JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

PERFORMANCE SURFACE

We represented the error as a vector in the space defined by the input signal. Let us look at the error in the space defined by the filter weights.

We saw that

$$e(k) = d(k) - y(k) = d(k) - \bar{W}^T \bar{X}(k) = d(k) - \bar{X}^T(k)\bar{W}$$

The square is

$$e^2(k) = \left(d(k) - W^T X(k)\right)^2 = d^2(k) - 2d(k)W^T X(k) + W^T X(k)\left(W^T X(k)\right)$$

$$= d^2(k) - 2d(k)X^T(k)W + W^T X(k) X(k)^T W$$

If x(k), d(k), e(k) are stationary r.p., the expected value is

$$J = \underline{\xi} = E\left[e^2(k)\right] = E\left[d^2(k)\right] - 2E\left[d(k)X^T(k)\right]W + W^T E\left[X(k)X^T(k)\right]W$$

$$= \underline{E\left[d^2(k)\right] - 2P^T W + W^T R \underline{W}}$$

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

$$P = E[d(k)X]$$

$$R = E[X(k)X^T(k)]$$

If we define as before.

Now can see that $E[e(k)^2]$ is QUADRATIC FUNCTION OF THE WEIGHTS, when the input is stationary.

We can also see that MSE can never be negative. The MSE surface is called the PERFORMANCE SURFACE. And it MUST be concave upwards.

JOSE C. PRINCIPE

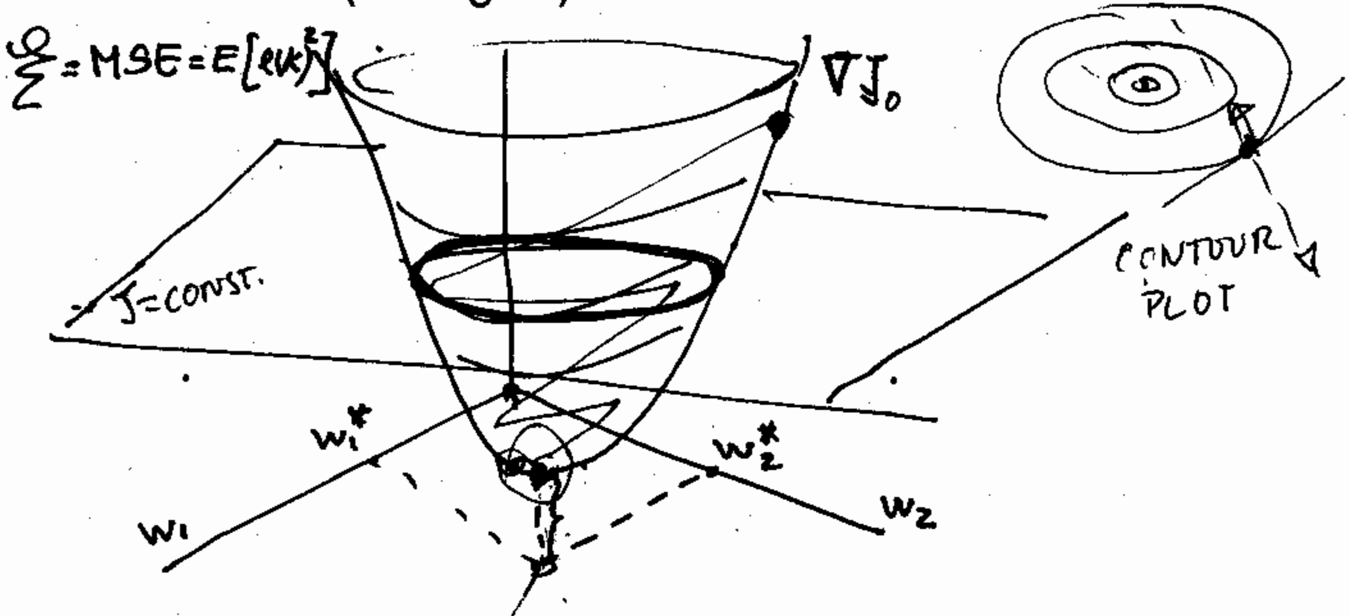
UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

In two dimensions (2 weights)



If we intercept eta by a plane parallel to w1,w2, we get an ellipse.

Another important remark is that the performance surface only has one minimum.

This is very important because we can use search techniques to find it, in alternative to computing it with an algebraic solution.

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

What is the value of the minimum?

The gradient of . eta is

$$\nabla(\xi) = \frac{\partial \xi}{\partial \bar{w}} = \begin{bmatrix} \frac{\partial \xi}{\partial w_0} \\ \vdots \\ \frac{\partial \xi}{\partial w_{L-1}} \end{bmatrix} = -2P + RW + W^T R$$

$$\overset{\shortparallel}{J} \qquad = \underbrace{-2P + 2\bar{R}\bar{W}}$$

The coordinates of the minimum are obtained by equating the gradient to zero.

$$-2P + 2RW = 0 \implies \boxed{W^* = R^{-1}P}$$

This means that the solution that corresponds to the minimum corresponds to the Wiener-Hopf solution.

NOTE: OUR J IS A SPECIAL CASE OF ξ

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

The value of the error at the minimum can be computed easily.

$$\xi_{min} = E\left[d^2(k)\right] + W^{*T} R W^* - 2 P^T W^*$$

$$= E\left[d^2(k)\right] + \left[R^{-1}P\right]^T \underbrace{R R^{-1}}_{I} P - 2 P^T R^{-1} P$$

$$= E\left[d^2(k)\right] + P^T R^{-1} P - 2 P^T R^{-1} P$$

$$= E\left[d^2(k)\right] - P^T R^{-1} P = \underline{E\left[d^2(k)\right] - P^T W^*}$$

So eta min depends on the energy of d(k), and crosscorrelation , and autocorrelation of input x(k).

Question:

If W is not exactly Wopt, what is the penalty in performance?

$$W = W^* + V$$

$$\xi = E\left[d^2(k)\right] - (W^* + V)^T R (W^* + V) - 2 P^T (W^* + V)$$

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDAt

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

$$\xi = \xi_{min} + V^T R V$$

Which means that the EXCESS MEAN SQUARE ERROR is a quadratic function of the deviation of the weights and depends only on the input signal statistics.

The form of the excess error makes us think of a change in co-ordinates to simply the expressions. The coordinate system should be centered at zeta min (Wopt), instead of (0,0).

When we do that, the new zeta becomes

$$\xi = \xi_{min} + (W - W^*)^T R (W - W^*) = \xi_{min} + V^T R V$$

where V=W-Wopt, and the gradient becomes

$$\frac{\partial \xi}{\partial V} = 2 R V$$

SAME AS BEFORE BECAUSE $V = W - W^*$

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

From the error, since it is non-negative we can conclude that

$$V^T R V \geq 0$$

which means that R must be positive semi-definite, (but can also equal zero). When the equality holds, R is singular (not full rank), and R^{-1} does not exist.

We saw that in this situation MORE THAN ONE solution exists. The iterative method can give one of the results.

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

EXAMPLE:

Design a N point FIR to match a pure delay with white noise excitation.

$$x(k) = A\, n(k) \qquad -\infty < k < \infty$$

$n(k) \rightarrow$ WHITE NOISE, ZERO MEAN, σ^2

$$d(k) = B\, n(k-\Delta) \qquad 0 \le \Delta \le N-1$$

The autocorrelation function for the input is

$$r_{ij} = E\{x(k-i)\, x(k-j)\} = A^2\sigma^2\delta(i-j)$$

$$\text{So} \quad R = A^2\sigma^2 I \qquad A^2\sigma^2\begin{bmatrix} 1 & 0 & \rightarrow \\ 0 & 1 & \\ \downarrow & & 1 \end{bmatrix}$$

The cross correlation function is

$$c_i = E\{d(k)\cdot x(k-i)\} = AB\, E\{n(k-\Delta)\, n(k-i)\}$$

$$= AB\sigma^2\delta(i-\Delta)$$

$$P = AB\sigma^2 I_\Delta \qquad I_\Delta = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix} \leftarrow \Delta \text{ ENTRY}$$

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

Therefore

$$W = R^{-1}P = (A^2\sigma I)^{-1} AB\sigma^2 I_\Delta = \frac{B}{A} I_\Delta$$

The filter has only one non-zero element that simply delays the input by delta samples and scales the input.

EXAMPLE II

Matching a rotated phasor. Assume input is a phasor and the desired signal another phasor, delayed by delta.

$$X(k) = A e^{j(wTk+\theta)}; \quad d(k) = B e^{jwT(k-\Delta)+\theta)}$$

$\omega, T = \text{CONST.}$

$\theta \to$ r.v. $[-\pi, \pi]$ UNIFORM

As before we can compute R and P

$$r_{i,j} = E\{x(k+j)\, x^*(k+i)\}$$

$$= \sum_{m=1}^{M} x(k+j)\, x^*(k+i)\, p(\theta_m)$$

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

So

$$r_{i,j} = \sum_{m=1}^{M} A e^{j\omega T((k+j)+\theta_m)} A e^{-j(\omega T((k+i)+\theta_m)} p(\theta_m)$$

$$= A^2 e^{j\omega T(j-i)} \sum_{m=1}^{M} e^{j\theta_m} e^{-j\theta_m} \frac{1}{M} = A^2 e^{j\omega T(j-i)}$$

LIKE WISE

$$C(i) = A B e^{j\omega T(i-\Delta)}$$

$$P = A B e^{j\omega T} I_\Delta \qquad I_\Delta = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ \vdots \end{bmatrix} \leftarrow \Delta \text{ ENTRY}$$

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

Now for N=1 (one coefficient filter) $\Delta = 0, N = 1$

$$W = R^{-1}P = (A^2 I)^{-1} A B e^{-j\omega T \Delta} = \frac{B}{A} e^{-j\omega T\Delta}$$

Now for N=2, $R = A^2 \begin{bmatrix} 1 & e^{j\omega T} \\ e^{-j\omega T} & 1 \end{bmatrix} \doteq \det[R] = 0$

i.e. R is singular. Many solutions!!!There are too many degrees of freedom and not enough constraints. We just need to provide a gain of B/A, and a phase shift of $-wT\Delta$. Any number of filters can do this.

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA:

904-335-8444

principe@brain.ee.ufl.edu

Source of nonuniqueness is the character of the input.

If $\bar{R} = \text{AVERAGE}\{X(k)X^T(k)\}$ or $R = E\{X(k)X^T(k)\}$

then $V^T R V = \text{Avg}\{V^T X(k) . X^T(k) V\}$

Let us call $V^T X(k) = s(k)$

s(k) is just the output of an N tap FIR filter whose impulse response is V, which we can call the difference filter.

Assume, V X(k)=0 for some choice of V, i.e. V0.

For this choice, any amount of V0 can be added to Wopt without increasing the MSE. The performance function is "blind" to a difference between Wopt and Wopt+kV0.(V0 is not observable).

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

When do situations like this happen?

$$X(k) = A\cos\left(\frac{2\pi n k}{N} + \theta\right)$$

$0 < n < N/2$

$N \rightarrow$ FILTER LENGTH

$\theta \rightarrow$ r.v.

IF $V_0 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ THEN $X^T(k)\,V_0 \equiv 0$.

Why does this V0 produce a zero s(k)? Notice that the frequency response of the filter given by V0 has transmission zeros at all multiples of f=1/N, including +/- n/N. So irrespective of the values of the input, the output will always be zero.

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

PROPERTIES OF THE PERFORMANCE SURFACE

Hardly ever we can work problems by hand as I did!!!!

To simplify things can decompose R in its eigenvectors.

The MODAL MATRIX allows a rotation of the coordinate system where R,P,W are described, and at the same time obtain a much better picture of the properties of the adaptation process.

REVIEW OF SIMPLE CONCEPTS:

Given a vector V in N dimension space, when V is multiplied by a NxN matrix A, V will be rotated and scaled.

For any matrix there are vectors, called EIGENVECTORS which have the property that they are not rotated when multiplied by A. Therefore, vector multiplication is the same as scalar multiplication by lameda.

$$\bar{A} \cdot \bar{V} = \lambda \bar{V}$$

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDAt

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

lameda is called an EIGENVALUE, V THE EIGEN VECTOR OF A.

There are at least N such eigenvector/eigenvalue combinations for an NxN matrix.

For our performace surface,

$$\xi = \xi_{MIN} + V^T R V$$

and assuming L+1 weights, R is (L+1)x(L+1).

We also saw that

- R is symmetric for real data (conjugate symmetric for complex). $r_{ij} = r_{ji}^*$
- R is positive semi-definite $V^T R V \geq 0$

This implies that the eigenvectors/eigenvalues of R are:

- R has N linearly independent eigenvectors Q_0, ...Q_L. Since their lenght is arbitrary, let us normalize it to 1.

$$\| Q_i \| = 1$$

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

- The N eigenvectors of R are orthogonal

$$Q_i^T Q_j = 0$$

so we call them orthonormal and write their inner product,

$$Q_i^T Q_j = \delta_{ij} \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$$

When one of the lamedas is repeated with multiplicity m, there are m corrresponding linearly independent eigenvectors, and they can be constructed to be orthogonal to each otehr and the others.

So, we can always write $\quad Q . Q^T = I \quad \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$

and so $\quad Q^{-1} = Q^T$

always exists.

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

- If the input is real valued, then N real eigenvalues can be found. Moreover they are always greater than zero. So we can write

$$R\,Q_n = \lambda\,Q_n \Rightarrow RQ = Q\lambda$$

or

$$R = Q\,\Lambda\,Q^{-1} = Q\,\Lambda\,Q^{T}.$$

where

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_{L+1} \end{bmatrix}$$

Q is called the MODAL matrix of R, and can be used to write equations in the "uncoupled" mode.

They are called the NORMAL form of R.

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDAt

904-335-8444

principe@brain.ee.ufl.edu

Usefulness of the MODAL matrix.

Given Q for R, let us define a coordinate transformation of the weight vector W

$$W = QW' \rightarrow \begin{cases} Q^T W = W' \leftarrow \\ Q^T P = P' \leftarrow \end{cases}$$

Now for the Wiener solution, $RW^* = P$

If we write $R = Q \Lambda Q^T \Rightarrow Q \Lambda \underbrace{Q^T W^*} = P \Rightarrow Q \Lambda W'^* = P$

and left multiply by Q^T

$$\boxed{\Lambda W'^* = P'}$$

What is the advantage of this form?

Since lameda is diagonal, the whole L+1 equations can be written independently

$$\lambda_i w_i^{*\prime} = p_i' \quad 0 \le i \le L$$

where $w_i^{*\prime}$, p_i^*

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDAt

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

are the i elements (scalars) of W' and P' respectively.

This means that each weight in the new coordinate system can be written only in terms of its eigenvalue and uncoupled correaltion coefficient. If $\lambda_i \neq 0$

$$w_i^{*\prime} = \frac{p_i'}{\lambda_i}$$

- If $\lambda_i=0$, then $w_i^{*\prime}$ is undetermined.

If we want to compute eta min,

$$\xi_{min} = E\{d_k^2\} - P^T W^* = E\{d_k^2\} - \underbrace{P^T Q}\, Q^T W^* = P^{T\prime} W^{*\prime}$$

$$= E\{d_k^2\} - \sum_{i=0}^{L} p_i' w_i^{*\prime} = E\{d_k^2\} - \sum_{i=0}^{L} \frac{|p_i'|^2}{\lambda_i}$$

Likewise for the excess mean square error. $(V = Q V')$

$$\xi = \xi_{min} + V^T R V \Rightarrow \xi - \xi_{min} = V^T R V$$

$$= V^T Q \Lambda Q^T V = V^{T\prime} \Lambda V' = \sum_{i=0}^{L} \lambda_i |v_i'|^2$$

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDAt

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

GEOMETRICAL SIGNIFICANCE OF EIGENVECTORS/EIGEN-VALUES.

Let us look at the projection of zeta in planes // to the weights.

$$\xi = \xi_{min} + V^T R V = \text{CONST}$$

which can be written as $V^T R V = \text{CONST}$

and represents an ellipse, centered at the origin of the voxv1 plane. The two normal lines to the ellipse are called the principal axis. The equation of these lines can be obtained noticing that the gradient is perpendicular to the contours (since it is perpendicular to eta).

$$\nabla = \begin{bmatrix} \frac{\partial(V^T R V)}{\partial v_0} \\ \vdots \\ \frac{\partial(V^T R V)}{\partial v_L} \end{bmatrix} = 2RV$$

From all the possible lines with this direction, we are interested in the ones that go through the origin $(V=0)$, SO μV.

THUS $2RV' = \mu V'$ OR $\left[R - \frac{\mu}{2} I\right] V' = 0$

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

V' represents the principal axis. This means that V' is an eigenvector of R.

- The eigenvector of the input correlation matrix DEFINE the PRINCIPAL axis of the error surface.

So the new uncorrelated axis are the principal axis of the error surface.

We arrive at this new representation through an AFFINE transform.

First we translated the origin

$$V = W - W^*$$

Then we rotated it

$$V' = Q^T V = Q^{-1} V$$

JOSE C. PRINCIPE

UNIVERSITY OF FLORIDA

EEL 6935- SPRING 90

904-335-8444

principe@brain.ee.ufl.edu

The eigenvalues also have an important geometrical significance

The gradient along a principal axis is

$$\frac{\partial \xi}{\partial v'_n} = 2 \lambda_n v'_n \quad \text{since} \quad \nabla = 2 \Lambda V'$$

$$\Longrightarrow \quad \frac{\partial^2 \xi}{\partial v_n'^2} = 2 \lambda_n$$

Thus the second derivative, which makes the rate of change of eta along a principal axis is proportional to the eigenvalues of R.

So knowing R, P we know:

- Minimum $\quad w^* = R^{-1} P$
- principal axis of eta
- rate of change of eta along principal axis

Therefore can sketch the error surface.