Filtratge lineal òptim

José Antonio Morán Moreno
Joan Claudi Socoró Carrié

PID_00175650

Universitat Oberta de Catalunya

Índex

Introducció

En el mòdul “Disseny de filtres discrets” hem estudiat les característiques principals dels filtres i les seves tècniques de disseny en el domini freqüencial. Com bé coneixeu, el domini de la freqüència resulta de gran utilitat en l’àmbit del processament de senyal, i permet la representació dels senyals en un domini alternatiu que ens permet extreure una informació vital per a moltes aplicacions de telecomunicacions.

Sabem que si volem fer una transmissió de diversos senyals sense interferències, senzillament podem posar la informació en bandes freqüencials separades i una operació de filtratge senzilla ens ajudarà a recuperar la informació de cadascun dels senyals en el receptor. Els filtres i les tècniques de filtratge ens permeten el disseny dels dispositius que permeten fer aquesta separació en bandes freqüencials i recuperar els senyals originals.

No obstant això, hi ha un bon nombre d’aplicacions i problemàtiques reals en què no tenim la sort de tenir els senyals en bandes separades, i això fa que el problema del filtratge freqüencial clàssic deixi de ser útil per a resoldre aquests problemes.

Podem separar dos senyals que es troben compartint la mateixa banda de freqüències? Hi ha tècniques de filtratge que ens permetin fer aquesta operació?

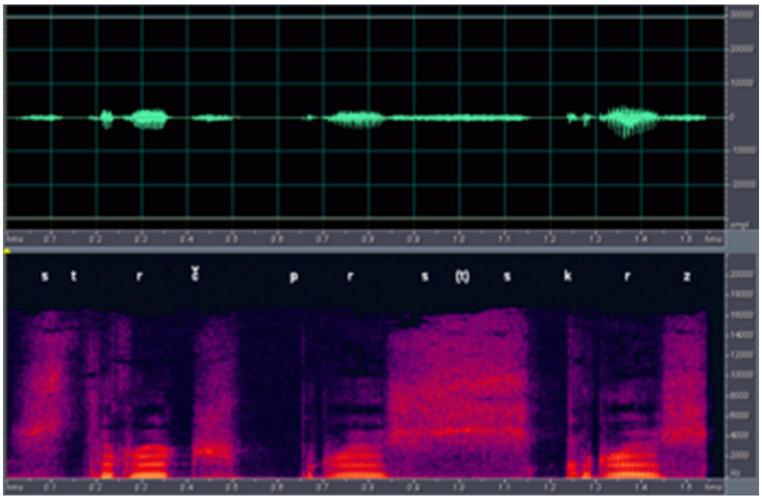
Pensem en l’exemple d’una transmissió d’àudio des d’un helicòpter. El senyal d’àudio es rep barrejat amb el soroll del rotor de l’helicòpter, que el contamina. Des d’una perspectiva de filtratge clàssic, el més bàsic que podríem fer és filtrar el senyal rebut en la banda de freqüències en què es troba el senyal de veu. Si sabem que la veu té un espectre entre 300 Hz i 3.400 Hz aproximadament, podríem filtrar el senyal rebut en aquesta banda i millorar la qualitat amb una reducció de la potència de soroll.

La pregunta que ens fem en aquest moment seria: **és això el màxim que podem fer, o hi ha alguna possibilitat de millorar les prestacions d’aquest filtratge bàsic?**

En aquest mòdul analitzarem altres possibilitats de filtratge, les que anomenarem *possibilitats de filtratge estadístic*. Dins de totes les possibilitats de filtratge estadístic, que són múltiples i variades, ens centrarem en algunes d’específiques, que són les més utilitzades en l’àmbit del processament de senyal i les seves aplicacions.

Encara que l'estudi de la teoria de senyal se sol fer sobre senyals determinístics, la realitat és que la majoria d'aplicacions, pràcticament la totalitat, es fan sobre senyals de naturalesa aleatòria. El senyal de veu humana, per molt que es trobi afitat en una banda freqüencial determinada, posseeix unes característiques freqüencials que varien amb el temps.

Figura 1



Font: es.wikipedia.org

Tal com s'observa en l'espectrograma, tot i que l'espectre està restringit freqüencialment, veiem com varia substancialment en funció dels sons que s'estan emetent. Podem imaginar que el filtratge freqüencial, malgrat que es pot utilitzar en aquesta aplicació, té grans limitacions pel fet que no aprofita totes les característiques de l'espai de senyal.

El filtratge òptim pretén assentar les bases teòriques per a poder abordar problemes de filtratge òptic basat en les característiques estadístiques dels senyals involucrats.

Gran nombre d'aplicacions de processament de senyal requereixen la utilització de tècniques estadístiques:

- Senyals contaminats per soroll extern (soroll del rotor en conversa per helicòpter).

- Senyal rebut per un sistema de comunicacions distorsionat pel canal.

- Imatge capturada per una càmera distorsionada pels efectes del moviment.

Els efectes distorsionants en aquests exemples d'aplicacions no són la majoria de vegades processos estacionaris, sinó que les seves característiques estadístiques varien amb el temps.

- El soroll del rotor de l'helicòpter o les revolucions del motor del cotxe canvien durant la comunicació.

- Els canals de radiofreqüència canvien les seves característiques en funció de les condicions de propagació (hora del dia, estació de l'any, pluja, etc.)

Tot això fa que les tècniques de filtratge hagin de treballar en el domini de l'espai de senyal estadístic, però que al seu torn disposin de la capacitat d'adaptar-se als canvis d'entorn. Com podem observar, les aplicacions reals requereixen tècniques òptimes de processament per a donar solució als problemes clàssics de comunicació i de processament de senyal.

1. El problema de l'estimació lineal

Quan abordem un problema de processament estadístic i es pretenen recuperar les característiques d'un senyal ens trobem dins d'un problema de la teoria de l'estimació.

La teoria de l'estimació és una teoria matemàtica molt àmplia que busca mètodes per a obtenir un valor aproximat d'un paràmetre determinat a partir de les dades proporcionades per una mostra:

- L'estimació d'un valor de temperatura en funció dels paràmetres atmosfèrics.

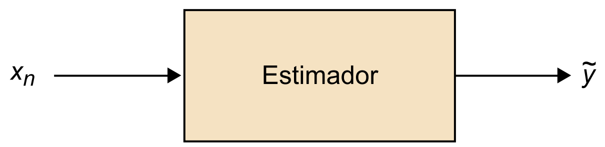
- L'estimació d'un senyal de veu en un enregistrament policial distorsionat.

- L'estimació del senyal transmès per un sistema de telecomunicacions en funció del senyal rebut.

- L'estimació de la velocitat d'un objecte en un radar.

En tots els exemples anteriors, com podem observar, resulten necessàries les mostres d'uns paràmetres i un paràmetre objectiu per a estimar. El sistema que s'encarregarà de fer l'estimació serà l'estimador.

Figura 2



Hi ha diferents criteris per a dissenyar el sistema d'estimació; els següents són alguns dels més coneguts:

- **Estimador MAP** (***maximum a posteriori***)
- **Estimador ML** (***maximum likelihood***)
- **Estimador MS** (***mean square***)
- **Estimador LMS** (***linear mean square***)

Els diferents tipus d'estimadors queden fora dels objectius d'aquesta assignatura. En qualsevol cas, principalment us heu de quedar amb la idea que els estimadors depenen del coneixement dels paràmetres estadístics de les variables involucrades. Si tenim un coneixement molt complet de la funció de probabilitat de les variables involucrades es poden desenvolupar tècniques d'estimació més fiables. Com a contrapartida, com més complex sigui el coneixement estadístic necessari, més gran serà el cost computacional de les operacions requerides per a fer l'estimació, i conseqüentment serà més complexa la implementació física de l'estimador, especialment si busquem aplicacions en temps real.

En aquest escenari veiem que resulta important el compromís entre la complexitat de l'estimador i el cost computacional d'aquest, especialment quan en aplicacions reals no podem tenir la certesa absoluta de la funció densitat de probabilitat real de les variables implicades.

És per això que en aquest mòdul ens centrarem en els **estimadors lineals**, que correspondran a estructures de filtres lineals que es dissenyaran seguint un criteri estadístic per a minimitzar l'error d'estimació.

2. La necessitat de mesurar magnituds en optimització

Quan s'aborden problemes de filtratge òptim és necessari disposar d'un criteri per a definir magnituds. En el cas de processament de senyal treballarem en espais vectorials, i conseqüentment, en un espai vectorial hem de tenir una manera de mesurar les magnituds dels vectors involucrats.

La **norma del vector** estarà definida segons el tipus de vectors amb els quals s'estigui treballant, i és diferent si estem en un espai vectorial R^N o en un espai de variables aleatòries. En qualsevol cas, la norma vectorial ha de complir una sèrie de propietats que permetin operar entre vectors i disposar de criteris vàlids de mesura de magnituds.

Les condicions bàsiques que ha de complir una **norma** són les següents.

Sigui V un **espai vectorial** sobre un **cos** $\mathbb{k}$ i $\vec{x}$ un vector de l'espai. Es diu que $\|\cdot\| : V \rightarrow \mathbb{R}$ és un operador que defineix la norma de $\vec{x}$, i escrivim $\|\vec{x}\|$, si compleix:

1) Per a tot $\vec{x}$ de V la seva norma ha de ser no negativa, i serà zero si, i només si, $\vec{x}$ és el vector zero: $0 < \|\vec{x}\|$ si $\vec{x} \neq \vec{0}$ i $\|\vec{x}\| = 0 \Leftrightarrow \vec{x} = \vec{0}$.

2) Per a tot $\vec{x}$ de V i per a tot k de $\mathbb{k}$ se satisfà que $\|k\vec{x}\| = \|k\| \cdot \|\vec{x}\|$.

3) Per a tots $\vec{x}$ i $\vec{y}$ de V es compleix que $\|\vec{x}+\vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ (desigualtat triangular).

Qualsevol operador que compleixi aquestes tres condicions, i en qualsevol geometria, serà un operador norma.

Pàgina web

http://es.wikipedia.org/wiki/norma_vectorial

Com veiem, els criteris d'una norma són criteris assenyats que faciliten l'operació entre vectors. La norma sempre ha de ser positiva, no té sentit tenir vectors amb norma negativa. Si multipliquem un vector per un factor k, la seva norma es veurà incrementada en el mateix factor; finalment, la tercera propietat correspon a la desigualtat triangular. La norma de la suma de vectors serà sempre inferior o igual a la suma de normes.

En les aplicacions de processament estadístic, quan es treballi amb variables aleatòries en problemes d'optimització necessitem un espai vectorial amb una norma ben definida que ens permeti fer les combinacions lineals dels vectors implicats i també les operacions de mesura de distàncies necessàries per a un procés d'optimització. Matemàticament, això es coneix com un espai de Hilbert.

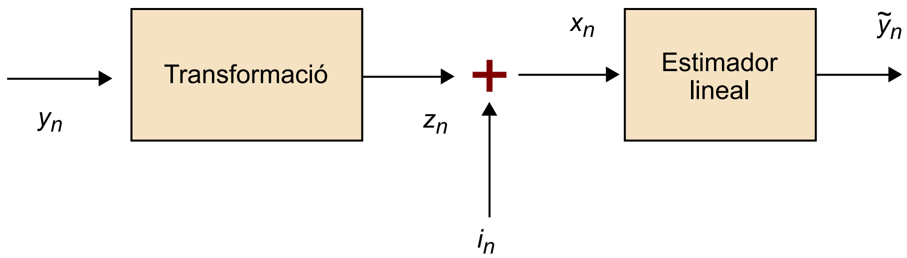
Vegeu també

No en farem aquí una exposició detallada, però podeu consultar les principals propietats d'un espai de Hilbert en l'annex d'aquest mateix mòdul.

3. Estimació lineal òptima

Per a abordar la teoria de l'estimació lineal òptima partirem d'un esquema general, a partir del qual s'aniran introduint els conceptes per a desenvolupar la teoria de l'estimació lineal òptima. El model de partida general serà el següent:

Figura 3



Ara podrem identificar els diferents senyals involucrats en el procés:

- y_n correspon al senyal que volem estimar. Aquest senyal no està disponible per a fer el processament però correspon al senyal objectiu. En el cas de l'aplicació de cancel·lació de soroll de l'helicòpter seria la veu real del pilot.

- z_n correspon a la veu del pilot que mesura el micròfon. En principi, pot ser un senyal molt similar però estarà distorsionat en més o menys mesura a causa dels efectes acústics de la cabina i també de les propietats mateixes del micròfon.

- i_n correspon als senyals interferència que se sumen; en aquest cas seria el soroll del rotor i altres sorolls que es puguin capturar a la cabina.

- x_n correspon al senyal real amb què treballarem, són les mostres reals que es capturen a través del micròfon i que corresponen al senyal original transformat amb totes les possibles interferències de soroll sumades a la cabina.

L'objectiu de l'estimador serà utilitzar les mostres de senyal capturades pel micròfon, x_n, per a fer la millor estimació possible del senyal original, y_n, seguint el criteri d'optimització definit.

Tenint això en compte, podem veure que la bondat de l'estimador dependrà de la relació estadística que tingui el senyal original amb el senyal transformat. Portant-ho a l'extrem, si el senyal transformat perd part d'informació en la transformació, aquesta informació no serà recuperable.

Un altre aspecte important en la recuperació del senyal original és la independència estadística dels senyals d'interferència respecte de l'original. Si el senyal d'interferència se superposa en el camp d'informació del senyal original això perjudicarà la qualitat de l'estimació. En canvi, si els senyals interferents són estadísticament independents del senyal original, el processador podrà aïllar els dos espais de senyal i fer una estimació de millor qualitat. En aquest exemple concret, si el soroll del rotor és estadísticament independent de la veu, l'estimació serà de millor qualitat, però si hi ha superposició en la informació, es perdrà qualitat en l'estimació.

A manera il·lustrativa, sabem que els diferents sons tenen característiques estadístiques diferents, i la veu varia les seves característiques en funció dels sons que emet. Si tenim un so que és molt diferent del rotor, com per exemple una vocal, podem esperar que l'estimació sigui de bona qualitat, però quan estiguem emetent sons fricatius, com per exemple la *r*, que té un component més similar al soroll de fons, podem esperar llavors una pèrdua de qualitat del senyal estimat.

En el cas concret de l'estimador lineal òptim, la variable estimada es generarà a partir de la combinació lineal de les mostres del senyal d'entrada, o el que és el mateix, com la sortida d'un filtre lineal de resposta impulsional finita FIR:

$$\tilde{y}[n] = \sum_{k=n_a}^{n_b} h^*[n-k]x[k]$$

Així mateix, en considerar que es treballa sobre espais vectorials, totes les operacions entre els senyals, ja siguin els d'entrada com les interferències, corresponen a combinacions lineals de variables dins de l'espai vectorial.

S'entendrà per **filtre lineal òptim** aquell que minimitza el valor esperat de la funció error al quadrat, és a dir:

$$E\left[\left(\tilde{y}[n] - y[n]\right)^2\right]_{MIN}$$

El problema d'optimització es resoldrà trobant els coeficients òptims del filtre estimador que minimitzen aquesta funció d'error.

4. Revisió de conceptes bàsics d'estadística i senyals aleatoris

Com a revisió dels conceptes bàsics de probabilitat i senyals aleatoris, en aquest apartat es presentaran les principals propietats que convé revisar per a abordar amb èxit la resta del mòdul. Es considera que disposeu de coneixements de probabilitat i estadística i els conceptes presentats en aquest punt no són més que una simple revisió dels coneixements que ja teniu.

> En l'àmbit de l'estadística i la teoria de la probabilitat, es defineixen els **processos estocàstics** com un concepte matemàtic que caracteritza una successió de variables aleatòries que evolucionen en funció d'una altra variable, que normalment correspon al temps.

En principi, en la definició completa d'un procés estocàstic seria necessari conèixer la funció densitat de probabilitat de cadascuna de les variables per cada instant de temps; evidentment, això resulta complex de conèixer en aplicacions reals, de manera que es definiran un subconjunt de processos estocàstics que s'adapten millor a les aplicacions reals.

Dins de l'àmbit de les sèries temporals hi ha diferents exemples de processos estocàstics:

- Senyals de l'àmbit de la telecomunicació i l'electrònica.
- Senyals d'aplicacions biomèdiques (encefalogrames, electrocardiogrames, etc.)
- Senyals procedents de processos sísmics o geològics.
- Evolució de la població d'una ciutat, municipi o país.
- Índexs de borsa
- Taques solars

I un llarg etcètera, són exemples de processos estocàstics que ens podem trobar en els àmbits d'aplicació.

La caracterització completa d'un procés estocàstic és tan complexa que difícilment podem trobar aplicacions reals en què sigui possible aquesta caracterització, de manera que haurem de treballar amb alguns casos especials de processos estocàstics amb unes característiques particulars que en faciliten l'ús en les aplicacions reals a partir de les dades conegudes.

Els processos estocàstics es poden definir en temps continu o en temps discret. En aquest cas ens centrarem en els processos estocàstics en temps discret, que són els que s'utilitzen en les aplicacions de processament del senyal digital.

4.1. Funció densitat de probabilitat, mitjana i variància

Suposem que x és una variable aleatòria amb funció densitat de probabilitat $p(x)$. La seva mitjana, variància i el seu moment de segon ordre estaran definits de la manera següent:

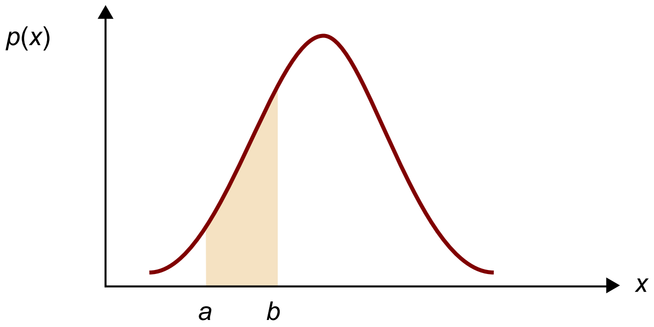
$m = E[x] = \int_{-\infty}^{\infty} x p(x) dx$ com la mitjana de la variable aleatòria

$\sigma 2 = var(x) = E[(x-m)^2] = \int_{-\infty}^{\infty} (x-m)^2 p(x) dx$ com la variància

$E[x2] = \int_{-\infty}^{\infty} x2 p(x) dx$ com el moment de segon ordre

La probabilitat que una variable aleatòria tingui un valor dins d'un interval de valors $[a, b]$ està definida per $\text{Prob}[a \le x \le b] = \int_{a}^{b} p(x) dx$

Figura 4



Com sabem, la funció densitat de probabilitat està sempre normalitzada a 1, atès que la probabilitat que la variable adquireixi qualsevol valor en tot el seu conjunt de valors sempre serà 1. Això ho podem representar de manera matemàtica així:

$$\int_{-\infty}^{\infty} p(x) dx = 1$$

Aquesta mateixa propietat ens permet determinar amb un senzill càlcul que:

$$\sigma^2 = var(x) = E\left[(x-m)^2\right] = E[x^2] - m^2$$

Les variables aleatòries poden estar definides sobre un domini continu o discret. En aquest cas considerem que la variable aleatòria està definida sobre un domini continu.

4.2. Processos estocàstics o senyals aleatoris

Un procés aleatori es defineix com una seqüència de variables aleatòries $\{x_0, x_1, x_2, \ldots, x_n\}$ en què n habitualment serà la variable temps, o mostres temporals. En aquest cas, la descripció completa del procés o senyal aleatori requeriria el coneixement de totes les funcions densitat de probabilitat conjuntes:

$$p(x_0, x_1, x_2, \ldots, x_n) \quad \text{per a} \quad n = 0, 1, 2, \ldots.$$

Podem intuir que un coneixement tan exhaustiu de la funció densitat de probabilitat conjunta serà pràcticament impossible de tenir en aplicacions reals, de manera que ens haurem de conformar amb un coneixement més limitat.

En les aplicacions que es presentaran en aquesta assignatura normalment es treballa amb senyals de mitjana zero. En qualsevol cas, si el senyal no presenta mitjana zero serà fàcil redefinir un senyal equivalent restant el valor mitjà i simplificar així tot el procés d'anàlisi, que es complicaria en el cas de senyals de mitjana diferent a zero. Una vegada finalitzat tot el tractament és tan fàcil com tornar a afegir la mitjana al resultat final.

Senyals de mitjana zero

Si $E[x] = m$, podem definir $x' = x - m$, en què x' és una variable de mitjana zero.

Per a treballar amb filtres lineals òptims, més endavant veurem que sovint haurem d'utilitzar la funció d'autocorrelació. La funció d'autocorrelació proporciona un indicador de la influència que té una variable aleatòria en un instant de temps sobre una altra que s'hi troba K mostres separada.

$$R_{xx}(n+k, n) = E\left[x_{n+k} x_n\right]$$

Veiem que la funció de correlació determina la informació estadística conjunta entre la variable en l'instant n i aquella desplaçada a $n + k$. Normalment anomenem *lag* la separació relativa entre variables aleatòries.

4.3. Processos estocàstics estacionaris

Com s'ha comentat anteriorment, es fa molt difícil arribar a obtenir el coneixement exhaustiu de la funció densitat de probabilitat dels processos estocàstics en aplicacions reals. No obstant això, els processos estacionaris representen un subconjunt que s'adapta molt millor a les condicions necessàries per a desenvolupar aplicacions.

Es considera que un procés estocàstic és **estacionari en sentit estricte** quan les funcions densitat de probabilitat de les variables aleatòries involucrades són invariants davant un canvi de temps. És a dir, si fem un canvi en l'índex temporal trobarem la mateixa seqüència exacta de variables aleatòries. Com podem imaginar, resulta complex en aplicacions reals demostrar que es compleix aquesta propietat, de tal manera que encara haurem de relaxar una mica les condicions en casos d'aplicacions reals.

Un procés estocàstic serà **estacionari en sentit ampli** quan els dos moment de primer ordre, és a dir, la mitjana i la covariància, es mantenen estables en el temps. És a dir, el valor de la mitjana serà constant per a qualsevol instant de temps i el valor de la correlació únicament dependrà de la diferència temporal entre variables, de manera que la correlació entre la mostra 1 i 3 serà igual entre la 5 i 7, o entre la 9 i l'11. Això simplifica enormement les condicions que ha de verificar un procés pel que fa a caracterització de les variables aleatòries involucrades i fa molt més senzill dur a terme aplicacions sota aquests supòsits.

$$R_{xx}(k) = E[x_{n+k}x_n] = E[x_{n'+k}x_{n'}]$$

Observem que en aquest cas la funció d'autocorrelació no depèn del valor temporal n sinó que quedarà completament caracteritzada pel valor del desplaçament o *lag*.

D'altra banda, en dissenyar filtres lineals, veurem que per al càlcul dels coeficients òptims tan sols ens serà necessari conèixer les mitjanes i les correlacions, de tal manera que els processos estocàstics en sentit ampli són suficients per a poder fer les aplicacions.

4.4. Processos ergòdics

Fins ara hem anat fent un procés de simplificació dels processos estocàstics i les seves característiques estadístiques per poder apropar-nos a aplicacions reals. Hi ha un altre pas que haurem de fer per a poder utilitzar tots aquests conceptes en el camp de les aplicacions reals.

Suposem que tenim una mostra de veu i que la podem caracteritzar com un procés estacionari en sentit ampli. El coneixement de les estadístiques que ens permetrien estimar les mitjanes i variàncies de les variables involucrades requeriria el coneixement de múltiples realitzacions del mateix procés. Això implicaria disposar en una aplicació de múltiples realitzacions del senyal de veu, cosa que no sempre serà possible en les aplicacions reals.

Imaginem el cas d'un senyal de veu que serveix de prova policial i que requereix la cancel·lació de soroll de fons per a extreure clarament la conversa. En la prova de veu, l'única de què disposarem, probablement la veu del subjecte objectiu estarà camuflada amb les interferències o altres veus que puguin aparèixer en l'enregistrament. Suposant que el procés és estacionari en sentit ampli, per a fer la resolució del problema necessitaríem l'estimació dels moments de primer ordre i de segon ordre del procés. En no disposar més que d'una realització temporal del procés, l'única possibilitat que ens queda és estimar les mitjanes estadístiques a partir de les mitjanes temporals, ja que en qualsevol altre cas seria impossible la resolució del problema, en no disposar més que d'una realització del procés. Aquells processos en què el càlcul de les mitjanes estadístiques es pot fer a partir de mitjanes temporals s'anomenen *processos ergòdics*. Haurem de fer aquesta suposició per a poder treballar amb aplicacions reals en què només coneguem una realització del procés, que d'altra banda són la majoria d'aplicacions que ens trobarem en els casos reals.

En conclusió, podem dir que en les aplicacions en què aplicarem el filtratge lineal òptim considerarem que treballem amb processos estacionaris de segon ordre i ergòdics, i això ho mantindrem així durant tot el curs a fi de resoldre els problemes de filtratge lineal òptim o de Wiener.

5. El filtre òptim o filtre de Wiener

El filtratge lineal òptim o filtre de Wiener correspon al disseny d'un filtre estadístic lineal que minimitza la funció de cost:

$$E\left[\left(e[n]\right)^2\right]\Big|_{MIN}$$

$e[n]$ és la funció error d'estimació, és a dir, la diferència entre el senyal desitjat i la sortida del filtre.

A fi d'endinsar-nos progressivament en les principals característiques del filtre lineal òptim abordarem el problema per etapes, i analitzarem en primer lloc les característiques d'un filtre d'ordre zero, és a dir, un filtre amb una resposta impulsional formada per un únic coeficient, per a seguir aprofundint-hi amb els conceptes bàsics per a filtres d'ordre major.

5.1. Filtre lineal òptim d'ordre zero

En el cas particular d'un filtre lineal òptim d'ordre zero, la resposta impulsional del filtre $h[n]$ seria un únic coeficient. La resolució del problema és determinar el valor d'aquest coeficient que fa que es minimitzi l'error quadràtic mitjà. Per a això, s'ha de fer la minimització de la funció error següent:

$$E\left[\left(y[n]-\hat{y}[n]\right)^2\right] \quad \text{en què} \quad \hat{y}[n]=ax[n]$$

La minimització d'aquesta funció error implicaria el càlcul de la derivada i igualar a zero. No seria necessari el càlcul de la segona derivada, ja que en ser una funció quadràtica tan sols disposa d'un mínim, i els màxims els tindrà per a valors del coeficient que tendeixen a infinit.

El desenvolupament de l'equació ens porta a l'expressió següent:

$$E\left[\left(y[n]-\hat{y}[n]\right)^2\right]=E\left[\left(y[n]-ax[n]\right)^2\right]$$

I trobarem la minimització derivant respecte del coeficient a i igualant a zero:

$$\frac{dE\left[\left(y[n]-ax[n]\right)^2\right]}{da}=-2E\left[\left(y[n]-ax[n]\right)x[n]\right]=0$$

La solució de l'equació seria:

$$E[y[n]x[n]] - aE[(x[n])^2] = 0$$

De tal manera que el valor òptim del coeficient és:

$$a = \frac{E[y[n]x[n]]}{E[(x[n])^2]}$$

El filtre lineal òptim d'ordre zero ens aporta una resolució simple però que ja permet interpretar alguns dels aspectes fonamentals dels estimadors lineals òptims. Veiem que el coeficient és directament proporcional a la correlació creuada que hi ha entre el senyal objectiu $y[n]$ i el senyal d'entrada $x[n]$. Si el senyal que pretenem estimar i el senyal origen no comparteixen cap informació comuna (la correlació seria zero), el coeficient del filtre òptim seria zero. Això vol dir que si no hi ha informació compartida entre un senyal i l'altre, la millor estimació que podem fer és zero. A mesura que hi hagi una correlació creuada superior entre el senyal objectiu i el senyal d'entrada, més bona serà l'estimació que podrem fer i menor serà el senyal error.

Nota

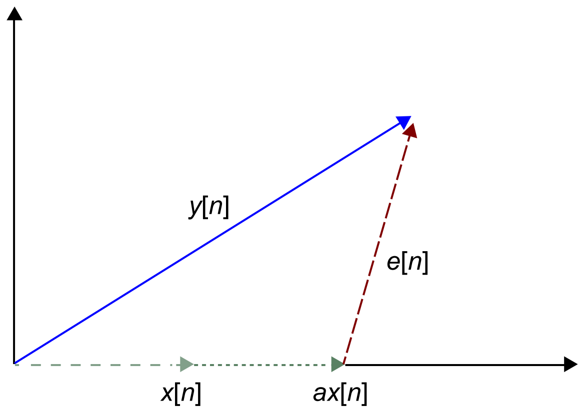
És important remarcar aquest aspecte: el fet que el filtre sigui òptim no vol dir que l'error sigui molt petit, sinó que l'error és el menor possible amb la informació de què es disposa. En el cas de senyals incorrelacionats el senyal d'error seria directament el senyal objectiu $y[n]$, atès que no es podria fer cap tipus d'estimació més bona en aquest cas.

Per aclarir aquests conceptes repassarem algunes claus de geometria euclidiana que permetin observar de manera visual aquests conceptes. Tenint en compte que les variables aleatòries amb un producte escalar ben definit formen un espai vectorial, podem fer ús de la geometria euclidiana per a representar els conceptes d'interès en el problema d'optimització lineal òptima. Per a això es representaran les variables aleatòries com a vectors i s'utilitzarà com a producte escalar la funció de correlació entre variables.

Exemples

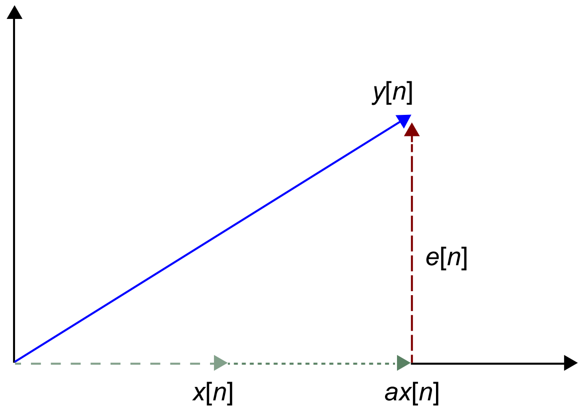
En la figura 5 podem observar en blau fosc (línia contínua) el senyal objectiu $y[n]$, en verd clar (discontínua ampla) el senyal d'entrada $x[n]$ i en verd fosc (puntejada) la sortida del filtre d'un coeficient $ax[n]$, amb el vector vermell (discontínua estreta) com a vector d'error.

Figura 5



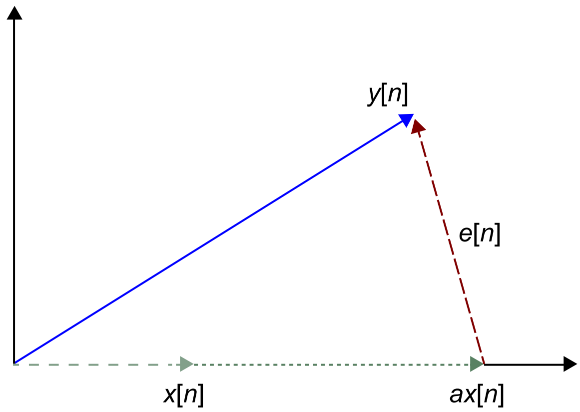
En aquest primer exemple veiem que el coeficient amplifica el vector però encara continua quedant una part del vector d'error que es projecta sobre l'eix x, de tal manera que el coeficient pot encara variar actuant sobre el vector d'error. Si a creix veiem que el vector d'error encara decreixerà.

Figura 6



En el segon exemple (figura 6) s'observa que el valor de a s'ha incrementat i el vector d'error té ara mateix el seu valor mínim, ja que qualsevol modificació del valor a farà que l'error creixi, com es pot observar en la figura 7.

Figura 7



Fent ús dels conceptes exposats prèviament, i de manera purament visual, determinem que obtindrem el filtre òptim lineal, aquell que minimitza el vector d'error, quan el vector d'error sigui perfectament ortogonal a l'espai d'informació d'entrada. Mentre el vector d'error tingui projecció sobre l'espai d'entrada, podrem continuar modificant el valor del coeficient i anar reduint l'error. En canvi, quan el vector d'error sigui perfectament ortogonal, ja no serà possible minimitzar més l'error i ens trobarem en el cas òptim.

Si recordem la solució matemàtica per al cas òptim:

$$\frac{dE\left[(y[n]-ax[n])^2\right]}{da} = -2E\left[(y[n]-ax[n])x[n]\right]=0$$

I a continuació reescrivim l'equació de la manera següent:

$$E\left[(y[n]-ax[n])x[n]\right]=0$$

$$E\left[e[n]x[n]\right]=0$$

Observem que trobem la solució òptima quan el vector d'error és ortogonal al vector d'entrada $x[n]$, és a dir, la mateixa conclusió a la qual s'ha arribat mitjançant la interpretació geomètrica dels vectors.

Aquest factor és el que es coneix com a **principi d'ortogonalitat**, que correspon a la resolució òptima lineal quan la funció de cost és l'error quadràtic mitjà. En lloc de plantejar totes les equacions i fer les derivades escaients, es pot plantejar el sistema d'equacions simplement forçant que l'error sigui ortogonal a cadascuna de les variables d'entrada de manera simultània, cosa que simplifica en gran mesura l'operativa de resolució d'aquest tipus de sistemes.

5.2. Filtre lineal òptim d'ordre *N*

Una vegada treballats els conceptes del cas simple, es tracta ara de generalitzar el problema a un filtre d'ordre *N*. Perquè el desenvolupament sigui el més complet possible el farem sobre els processos estocàstics complexos, de tal manera que haurem resolt el cas més complet possible, amb els processos reals com a cas particular de la nostra anàlisi.

En aquest exemple d'anàlisi desenvoluparem primer l'anàlisi matemàtica completa per obtenir el sistema d'equacions que s'haurà de resoldre, i finalment plantejarem les mateixes equacions a partir del principi d'ortogonalitat, i així veurem la utilitat del desenvolupament fet en l'apartat anterior.

En el cas d'un filtre d'ordre *N* de coeficients complexos i en un espai vectorial de processos estacionaris i ergòdics complexos, tenim la funció de cost següent:

$$E[|e|^2]\big|_{MIN} = E\left[\left(y - \vec{h}^H \vec{x}\right)\left(y - \vec{h}^H \vec{x}\right)^H\right]\Big|_{MIN}$$

Per al desenvolupament d'aquest càlcul, s'utilitza la notació matricial de la manera següent:

$$\hat{y} = \vec{h}^H \vec{x}$$
$$e(y - \hat{y}) = (y - \vec{h}^H \vec{x})$$

$$\hat{y} = \hat{y}[n] \quad y = y[n] \quad e = e[n]$$

$$\vec{x} = \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-N] \end{bmatrix} \quad \vec{h} = \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[N] \end{bmatrix}$$

Observeu que la sortida del filtre correspon al producte escalar dels coeficients del filtre per un vector de mostres d'entrada girades en el temps i desplaçades. Aquest producte escalar no és més que la representació vectorial del sumatori de convolució per a la sortida del filtre $\hat{y}[n] = \sum_{k=0}^{N} h^*[i]x[n-k]$, amb * com a operador conjugat.

Vegeu també

Podeu revisar l'annex per a fer un repàs dels operadors vectorials utilitzats en el desenvolupament.

La minimització de la funció de cost *J*, definida com el valor esperat de l'error quadràtic mitjà, implica la minimització de l'equació següent:

$$E[|e|^2]\,|_{MIN} = E[yy^H] - \vec{h}^H E[\vec{x}y^H] - E[y\vec{x}^H]\vec{h} + \vec{h}^H E[\vec{x}\vec{x}^H]\vec{h}|_{MIN}$$

En aquest cas, en ser un problema de dimensió *N*, la minimització de la funció de cost implicarà la derivada vectorial igualada a zero. En ser una funció quadràtica, la resolució de l'equació ens donarà el mínim de la funció:

$$\frac{\partial E[|e|^2]}{\partial \vec{h}^*} = -E[\vec{x}y^H] + E[\vec{x}\vec{x}^H]\vec{h} = 0$$

I la resolució ens porta a:

$$\begin{aligned}
&-E[y^*\vec{x}] + E[\vec{x}\vec{x}^H]\vec{h} = 0\\
&\vec{h}_{opt} = E[\vec{x}\vec{x}^H]^{-1}E[y^*\vec{x}] = \bar{\bar{R}}_{xx}^{-1}\vec{x}_{yx}\\
&\vec{r}_{yx} = E[y^*\vec{x}]^* = E\begin{bmatrix} y^*[n]x[n] \\ y^*[n]x[n-1] \\ \vdots \\ y^*[n]x[n-N] \end{bmatrix} = \begin{bmatrix} r_{yx}^*(0) \\ r_{yx}^*(1) \\ \vdots \\ r_{yx}^*(N) \end{bmatrix}\\
&\bar{\bar{R}}_{xx} = E[\vec{x}\vec{x}^H] = E\begin{bmatrix} x[n]x^*[n] & x[n]x^*[n-1] & \cdots & x[n]x^*[n-N] \\ x[n-1]x^*[n] & x[n-1]x^*[n-1] & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x[n-k]x^*[n] & \cdots & \cdots & x[n-N]x^*[n-N] \end{bmatrix} =\\
&\qquad = \begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \dots & r_{xx}[N] \\ r_{xx}[-1] & r_{xx}[0] & & \vdots \\ \vdots & & \ddots & \vdots \\ r_{xx}[-N] & \dots & \dots & r_{xx}[0] \end{bmatrix}
\end{aligned}$$

S'observa que en el cas d'un sistema de dimensió *N* la resolució del problema demana la resolució d'un sistema d'equacions d'ordre *N* en què s'involucra el vector de correlació creuada entre el senyal objectiu i cadascuna de les entrades del sistema, i també la matriu d'autocorrelació de totes les dades del sistema.

En el cas de treballar amb senyals reals, la solució seria simplement un cas particular de l'exemple anterior en el qual els complementaris desapareixen en estar sobre l'espai dels senyals reals.

$$\vec{r}_{yx} = E[y\vec{x}] = E\begin{bmatrix} y[n]x[n] \\ y[n]x[n-1] \\ \vdots \\ y[n]x[n-N] \end{bmatrix} = \begin{bmatrix} r_{yx}(0) \\ r_{yx}(1) \\ \vdots \\ r_{yx}(N) \end{bmatrix}$$

$$\bar{\bar{R}}_{xx} = E\left[\vec{x}\,\vec{x}^T\right] = E\begin{bmatrix} x[n]x[n] & x[n]x[n-1] & \cdots & x[n]x[n-N] \\ x[n-1]x[n] & x[n-1]x[n-1] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ x[n-N]x[n] & \cdots & \cdots & x[n-N]x[n-N] \end{bmatrix} =$$

$$= \begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \ldots & r_{xx}[N] \\ r_{xx}[1] & r_{xx}[0] & & \vdots \\ \vdots & & \ddots & \\ r_{xx}[N] & \ldots & & r_{xx}[0] \end{bmatrix}$$

Una vegada desenvolupat el cas de dimensió N, plantejarem el cas d'un filtre d'ordre 1 (dos coeficients) per acabar d'aprofundir en els conceptes presentats fins al moment. La interpretació sobre un espai bidimensional permetrà completar els conceptes necessaris per a comprendre el procés d'optimització lineal.

En el cas concret de dos coeficients i per al cas real, el vector de correlació creuada i la matriu d'autocorrelació correspondrien a les equacions següents:

$$\begin{bmatrix} r_{yx}[0] \\ r_{yx}[1] \end{bmatrix} = \begin{bmatrix} r_{xx}[0] & r_{xx}[1] \\ r_{xx}[-1] & r_{xx}[0] \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \end{bmatrix}$$

$$\begin{bmatrix} E[y[n]x[n]] \\ E[y[n]x[n-1]] \end{bmatrix} = \begin{bmatrix} E[x[n]x[n]] & E[x[n]x[n-1]] \\ E[x[n-1]x[n]] & E[x[n-1]x[n-1]] \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \end{bmatrix}$$

En aquest moment convé utilitzar aquesta matriu simple per a plantejar un parell d'escenaris que ens portaran a comprendre en més profunditat els conceptes exposats.

En un primer escenari, imaginarem un sistema de dos coeficients de l'estil $\hat{y}[n] = h[0]x[n] + h[1]x[n-1]$. En aquest cas es tracta d'un sistema amb una resposta impulsional de dos coeficients en què considerarem que $x[n]$ i $x[n–1]$ són senyals incorrelacionats, o el que és el mateix en els conceptes d'espais vectorials utilitzats, dos senyals ortogonals.

En aquest cas concret, el sistema d'equacions per resoldre seria:

$$\begin{bmatrix} E[y[n]x[n]] \\ E[y[n]x[n-1]] \end{bmatrix} = \begin{bmatrix} E[x[n]x[n]] & 0 \\ 0 & E[x[n-1]x[n-1]] \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \end{bmatrix}$$

El sistema d'equacions en aquest cas presentaria una solució trivial en què els coeficients serien:

$$h[0] = \frac{E[y[n]x[n]]}{E[x[n]x[n]]} \text{ i } h[1] = \frac{E[y[n]x[n-1]]}{E[x[n-1]x[n-1]]}$$

Si parem esment als coeficients, veiem que corresponen exactament al coeficient de projecció observat en el cas d'una variable. És a dir, el filtre ideal resultarà de projectar la informació desitjada sobre cadascuna de les variables d'entrada de manera independent i fer la combinació lineal de les aportacions de totes les variables d'entrada involucrades. En el cas de senyals d'entrada ortogonals no serà necessari resoldre cap sistema d'equacions, ja que els diferents senyals d'entrada no s'interfereixen entre si, de manera que la solució implica projectar sobre cadascun dels vectors d'entrada i sumar-ne les aportacions. La principal conclusió d'aquest exemple és que si tenim un espai de senyal d'entrada format per una base ortogonal, els coeficients del filtre es poden calcular de manera directa mitjançant les projeccions sobre cadascun dels elements d'entrada.

En l'escenari més genèric en què les dues variables d'entrada estan correlacionades serà necessari desacoblar la informació comuna que presenten totes dues mitjançant la resolució del sistema, que implicarà aplicar la inversa de la matriu de correlació a fi d'eliminar la informació creuada que hi ha en els diferents vectors d'entrada.

La solució en aquest cas seria la següent, que requereix la resolució de la matriu inversa per a desacoblar la informació conjunta dels senyals d'entrada:

$$\begin{bmatrix} E[x[n]x[n]] & E[x[n]x[n-1]] \\ E[x[n-1]x[n]] & E[x[n-1]x[n-1]] \end{bmatrix}^{-1} \begin{bmatrix} E[y[n]x[n]] \\ E[y[n]x[n-1]] \end{bmatrix} = \begin{bmatrix} h[0] \\ h[1] \end{bmatrix}$$

5.3. El principi d'ortogonalitat per a un filtre d'ordre *N*

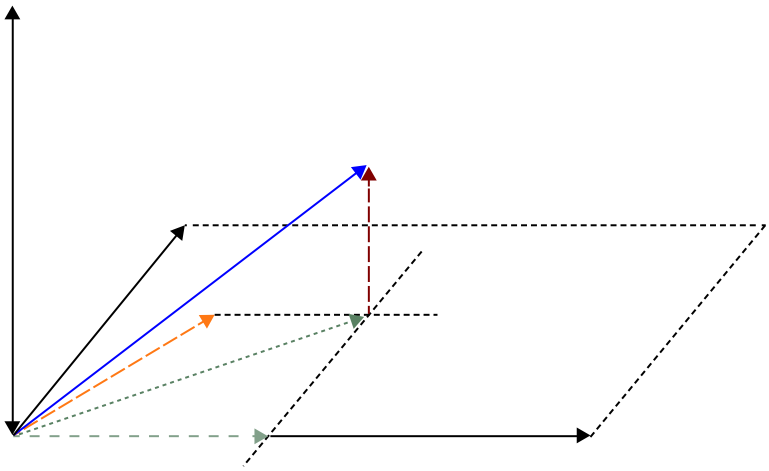
Tal com hem pogut observar en els subapartats anteriors, el desenvolupament matemàtic necessari per a abordar la solució del filtre de Wiener requereix la resolució d'un problema d'optimització aplicant derivades vectorials, que sense ser extremadament complexes en la resolució, sí que comporten una certa complexitat i poden induir a l'error.

La interpretació geomètrica presentada en el cas de dimensió 1 pot ser generalitzada a dimensió N per a portar-nos a un plantejament més senzill, directe i intuïtiu del sistema d'equacions. En el cas de dimensió 2, ens trobem amb la necessitat d'estimar un vector a partir d'una base de dos components.

Tal com s'observa en la figura 8, imaginem que pretenem fer una estimació del vector objectiu $y[n]$ a partir d'una combinació lineal de $x[n]$ (verd clar) i $x[n-1]$ (carabassa). Sense necessitat de recórrer a una resolució complexa d'un sistema d'equacions, i aplicant criteris purament geomètrics, veiem que el vector d'error tindrà mòdul mínim quan sigui ortogonal al pla format pels vectors

$x[n]$ i $x[n–1]$. Veiem igualment que els dos vectors, en tenir interferència creuada, se sumen aplicant la regla del paral·lelogram, que és una manera gràfica d'eliminar la informació creuada que hi ha entre tots dos vectors.

Figura 8



Exemple

Aquest exemple permet ampliar el principi d'ortogonalitat a un espai de dimensió N i afirmar que obtindrem la minimització de l'error quadràtic mitjà quan el vector d'error sigui ortogonal a tots i cadascun dels vectors de l'espai sobre el qual es fa la projecció, de tal manera que obtindrem la solució quan es compleixi la condició següent:

$$J = E\left[\left|y[n]-\hat{y}[n]\right|^2\right] = E\left[\left(y[n]-\sum_{k=0}^{N} x[n-k]h^*[k]\right)\left(y[n]-\sum_{k=0}^{N} x[n-k]h^*[k]\right)^*\right]$$

$$\frac{\partial J}{\partial h^*[i]} = -E\left[x[n-i]\left(y[n]-\sum_{k=0}^{N} x[n-k]h^*[k]\right)^*\right] = -E\left[x[n-i]e^*[n]\right] = 0$$

$$E\left[x[n-i]e^*[n]\right] = 0 \quad \forall\, i \in [0,1,\ldots,N]$$

En l'última fila veiem que tindrem la solució del sistema quan els productes escalars entre el vector d'error i els diferents valors del senyal d'entrada siguin 0, o el que és el mateix, quan el vector d'error sigui ortogonal a tots els components de l'entrada del filtre ($x[n]$, $x[n–1]$, ..., $x[n–K]$). La solució òptima s'obté quan el vector d'error sigui ortogonal al vector de dades d'entrada:

$$\left.\begin{aligned}
\langle x[n]e[n]\rangle &= E\left[x[n]e^*[n]\right] = 0 \\
\langle x[n-1]e[n]\rangle &= E\left[x[n-1]e^*[n]\right] = 0 \\
&\vdots \\
\langle x[n-N]e[n]\rangle &= E\left[x[n-N]e^*[n]\right] = 0
\end{aligned}\right\}$$

Aquest criteri es podrà aplicar per a qualsevol tipus de filtratge lineal òptim que s'optimitzi segons el criteri dels mínims quadrats, de manera que veiem que el principi d'ortogonalitat representa una conceptualització força útil en aplicacions de filtratge òptim, i més endavant, en filtratge adaptatiu.

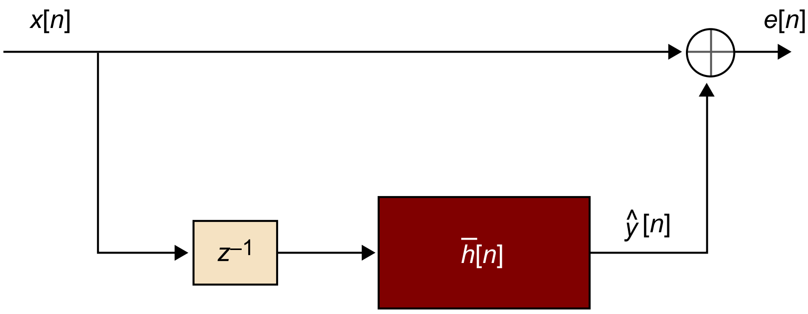
6. El predictor lineal

Dins de les aplicacions de processament avançat, n'hi ha una d'especial utilitat dins del camp de les telecomunicacions pel seu ús en tècniques de compressió de dades i aplicacions de veu, entre altres.

El predictor lineal és un cas particular de filtratge òptim en què el senyal objectiu és una mostra avançada del senyal d'entrada $y[n] = x[n+k]$, en què k serà l'ordre del predictor. El predictor d'ordre 1 prediu la mostra següent, el predictor d'ordre 2, $x[n+2]$, i així per a qualsevol ordre de predicció. En la majoria d'aplicacions pràctiques, l'ús del predictor se sol centrar en el predictor d'ordre 1, i com s'ha dit, és molt útil en tècnica de compressió de veu i també en tècniques d'equalització de canal.

En la figura següent observem la representació d'un predictor lineal.

Figura 8



Pel que fa a la resolució de les equacions de filtratge òptim, la problemàtica del predictor lineal no aporta cap complexitat afegida respecte de les anteriors: més aviat es tracta d'un cas particular del filtre de Wiener de gran utilitat en l'àmbit del processament.

L'única diferència és que el vector de correlació creuada es transforma en un vector d'autocorrelació desplaçat de la manera següent:

$$E[y\vec{x}] = E\begin{bmatrix} y^*[n]x[n] \\ y^*[n]x[n-1] \\ \vdots \\ y^*[n]x[n-N] \end{bmatrix} = \begin{bmatrix} r^*_{yx}(0) \\ r^*_{yx}(1) \\ \vdots \\ r^*_{yx}(N) \end{bmatrix} = \begin{bmatrix} r^*_{xx}(k) \\ r^*_{xx}(k+1) \\ \\ r^*_{xx}(k+N) \end{bmatrix}$$

Amb això, el sistema d'equacions per resoldre en el cas d'un predictor seria el mateix substituint el vector de correlació creuada segons l'equació anterior.

7. Conclusions

En aquest mòdul hem pogut estudiar les principals característiques del filtratge lineal òptim, i hem assentat les bases per a la resolució en aplicacions pràctiques, però sobretot com a introducció dels conceptes fonamentals del processament adaptatiu, atès que les característiques canviants de les aplicacions reals ens portaran en la majoria de casos a implementacions adaptatives.

El principal avantatge dels estimadors lineals respecte d'un altre tipus d'estimadors més complexos és que per a la resolució dels coeficients òptims del filtre tan sols és necessari el coneixement de les funcions estadístiques de primer i segon ordre (mitjanes i correlacions). Atès que en moltes aplicacions tan sols es disposa d'una única realització del procés, per a poder fer l'estimació de les funcions estadístiques haurem de recórrer a l'ús de les mostres temporals, és a dir, calcular les mitjanes i les correlacions a partir de les mostres temporals del procés. Recordem que aquest aspecte és només possible si el procés és estacionari i ergòdic. Serà en aquest cas quan resoldrem el problema utilitzant un càlcul estadístic de la mitjana i la correlació mitjançant les equacions següents:

$$E[x] = \lim_{L\to\infty}\frac{1}{2L+1}\sum_{n=-L}^{L}x[n] \quad E[yx] = \lim_{L\to\infty}\frac{1}{2L+1}\sum_{n=-L}^{L}x[n]y[n]$$

En el cas de tenir processos limitats en el temps, com serà habitual, restringirem els límits del sumatori a la finestra de dades de què disposem, atès que aquesta serà la millor estimació possible que podrem fer.

8. Exemples pràctics d'ús del filtratge lineal òptim

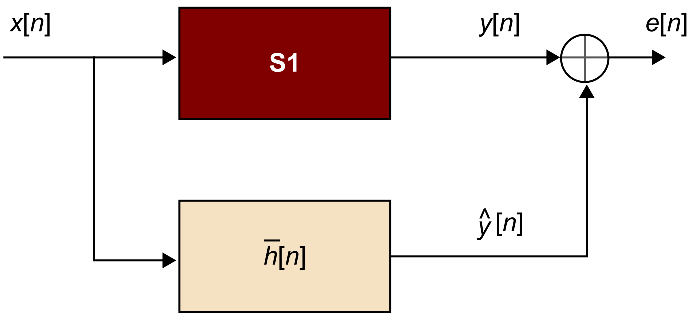
En aquest apartat, abans de finalitzar el mòdul, abordarem alguns exemples típics d'aplicacions lineals òptimes en el cas de filtres adaptatius a fi d'il·lustrar amb aquests exemples alguns dels conceptes més interessants presentats en el mòdul.

8.1. Identificació d'un sistema

Suposem que disposem d'un sistema S1, que és una caixa negra de la qual podem conèixer la resposta. És a dir, si hi introduïm un senyal d'entrada, podem mesurar com és la sortida del sistema. El sistema en si podria ser un equip electromecànic, una sala acústica, un canal de comunicacions o qualsevol altre exemple que se'ns pugui ocórrer.

El diagrama de blocs de disseny correspondrà a l'esquema que es presenta a la figura 10:

Figura 10



La idea és fer una estimació dels coeficients lineals que més s'aproximen a la resposta real del filtre. Si el filtre fos un sistema lineal i l'ordre de l'estimador fos prou gran, l'estimació seria pràcticament perfecta sempre que s'escollissin apropiadament els senyals de test del sistema.

Quin tipus de senyal d'entrada seria convenient utilitzar per a la implementació del filtre òptim?

Si ens interessa fer una estimació completa de la resposta del filtre, necessitem escollir com a senyal d'entrada $x[n]$ un senyal que sigui ric en freqüència, és a dir, que sigui capaç d'estimular el sistema en tota la varietat possible de senyals d'entrada. Recordem novament que el filtre òptim farà la millor estimació possible amb la informació de la qual disposa, però que això no vol dir que l'estimació sigui bona.

Exemple

Si el senyal d'excitació és un senyal sinusoïdal que tan sols excita una determinada freqüència, el filtre de Wiener, per alt que sigui l'ordre del filtre, tan sols podrà determinar la resposta estimada a la freqüència del senyal d'entrada.

Tenint en compte aquest aspecte, veiem que resulta important en problemes d'identificació de sistemes que el senyal d'entrada cobreixi les màximes dimensions de l'espai vectorial sobre el qual es treballa, i això, traduït al camp del processament de senyal, seria que tingui components de totes les freqüències possibles.

Exemple

Un exemple de senyal d'entrada seria la funció delta discreta $\delta[n]$, que com bé sabem té un espectre freqüencial pla, o un soroll blanc, que és capaç de cobrir tot el marge freqüencial. Si el sistema és un motor o una sala acústica, una delta no seria una funció pràctica, atès que en el món analògic no la podríem implementar, amb la qual cosa seria més convenient treballar amb una funció de tipus soroll blanc.

Com sabrem com és l'ordre apropiat del filtre?

Aquesta pregunta també resulta interessant, atès que *a priori* no tenim cap informació que ens indiqui com serà l'ordre apropiat. Si per algun motiu coneixem aquest paràmetre, procedirem a escollir l'ordre del filtre que s'adapti al coneixement que tinguem de la realitat.

El que ocorrerà si escollim un ordre menor serà que el sistema no es podrà adaptar completament i tindrem un error d'estimació més gran i una identificació del sistema de pitjor qualitat. Una bona solució és fer diferents avaluacions incrementant l'ordre del filtre fins que l'error d'estimació no decreixi.

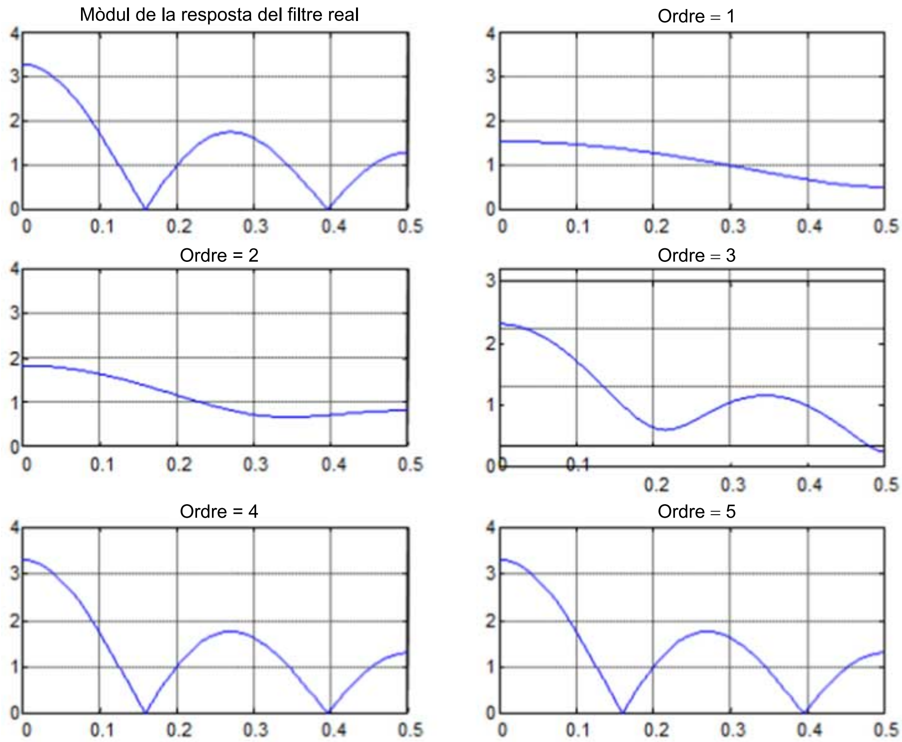
Si escollim un ordre superior al del sistema real, l'estimació serà bona però els últims coeficients de la resposta impulsional del filtre estimat seran zero. *A priori*, això no representa un gran problema des del punt de vista pràctic, però sí des del punt de vista computacional, atès que la resolució del filtre òptim requereix el càlcul de la inversa d'una matriu, una operació computacionalment costosa, i el cost de la qual augmenta de manera exponencial. Si la implementació es fa en un DSP o una FPGA, això pot representar un problema per al maquinari del sistema.

En la figura 11 es mostra la resposta freqüencial estimada mitjançant un filtre de Wiener per a un sistema amb resposta impulsional real $h[n]$ = [1 0,5 0,3 0,5 1]. En aquest cas s'ha utilitzat com a senyal d'entrada soroll blanc gaussià i s'ha resolt el filtre de Wiener mitjançant l'estimació temporal de les funcions de correlació.

S'observa que en el cas del filtre d'ordre 1 (2 coeficients) l'estimació del sistema és força pobra, però fa una estimació de tipus passabaix a fi d'aproximar-se de la millor manera possible a les bandes energètiques predominants del sistema real. Aquest factor va ocorrent a mesura que augmenten els coeficients, i veiem que la forma es va modulant fins a adaptar-se completament en el filtre d'ordre

4 (5 coeficients). Si augmentem encara més l'ordre del filtre veiem que ja no hi ha cap millora en l'estimació, ja que l'últim coeficient de la resposta es farà zero.

Figura 11



Annex

Espais de Hilbert

Un espai de Hilbert és una generalització d'un espai euclidià que ens permet ampliar els conceptes de geometria euclidiana a altres tipus de vectors. L'espai de Hilbert està format per un parell (V, <,>), en què V és un espai vectorial ben definit (de dimensió finita o infinita) sobre el cos dels nombres reals o complexos, i <,> és un producte interior sobre V, amb la condició addicional que l'espai ha de ser mètric complet sobre la distància induïda pel producte intern.

Que <,> sigui un producte intern sobre V implica que $\langle\ ,\ \rangle: V \times V \to K$, en què $K \in \{\mathbb{R}, \mathbb{Q}\}$, que és el cos subjacent sobre el qual està definit el producte escalar. L'operador <,> compleix les condicions següents per a tots els elements $x, y, z \in V$ i per a tots els escalars $a, b \in K$.

- $< x, y > = \overline{< y, x >}$ en què la barra denota conjugació complexa.
- $< ax + by, z > = a < x, z > + b < y, z >$
- $< x, x > \geq 0$, i la igualtat només es compleix si, i només si, $x = 0$.

Sobre aquest producte intern ben definit s'indueix una norma, que serà l'operador que servirà per a poder mesurar distàncies dins de l'espai vectorial, definida com $\| x \| = \sqrt{\langle x, x \rangle}$ que compleix les propietats següents:

- $\| x \| = 0$ si, i només si, $x = 0$.
- $\| ax \| = |a| \| x \|$
- $|< x, y >| \leq \| x \| \, \| y \|$ (desigualtat de Cauchy-Schwarz).
- $\| x + y \| \leq \| x \| + \| y \|$ (desigualtat triangular).

Sense entrar en més profunditat en els detalls matemàtics exposats, ens hem de quedar amb el fet que els espais de Hilbert permeten fer mesures i treballar amb distàncies segons una norma ben definida que ens permetrà abordar problemes d'optimització sobre espais vectorials de variables aleatòries.

Regles de derivació amb variable complexa i amb funcions escalars de vectors complexos

Notació

Variable complexa: $x = x_R + jx_I \in C$ on $x_R = \mathrm{Re}[x] \in \Re$, $x_I = \mathrm{Im}[x] \in \Re$

Complex conjugat: $x^* = x_R - jx_I \in C$

Derivació de variable complexa	Exemples
$\frac{\partial}{\partial x} = \frac{1}{2}\left(\frac{\partial}{\partial x_R} - j\frac{\partial}{\partial x_I}\right)$	$\frac{\partial x}{\partial x} = \frac{1}{2}\left(\frac{\partial(x_R + jx_I)}{\partial x_R} - j\frac{\partial(x_R + jx_I)}{\partial x_I}\right) = \frac{1}{2}(1 - j \times j) = 1$ $\frac{\partial x^*}{\partial x} = \frac{1}{2}\left(\frac{\partial(x_R - jx_I)}{\partial x_R} - j\frac{\partial(x_R - jx_I)}{\partial x_I}\right) = \frac{1}{2}(1 - j \times (-j)) = 0$ $\frac{\partial x^* x}{\partial x} = \frac{\partial \lvert x\rvert^2}{\partial x} = \frac{\partial(x_R^2 + x_I^2)}{\partial x} = \frac{1}{2}\left(\frac{\partial(x_R^2 + x_I^2)}{\partial x_R} - j\frac{\partial(x_R^2 + x_I^2)}{\partial x_I}\right) =$ $= \frac{1}{2}(2x_R - j \times (2x_I)) = x^*$
$\frac{\partial}{\partial x^*} = \frac{1}{2}\left(\frac{\partial}{\partial x_R} + j\frac{\partial}{\partial x_I}\right)$	$\frac{\partial x}{\partial x^*} = \frac{1}{2}\left(\frac{\partial(x_R + jx_I)}{\partial x_R} + j\frac{\partial(x_R + jx_I)}{\partial x_I}\right) = \frac{1}{2}(1 + j \times j) = 0$ $\frac{\partial x^*}{\partial x^*} = \frac{1}{2}\left(\frac{\partial(x_R - jx_I)}{\partial x_R} + j\frac{\partial(x_R - jx_I)}{\partial x_I}\right) = \frac{1}{2}(1 + j \times (-j)) = 1$ $\frac{\partial x2}{\partial x^*} = \frac{\partial(x_R^2 + j2x_I x_R - x_I^2)}{\partial x^*} = \frac{1}{2}(2x_R + j2x_I + j \times (2jx_R - 2x_I)) = 0$ $\frac{\partial x^* x}{\partial x^*} = \frac{\partial \lvert x\rvert^2}{\partial x^*} = \frac{\partial(x_R^2 + x_I^2)}{\partial x^*} = \frac{1}{2}\left(\frac{\partial(x_R^2 + x_I^2)}{\partial x_R} + j\frac{\partial(x_R^2 + x_I^2)}{\partial x_I}\right) =$ $= \frac{1}{2}(2x_R + j \times (2x_I)) = x$

Notació

Variable vector complex: $\mathbf{x} = [x_1 \ x_2 \ \cdots \ x_N]^T \in C^{N\times 1}$ vector columna, on $x_k \in C, \ k = 1,2,\ldots,N$

Constant vector complex: $\mathbf{a} = [a_1 \ a_2 \ \cdots \ a_N]^T \in C^{N\times 1}$ vector columna, on $a_k \in C, \ k = 1,2,\ldots,N$

$\mathbf{b} = [b_1 \ b_2 \ \cdots \ b_N]^T \in C^{N\times 1}$ vector columna, on $b_k \in C, \ k = 1,2,\ldots,N$

Constant vector complex: $\mathbf{a} = [a_1 \ a_2 \ \cdots \ a_N]^T \in C^{N\times 1}$ vector columna, on $a_k \in C, \ k = 1,2,\ldots,N$

$\mathbf{b} = [b_1 \ b_2 \ \cdots \ b_N]^T \in C^{N\times 1}$ vector columna, on $b_k \in C, \ k = 1,2,\ldots,N$

Constant matriu complexa: $\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1N} \\ r_{21} & r_{22} & & r_{2N} \\ \vdots & & \ddots & \vdots \\ r_{N1} & r_{N2} & \cdots & r_{NN} \end{bmatrix} \in C^{N\times N}$ matriu quadrada, on $r_{ij} \in C, \ i = 1,2,\ldots,N; \ j = 1,2,\ldots,N$

Operador hermític (transposició+conjugació): $\mathbf{x}^H = (\mathbf{x}^T)^* = (\mathbf{x}^*)^T = [x_1^* \ x_2^* \ \cdots \ x_N^*]^T \in C^{N\times 1}$

Derivació vectorial (gradient)	Exemples
$\frac{\partial}{\partial \mathbf{x}} = \left[\frac{\partial}{\partial x_1} \;\; \frac{\partial}{\partial x_2} \;\; \cdots \;\; \frac{\partial}{\partial x_N}\right]^T$	$\frac{\partial\left(2+x_1x_2-x_1^2\right)}{\partial \mathbf{x}} = \begin{bmatrix}\partial\left(2+x_1x_2-x_1^2\right)/\partial x_1 \\ \partial\left(2+x_1x_2-x_1^2\right)/\partial x_2\end{bmatrix} = \begin{bmatrix} x_2-2x_1 \\ x_1\end{bmatrix}$
$\frac{\partial}{\partial \mathbf{x}^*} = \left[\frac{\partial}{\partial x_1^*} \;\; \frac{\partial}{\partial x_2^*} \;\; \cdots \;\; \frac{\partial}{\partial x_N^*}\right]^T$	$\frac{\partial\left(2+x_1^*x_2-x_1^2\right)}{\partial \mathbf{x}^*} = \begin{bmatrix}\partial\left(2+x_1^*x_2-x_1^2\right)/\partial x_1^* \\ \partial\left(2+x_1^*x_2-x_1^2\right)/\partial x_2^*\end{bmatrix} = \begin{bmatrix} x_2 \\ 0\end{bmatrix}$
$\frac{\partial(a_1)}{\partial \mathbf{x}} = \mathbf{0}$	$\frac{\partial(1+j)}{\partial \mathbf{x}} = \begin{bmatrix}\partial(1+j)/\partial x_1 \\ \partial(1+j)/\partial x_2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} = \mathbf{0}$
$\frac{\partial(a_1)}{\partial \mathbf{x}^*} = \mathbf{0}$	$\frac{\partial(1+j)}{\partial \mathbf{x}^*} = \begin{bmatrix}\partial(1+j)/\partial x_1^* \\ \partial(1+j)/\partial x_2^*\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} = \mathbf{0}$
$\frac{\partial(\mathbf{a}^H\mathbf{x})}{\partial \mathbf{x}} = \mathbf{a}^*$	$\frac{\partial\left(\left[1+j \;\; 2-j\right]\begin{bmatrix}x_1\\x_2\end{bmatrix}\right)}{\partial \mathbf{x}} = \begin{bmatrix}\partial\left((1+j)x_1+(2-j)x_2\right)/\partial x_1 \\ \partial\left((1+j)x_1+(2-j)x_2\right)/\partial x_2\end{bmatrix} = \begin{bmatrix}1+j\\2-j\end{bmatrix}$
$\frac{\partial(\mathbf{x}^H\mathbf{a})}{\partial \mathbf{x}^*} = \mathbf{a}$	$\frac{\partial\left(\left[x_1^* \;\; x_2^*\right]\begin{bmatrix}1+j\\2-j\end{bmatrix}\right)}{\partial \mathbf{x}^*} = \begin{bmatrix}\partial\left((1+j)x_1^*+(2-j)x_2^*\right)/\partial x_1^* \\ \partial\left((1+j)x_1^*+(2-j)x_2^*\right)/\partial x_2^*\end{bmatrix} = \begin{bmatrix}1+j\\2-j\end{bmatrix}$
$\frac{\partial(\mathbf{a}^H\mathbf{x})}{\partial \mathbf{x}^*} = \frac{\partial(\mathbf{x}^H\mathbf{a})}{\partial \mathbf{x}} = \mathbf{0}$	$\frac{\partial\left(\left[x_1^* \;\; x_2^*\right]\begin{bmatrix}1+j\\2-j\end{bmatrix}\right)}{\partial \mathbf{x}} = \begin{bmatrix}\partial\left((1+j)x_1^*+(2-j)x_2^*\right)/\partial x_1 \\ \partial\left((1+j)x_1^*+(2-j)x_2^*\right)/\partial x_2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} = \mathbf{0}$
$\frac{\partial(\mathbf{x}^H\mathbf{R}\mathbf{x})}{\partial \mathbf{x}} = (\mathbf{x}^H\mathbf{R})^T$	$\frac{\partial\left(\left[x_1^* \;\; x_2^*\right]\begin{bmatrix}r_{11} & r_{12}\\r_{21} & r_{22}\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}\right)}{\partial \mathbf{x}} = \frac{\partial\left(r_{11}x_1^*x_1+r_{12}x_1^*x_2+r_{21}x_2^*x_1+r_{22}x_2^*x_2\right)}{\partial \mathbf{x}} = $ $= \begin{bmatrix}\partial/\partial x_1\\ \partial/\partial x_2\end{bmatrix} = \begin{bmatrix}r_{11}x_1^*+r_{21}x_2^*\\ r_{12}x_1^*+r_{22}x_2^*\end{bmatrix} = \begin{bmatrix}r_{11} & r_{21}\\ r_{12} & r_{22}\end{bmatrix}\begin{bmatrix}x_1^*\\x_2^*\end{bmatrix} = \mathbf{R}^T\mathbf{x}^* = (\mathbf{x}^H\mathbf{R})^T$
$\frac{\partial(\mathbf{x}^H\mathbf{R}\mathbf{x})}{\partial \mathbf{x}^*} = \mathbf{R}\mathbf{x}$	$\frac{\partial\left(\left[x_1^* \;\; x_2^*\right]\begin{bmatrix}r_{11} & r_{12}\\r_{21} & r_{22}\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}\right)}{\partial \mathbf{x}^*} = \frac{\partial\left(r_{11}x_1^*x_1+r_{12}x_1^*x_2+r_{21}x_2^*x_1+r_{22}x_2^*x_2\right)}{\partial \mathbf{x}^*} = $ $= \begin{bmatrix}\partial/\partial x_1^*\\ \partial/\partial x_2^*\end{bmatrix} = \begin{bmatrix}r_{11}x_1+r_{12}x_2\\ r_{21}x_1+r_{22}x_2\end{bmatrix} = \begin{bmatrix}r_{11} & r_{12}\\ r_{21} & r_{22}\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \mathbf{R}\mathbf{x}$

Demostració de l'aplicació de les regles de derivació en l'obtenció de la solució òptima del filtre de Wiener

En l'apartat 5.2. *Filtre lineal òptim d'ordre N* d'aquest mòdul es parteix de la següent equació:

$$E\left[|e|^2\right]_{MIN} = E\left[yy^H\right] - \vec{h}^H E\left[\vec{x}y^H\right] - E\left[y\vec{x}^H\right]\vec{h} + \vec{h}^H E\left[\vec{x}\vec{x}^H\right]\vec{h}\Big|_{MIN}$$

El que es busca és trobar el valor de $\vec{h}^*$ que minimitzi aquesta funció de cost, per tant, el que cal fer és:

$$\frac{\partial E[|e|^2]}{\partial \vec{h}^*} = \underbrace{\frac{E[yy^H]}{\partial \vec{h}^*}}_{(1)} - \underbrace{\frac{\vec{h}^H E[\vec{x}\,y^H]}{\partial \vec{h}^*}}_{(2)} - \underbrace{\frac{E[y\vec{x}^H]\vec{h}}{\partial \vec{h}^*}}_{(3)} + \underbrace{\frac{\vec{h}^H E[\vec{x}\,\vec{x}^H]\vec{h}}{\partial \vec{h}^*}}_{(4)} = 0$$

Vegem les regles de derivació que cal aplicar en cada cas:

$$(1) \rightarrow \left\{\frac{\partial(a_1)}{\partial \mathbf{x}^*} = \mathbf{0}\right\} \rightarrow \frac{E[yy^H]}{\partial \vec{h}^*} = 0$$

$$(2) \rightarrow \left\{\frac{\partial(\mathbf{x}^H\mathbf{a})}{\partial \mathbf{x}^*} = \mathbf{a}\right\} \rightarrow \frac{\vec{h}^H E[\vec{x}\,y^H]}{\partial \vec{h}^*} = E[\vec{x}\,y^H]$$

$$(3) \rightarrow \left\{\frac{\partial(\mathbf{a}^H\mathbf{x})}{\partial \mathbf{x}^*} = \frac{\partial(\mathbf{x}^H\mathbf{a})}{\partial \mathbf{x}} = \mathbf{0}\right\} \rightarrow \frac{E[y\vec{x}^H]\vec{h}}{\partial \vec{h}^*} = 0$$

$$(4) \rightarrow \left\{\frac{\partial(\mathbf{x}^H\mathbf{R}\mathbf{x})}{\partial \mathbf{x}^*} = \mathbf{R}\mathbf{x}\right\} \rightarrow \frac{\vec{h}^H E[\vec{x}\,\vec{x}^H]\vec{h}}{\partial \vec{h}^*} = E[\vec{x}\,\vec{x}^H]\vec{h}$$

Per tant, s'arriba al resultat final que apareix al mateix l'apartat:

$$\frac{\partial E[|e|^2]}{\partial \vec{h}^*} = -E[\vec{x}\,y^H] + E[\vec{x}\,\vec{x}^H]\vec{h} = 0$$