vec{P}^H \tilde{R}^{-1} \vec{P} + (\vec{w} - \tilde{R}^{-1}\vec{P})^H \tilde{R} (\vec{w} - \tilde{R}^{-1}\vec{P})$$

$$= J_{min} + (\vec{w} - \vec{w}_o)^H \tilde{R} (\vec{w} - \vec{w}_o)$$

where $J_{min} = J(\vec{w}_o) = \sigma_d^2 - \vec{P}^H \tilde{R}^{-1} \vec{P}$

and $\vec{w}_o = \tilde{R}^{-1}\vec{P}$

Since $\tilde{R}$ is Hermitian matrix, the eigen-decomposition can be done for $\tilde{R}$ such that $\tilde{R} = \tilde{Q}\tilde{\Lambda}\tilde{Q}^H$, where $\tilde{\Lambda}$ is a diagonal matrix consisting of eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_M$ and $\tilde{Q}$ is the orthogonal matrix with its columns equivalent to eigenvectors $\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_M$.

Hence $J = J_{min} + (\vec{w} - \vec{w}_o)^H \tilde{Q}\tilde{\Lambda}\tilde{Q}^H (\vec{w} - \vec{w}_o)$

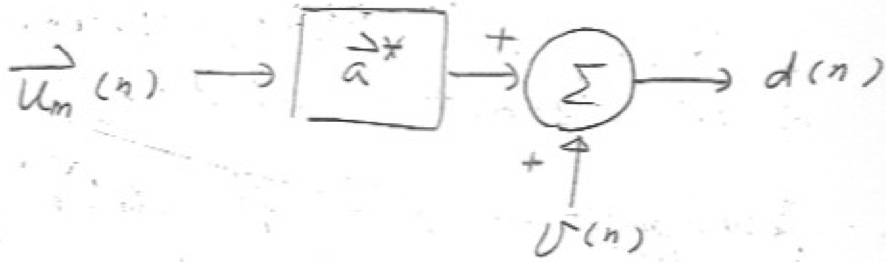
$$= J_{min} + \vec{v}^H \tilde{\Lambda} \vec{v}, \text{ where}$$

$\vec{v} = Q^H (\vec{w} - \vec{w}_o)$

$$J = J_{min} + \sum_{k=1}^{M} \lambda_k v_k v_k^* = J_{min} + \sum_{k=1}^{M} \lambda_k |v_k|^2$$

where v_k is the k^{th} component of the vector $\vec{v}$. It is the canonical form.

Example:



$$d(n) = \vec{a}^H \vec{u}_m(n) + v(n), \quad \text{where}$$

$\vec{a}$ denotes the unknown parameter vector of the model; $\vec{u}_m(n)$ denotes the input vector or regressor, $v(n)$ is the additive noise with variance σ_v^2.

$$\vec{u}_m(n) = [u(n), u(n-1), \dots, u(n-m+1)]^T$$

$$\vec{a} = [a_0, a_1, a_2, \dots, a_{m-1}]^T$$

$$\sigma_d^2 = E[d(n)d^*(n)]$$

$$= \sigma_v^2 + \vec{a}^H \tilde{R}_m \vec{a}, \quad \text{where } \tilde{R}_m = E[\vec{u}_m(n)\vec{u}_m^H(n)]$$

Given the autocorrelation function

$r(k) = E[u(n)\, u^*(n-k)]$ as follows:

$r(0) = 1.15$, $r(1) = 0.5$, $r(2) = 0.1$,

$r(3) = -0.05$

and the cross-correlation function

$p(-k) = E[u(n-k)\, d^*(n)]$ as follows:

$p(0) = 0.5272$, $p(-1) = -0.4458$, $p(-2) = -0.1003$

$p(-3) = -0.0126$.

Therefore, the optimum tap-weight vector

is $\underset{M\times 1}{\vec{w}_o} = \underset{M\times M}{\tilde{R}_M^{-1}}\ \underset{M\times 1}{\vec{p}}$

(i) for $M = 1 \Rightarrow \vec{w}_o = 0.4793$

(ii) for $M = 2 \Rightarrow \vec{w}_o = [\,0.8360,\ -0.7853\,]^T$

(iii) for $M = 3 \Rightarrow \vec{w}_o = [0.8719,\ -0.9127,\ 0.2444]^T$

(iv) for $M = 4 \Rightarrow \vec{w}_o = [0.8719,\ -0.9129,\ 0.2444,\ 0]$

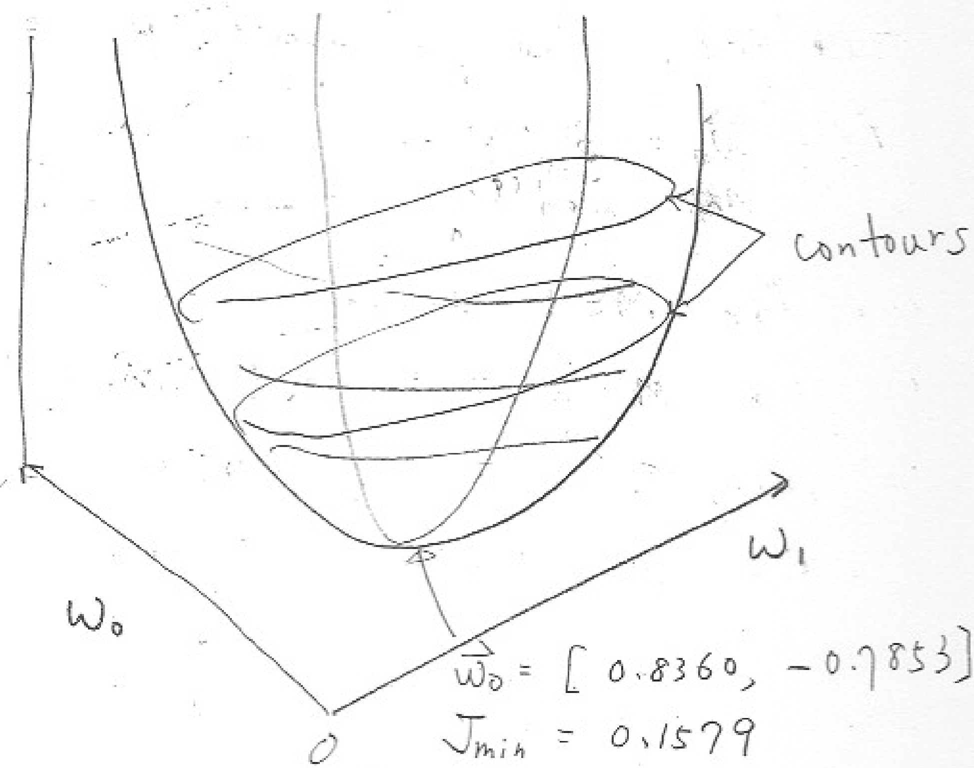
$J_{min} = \sigma_v^2 + \vec{a}^H \tilde{R}_M \vec{a}$, will reach and stay in the minimum error when we increase model order M from 1 to 3.

For filter length $M = 2$, the cost function is

$$J(w_0, w_1) = 0.9486 - 2\,[0.5272, -0.4458] \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} + [w_0 \; w_1] \begin{bmatrix} 1.1 & 0.5 \\ 0.5 & 1.1 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$$= 0.9486 - 1.0544\, w_0 + 0.8961\, w_1 + w_0 w_1 + 1.1\,(w_0^2 + w_1^2),$$

which describes the error performance surface when $M = 2$.



bowl-shaped J versus w_0, w_1

The global minimum occurs when the tap weights are optimum values such that

$$\vec{w}_0 = \begin{bmatrix} 0.8360 \\ -0.7853 \end{bmatrix},$$

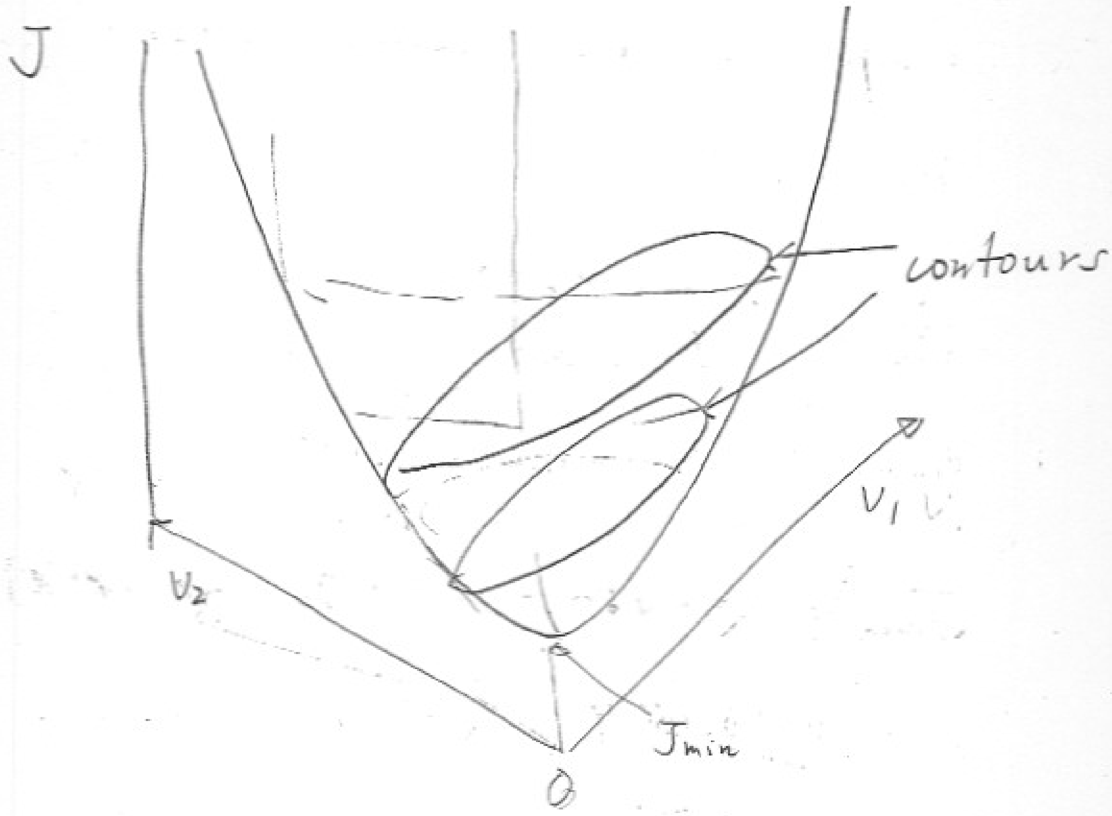
which leads to a minimum cost,

$$J_{min} = 0.9486 - [0.5272, -0.4458] \begin{bmatrix} 0.8360 \\ -0.7853 \end{bmatrix}$$

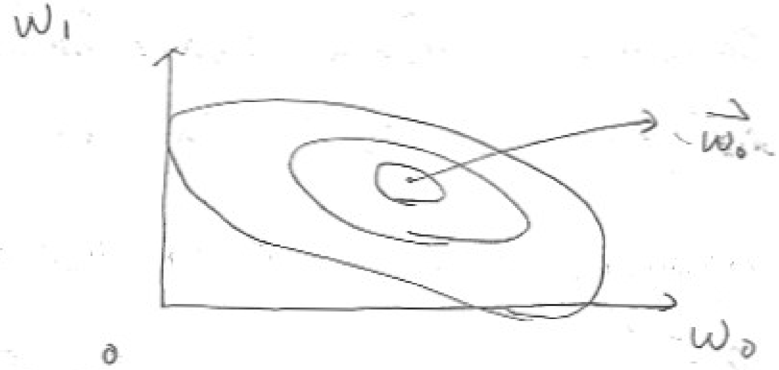
$$= 0.1579$$

※ Canonical form Transformation of the Error-Performance Surface:

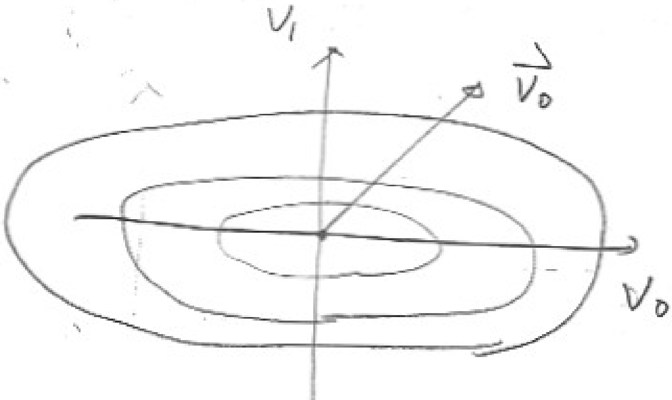
The canonical form transformation in Eq. (2.54) is just to do $\vec{w}_0$-displacement and eigenvector-(of $\tilde{R}$) rotation of the original $\vec{w}$-space. The new space spanned by $\vec{v}$ will give the regular-elliptical contours when the cost function J is sliced.



bowl-shaped J versus V_1, V_2



contours sliced in J



contours sliced in J

2.8 Linearly Constrained Minimum-Variance Filter

The aformentioned adaptive system doesn't have any additional constraint besides the cost function. However, in the real applications, contraints will be imposed very often. In this section, we would lik to discuss filter design that minimizes a mean-square criterion subject to a spesific constraint. For example, the requirement may be that of minimizing the average output power of a linear filter while the response of the filter measured at some spesific frequency of interest is constrained to remain constant.

Consider a linear transversal filter as depicted in Fig 2.9. The filter response can be given by

$$y(n) = \sum_{k=0}^{M-1} w_k^* u(n-k)$$

If the input is a sinusoidal excitation,

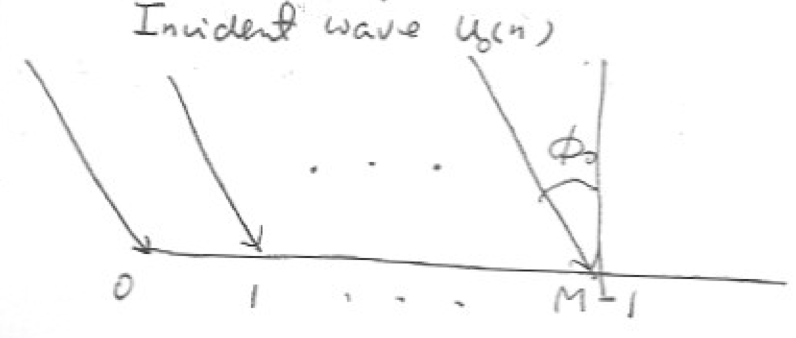
$$u(n) = e^{j\omega n}, \quad \text{then}$$

$$y(n) = e^{j\omega n} \sum_{k=0}^{M-1} w_k^* e^{-j\omega k}.$$

The constrained optimization problem can be stated as follows:

Find the optimum set of filter coefficients $w_{o0}, w_{o1}, \ldots, w_{o,M-1}$ to minimize the mean-square value of the filter output $y(n)$, subject to the linear constraint $\sum_{k=0}^{M-1} w_k^* e^{-j\omega_0 k} = g$, where ω_0 is a prescribed value of the normalized angular frequency ω; $-\pi \leq \omega \leq \pi$ and g is a complex-valued gain.

Consider the figure below,



which consists of a linear array of uniformly spaced antenna elements with adjustable weights. The array is illuminated by an isotropic source (point source) located in the far field such that at time n, a plane wave impinges on the array along a direction specified by the angle ϕ_0 with respect to the perpendicular to the array. It is also assumed that the interelement spacing of the array is less than $\lambda/2$, where λ is the wavelength of the transmitted signal. The resulting beamformer output is given by

$$y(n) = u_0(n) \sum_{k=0}^{M-1} w_k^* e^{-jk\theta_0},$$

where the direction of arrival, signified by the electrical angle θ_0, is related to the actual angle of incidence ϕ_0; $u_0(n)$ is the electrical signal picked up by the antenna element labeled $\emptyset$ in Figure 2.10 that is the point of reference; and the w_k denote the weights of the beamformer. The spatial version of the constrained optimization problem may be stated as:

Search for the optimum set of weights $w_{o\emptyset}, w_{o1}, \cdots, w_{o,M-1}$ that minimizes the mean-square value of the output subject to the linear constraint

$$\sum_{k=0}^{M-1} w_k^* e^{-jk\theta_0} = g,$$

where g is a complex-value gain.

To solve this constrained optimization problem, the method of Lagrange Multipliers is applied.

A real-valued cost function can be defined as

$$J = \underbrace{E[|y(n)|^2]}_{\text{output energy}} + \underbrace{Re\left\{\lambda^* \left(\sum_{k=0}^{M-1} w_k^* e^{-j\theta_0 k} - g\right)\right\}}_{\text{linear constraint}}$$

where λ is a complex Lagrange Multiplier.

The k^{th} element of the gradient vector $\nabla_k J$ is

$$\nabla_k J = 2 \sum_{i=0}^{M-1} w_i r(i-k) + \lambda^* e^{-j\theta_0 k}$$

Then the condition for optimality of the beamformer is described by $\nabla_k J = 0$, or

$$\sum_{i=0}^{M-1} w_{oi}\, r(i-k) = -\frac{\lambda^*}{2} e^{-j\theta_0 k}, \quad k = 0, 1, 2, \ldots M-1$$

$$\Rightarrow \vec{R}\,\vec{w}_0 = -\frac{\lambda^*}{2} \vec{s}(\theta_0), \quad \text{where}$$

$$\vec{w}_0 \equiv [w_{o0}, w_{o1}, \ldots w_{oM-1}]^T$$

$$\vec{s}(\theta_0) \equiv [1, e^{-j\theta_0}, \ldots, e^{-j(M-1)\theta_0}]^T.$$

$$\Rightarrow \vec{w}_0 = -\frac{\lambda^*}{2} \tilde{R}^{-1} \vec{s}(\theta_0)$$

The remaining problem is, to find the undetermined λ. Since the linear constraint provides $\vec{w}_0^H \vec{s}(\theta_0) = g$,

$$\left(-\frac{\lambda^*}{2} \tilde{R}^{-1} \vec{s}(\theta_0)\right)^H \vec{s}(\theta_0) = g.$$

$$\Rightarrow \quad -\frac{\lambda}{2} \vec{s}^H(\theta_0) \tilde{R}^{-1} \vec{s}(\theta_0) = g$$

$$\Rightarrow \quad \lambda = -\frac{2g}{\vec{s}^H(\theta_0) \tilde{R}^{-1} \vec{s}(\theta_0)}$$

$$\Rightarrow \quad \vec{w}_0 = \frac{g^* \tilde{R}^{-1} \vec{s}(\theta_0)}{\vec{s}^H(\theta_0) \tilde{R}^{-1} \vec{s}(\theta_0)}$$

The $\vec{w}_0$ above characterizes the weight vector of a linearly constrained minimum-variance (LCMV) beamformer.

From this example, it is clear that the constrained adaptive filter design is equivalent to the Lagrange Multiplier optimization problem. The procedure is as follows:

① Define the Lagrange multiplier cost function J as

$$J = f(\vec{w}) + \lambda g(\vec{w}),$$

where $f(\vec{w})$ is the objective function to be maximized or minimized; λ is a parameter (Lagrange multiplier) to be determined; $g(\vec{w})$ is the constraint function, i.e, $g(\vec{w}) = 0$ is the given constraint.

② $$\nabla_k J \Big|_{\vec{w} = \vec{w}_o} = 0, \quad k = 0, 1, 2, \cdots, M-1$$

③ Combine the simultaneous equations in ② with $g(\vec{w}) = 0$ to determine λ

④ Substitute the determined λ into ②
to find the optimal solution $\vec{w}_o$.

Example: Find the optimum tap-weight vector
$\vec{w} = [w_\emptyset, w_1, \cdots, w_{M-1}]^T$ such that
$\vec{w}^H \tilde{R} \vec{w}$ is maximized and $\vec{w}$ is a
uni-norm vector, i.e., $\vec{w}^H \vec{w} = 1$. (constraint).

Solution: This is so-called Rayleigh quotient $\left(\frac{\vec{w}^H \tilde{R} \vec{w}}{\vec{w}^H \vec{w}}\right)$ optimization problem. First a Lagrange multiplier - cost function is stated as $J = \vec{w}^H \tilde{R} \vec{w} + \lambda (\vec{w}^H \vec{w} - 1)$

Then $\nabla_k J = 0 \Rightarrow 2 \tilde{R} \vec{w} + 2\lambda \vec{w} = 0$

$\Rightarrow \tilde{R} \vec{w} = -\lambda \vec{w}$

$\Rightarrow -\lambda$ is the maximum eigen value of $\tilde{R}$, and hence $\vec{w}_{opt}$ is the eigenvector associated with the largest eigen value