WIDE-SENSE LINEAR MEAN SQUARE ESTIMATION AND PREDICTION

Bernard Picinbono
Laboratoire des Signaux et Systèmes
Supélec, Plateau de Moulon, 91190 Gif-sur-Yvette
France

ABSTRACT

Optimum mean square estimation of a random variable y in terms of an observation vector $\mathbf{x}$ is realized by the conditional expectation value. When $\mathbf{x}$ and y are real and jointly normal this expectation is linear in $\mathbf{x}$. This is no longer the case when $\mathbf{x}$ and y are complex and jointly normal and the expectation is linear in $\mathbf{x}$ and in its complex conjugate $\mathbf{x}^*$, which introduces a widely linear procedure. The purpose of this paper is to study the properties of widely linear systems for estimation and prediction. The structure of such systems is calculated and the gain in performance is analyzed. The results are applied to autoregressive signals, which introduces widely linear prediction.

I. INTRODUCTION

Mean square estimation (MSE) is one of the most fundamental techniques of Statistical Signal Processing [1]. The basic problem can be stated as follows: let y be a scalar random variable to be estimated (estimandum) in terms of an observation which is a random vector $\mathbf{x}$. The estimate $\hat{y}$ that minimizes the MS error is then the conditional expectation value $E[y \mid \mathbf{x}]$ [2]. This result is usually given when $\mathbf{x}$ and y are real. However it remains valid when these quantities are complex valued [3]. If $\mathbf{x}$ and y are zero-mean, real, and jointly *normal*, the estimate $\hat{y}$ becomes linear and can be expressed as $\mathbf{h}^T\mathbf{x}$. This is no longer true when $\mathbf{x}$ and y are *complex* and jointly normal, and it appears that the conditional expectation takes the form

$$\hat{y} = \mathbf{u}^H\mathbf{x} + \mathbf{v}^H\mathbf{x}^*, \qquad (1.1)$$

where $\mathbf{u}$ and $\mathbf{v}$ are complex vectors, H means the complex conjugation and transposition (Hermitian transposition), and $\mathbf{x}^*$ is the complex conjugate of $\mathbf{x}$. In fact the conditional expectation $E[y \mid \mathbf{x}]$ can be written as $E[y_1 \mid \mathbf{x}_1, \mathbf{x}_2] + jE[y_2 \mid \mathbf{x}_1, \mathbf{x}_2]$, where y_1, y_2, $\mathbf{x}_1$ and $\mathbf{x}_2$ are the real and imaginary parts of y and $\mathbf{x}$. As a consequence $E[y \mid \mathbf{x}]$ is linear both in $\mathbf{x}_1$ and $\mathbf{x}_2$ with complex coefficients. Writing these vectors in terms of $\mathbf{x}$ and $\mathbf{x}^*$, $E[y \mid \mathbf{x}]$ takes the form (1.1).

It is clear that $\hat{y}$ given by (1.1) is no longer a linear function of $\mathbf{x}$. However the moment of order k of $\hat{y}$ is explicitly defined from the moments of the same order of $\mathbf{x}$ and $\mathbf{x}^*$. This is a reason why $\hat{y}$ is called *widely linear*, or linear in the wide sense. The purpose of this paper is to study the structure and properties of widely linear MSE, or WLMSE. It is obvious that WLMSE appears only when complex data are used and becomes again strictly linear in the real case. However the use of complex data is very common in many fields of Signal Processing such as spectral analysis and antenna or beam forming design [4].

II. WIDELY LINEAR MSE

The problem is then to find the vectors $\mathbf{u}$ and $\mathbf{v}$ in such a way that (1.1) gives the minimum mean square error $E[\,|y - \hat{y}|^2]$. For this purpose the first point to note is that the set of scalar complex random variables $z(\omega)$ in the form

$$z(\omega) = \mathbf{a}^H\mathbf{x}(\omega) + \mathbf{b}^H\mathbf{x}^*(\omega), \qquad (2.1)$$

where $\mathbf{a}$ and $\mathbf{b}$ belong to $\mathbb{C}^N$, constitutes a linear space. It becomes a Hilbert subspace with the scalar product $<z, z'> = E(z^*z')$. As a result the WLMSE $\hat{y}$ is the projection of y onto this subspace and is characterized by the orthogonality principle

$$(y - \hat{y}) \perp\!\!\!\perp \mathbf{x} \ ; \ (y - \hat{y}) \perp\!\!\!\perp \mathbf{x}^*. \qquad (2.2)$$

The symbol $\perp\!\!\!\perp$ means that all the components of $\mathbf{x}$ or $\mathbf{x}^*$ are orthogonal to $y - \hat{y}$ with the previous scalar product. As a consequence, these equations can be written in terms of expectations, which yields

$$E(\hat{y}^*\mathbf{x}) = E(y^*\mathbf{x}) \qquad (2.3)$$

$$E(\hat{y}^*\mathbf{x}^*) = E(y^*\mathbf{x}^*). \qquad (2.4)$$

Replacing $\hat{y}$ by (1.1) gives

$$\Gamma\mathbf{u} + \mathbb{C}\mathbf{v} = \mathbf{r} \qquad (2.5)$$

$$\mathbb{C}^*\mathbf{u} + \Gamma^*\mathbf{v} = \mathbf{s}^*, \qquad (2.6)$$

where

$$\Gamma = E[\mathbf{x}\mathbf{x}^H] \ ; \ \mathbb{C} = E[\mathbf{x}\mathbf{x}^T] \qquad (2.7)$$

$$\mathbf{r} = E[y^*\mathbf{x}] \ ; \ \mathbf{s} = E[y\mathbf{x}]. \qquad (2.8)$$

The solution of (2.7) and (2.8) can be expressed as

$$\mathbf{u} = [\Gamma - \mathbb{C}\Gamma^{-1*}\mathbb{C}^*]^{-1}[\mathbf{r} - \mathbb{C}\Gamma^{-1*}\mathbf{s}^*] \qquad (2.9)$$

$$\mathbf{v} = [\Gamma^* - \mathbb{C}^*\Gamma^{-1}\mathbb{C}]^{-1}[\mathbf{s}^* - \mathbb{C}^*\Gamma^{-1}\mathbf{r}]. \qquad (2.10)$$

0-7803-2431-5/95 $4.00 © 1995 IEEE

The corresponding mean square error is also deduced from the projection theorem by

$$\varepsilon^2 = E[\,|y|^2] - E[\,|\hat{y}|^2]. \qquad (2.11)$$

By using (1.1) and the equations (2.5) and (2.6), we obtain

$$\varepsilon^2 = E[\,|y|^2] - (\mathbf{u}^H\mathbf{r} + \mathbf{v}^H\mathbf{s}^*). \qquad (2.12)$$

This error is smaller than ε_L^2, the error that is obtained with a strictly LMSE and equal to $E[\,|y|^2] - \mathbf{r}^H\Gamma^{-1}\mathbf{r}$. The advantage of the wide sense linear procedure over the strict sense one is characterized by the quantity $\delta\varepsilon^2 = \varepsilon_L^2 - \varepsilon^2$ which can be expressed as

$$\delta\varepsilon^2 = [\mathbf{s}^* - \mathbb{C}^*\Gamma^{-1}\mathbf{r}]^H[\Gamma^* - \mathbb{C}^*\Gamma^{-1}\mathbb{C}]^{-1}[\mathbf{s}^* - \mathbb{C}^*\Gamma^{-1}\mathbf{r}], \qquad (2.13)$$

This quantity is non-negative. In fact the matrix $[\Gamma^* - \mathbb{C}^*\Gamma^{-1}\mathbb{C}]$ is positive definite if $\mathbf{x}$ is not real. and consequently $\delta\varepsilon^2 = 0$ is possible only when $\mathbf{s}^* - \mathbb{C}^*\Gamma^{-1}\mathbf{r} = \mathbf{0}$.

At this step it is worth pointing out that all the previous calculations could be realized by using only real quantities. For this purpose it suffices to write (1.1) in the form

$$\hat{y} = \mathbf{h}_1^T\mathbf{x}_1 + \mathbf{h}_2^T\mathbf{x}_2\,, \qquad (2.14)$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are the real and imaginary parts of $\mathbf{x}$, and to split this equation in two real equations because $\mathbf{h}_1$ and $\mathbf{h}_2$ are complex and to calculate the four vectors $\mathbf{h}$ appearing in those equations. However by doing so the compact expression of the complex quantity $\hat{y}$ is less obvious and (2.13) takes a much more complex form. Furthermore the comparison with the strictly linear procedure characterized by $\mathbf{v} = 0$ is less obvious with real quantities than with (1.1). Finally in the strictly linear case, nobody has the idea of transforming the complex Wiener-Hopf equation $\Gamma\mathbf{u} = \mathbf{c}$ in a set of two real equations.

III. EXAMPLES

3.1 Jointly Circular Case

This situation is characterized by

$$\mathbb{C} = \mathbb{0} \quad ; \quad \mathbf{s} = \mathbf{0}. \qquad (3.1)$$

This assumption is well-known in the normal case (see [2], p. 118) and is sometimes used in the definition of complex normal random vectors. Furthermore one can show that under some conditions the Fourier components of stationary signals are complex circular random variables. The analytic signal of a real stationary signal is also second-order circular. The term of circularity comes from the fact that if (3.1) holds, the random vectors $\mathbf{x}$ and $\mathbf{x}\exp(j\alpha)$ have the same second-order properties for any α. Note that (3.1) characterizes only second-order circularity, and the concept can be extended when using higher-order statistics. Note also that (3.1) means a joint circularity and is then an assumption on $\mathbf{x}$ and y.

It immediately results from (2.10) that (3.1) implies $\mathbf{v} = \mathbf{0}$. Similarly (2.9) gives $\mathbf{u} = \Gamma^{-1}\mathbf{r}$. Thus the assumption of joint circularity implies that the WLMSE (1.1) becomes strictly linear. It is also clear that (2.13) gives $\delta\varepsilon^2 = 0$, and the conclusion is that, in the case of a joint circularity, a strictly linear system is sufficient to reach the best performance. This is one of the arguments justifying the interest of circularity.

3.2. Circular Observation

Suppose now that the second assumption (3.1) is deleted. This means that circularity is only valid for the observation and is characterized by $\mathbb{C} = \mathbb{0}$, while no specific assumption is introduced for the estimandum y. In this case (2.9) and (2.10) are greatly simplified and become

$$\mathbf{u} = \Gamma^{-1}\mathbf{r} \quad ; \quad \mathbf{v}^* = \Gamma^{-1}\mathbf{s}. \qquad (3.2)$$

This means that the term $\mathbf{u}^H\mathbf{x}$ in (1.1) is the same as the one obtained when using strictly linear estimation. This fact can be explained by noting that the circularity assumption implies that the vectors $\mathbf{x}$ and $\mathbf{x}^*$ are uncorrelated. Thus the Hilbert subspaces generated by $\mathbf{x}$ and $\mathbf{x}^*$ are orthogonal and taking into account $\mathbf{x}^*$ does not change the term coming from $\mathbf{x}$ only. This also explains the simplification of (2.13) that becomes

$$\delta\varepsilon^2 = \mathbf{s}^H\Gamma^{-1}\mathbf{s}. \qquad (3.3)$$

Thus a non-zero vector $\mathbf{s}$ necessarily implies an increment of the performance of estimation when using WLMSE instead of LMSE.

3.3. Case of a Real Estimandum y

The estimation of a real quantity from complex data appears in many situations as, for example, when the observation comes from Fourier components of a real signal some real parameters of which have to be estimated. Suppose then that y is real, $\mathbf{x}$ still being complex. This obviously implies that $\mathbf{r} = \mathbf{s}$ in (2.8). It results from (2.9) and (2.10) that $\mathbf{u} = \mathbf{v}^*$ and consequently

$$\hat{y} = 2\,\mathrm{Re}(\mathbf{u}^H\mathbf{x}). \qquad (3.4)$$

Similarly the estimation error takes the form

$$\varepsilon^2 = E(y^2) - 2\,\mathrm{Re}(\mathbf{u}^H\mathbf{r}) \qquad (3.5)$$

The main property of the estimate (3.4) is that it is real, while there is no reason for the strictly linear estimate to be real, which is not convenient when estimating a real quantity.

The advantage of WLMSE with respect to LMSE is even more clear when the observation $\mathbf{x}$ is circular. In fact, as seen previously in subsection 3.2, the vector $\mathbf{u}$ is the same as the one that must be used to realize the MSE of y with the strictly linear method. Thus, by using this vector, the two estimators become

$$\hat{y}_L = \mathbf{u}^H \mathbf{x} \tag{3.6}$$

$$\hat{y}_{WL} = 2\,\mathrm{Re}(\mathbf{u}^H \mathbf{x}) \tag{3.7}$$

and the corresponding errors are

$$\varepsilon_L^2 = E(y^2) - \mathbf{u}^H \mathbf{r} \tag{3.8}$$

$$\varepsilon_{WL}^2 = E(y^2) - 2\mathbf{u}^H \mathbf{r}. \tag{3.9}$$

Note that the quantity $\mathbf{u}^H\mathbf{r}$ is positive because it is equal to $\mathbf{u}^H \Gamma\, \mathbf{u}$ and Γ is a positive definite matrix. In conclusion, the wide sense linear estimator (3.7) provides a real estimate and a decrease of the error that is twice as great as the strictly linear estimate, which in general is complex. It is also clear that $\delta\varepsilon^2 = \mathbf{u}^H \mathbf{r}$.

These results can especially be applied to the classical non-causal Wiener filtering. Let $x(t)$ and $y(t)$ be two jointly stationary continuous-time signals. Suppose further that $y(t)$ is real and that $x(t)$ is complex and second-order circular, which means that $E[x(t)x(t-\tau)] = 0$. This is, for example, the case of the analytical signal of a real stationary signal (see [2], p. 230). The strictly linear estimate of $y(t)$ can be expressed as (see [2], p. 450)

$$\hat{y}_L(t) = \int h(t-\theta)\, x(\theta)\, d\theta, \tag{3.10}$$

and $h(t)$ is determined by the orthogonality equation

$$\int h(\theta)\, \gamma_x(\tau - \theta)\, d\theta = \gamma_{yx}(\tau), \tag{3.11}$$

where $\gamma_x(\tau) = E[x(t)x^*(t-\tau)]$ and $\gamma_{yx}(\tau) = E[y(t)x^*(t-\tau)]$. A Fourier transformation yields the frequency response

$$H(\nu) = [\Gamma_x(\nu)]^{-1}\Gamma_{yx}(\nu), \tag{3.12}$$

where $\Gamma_x(\nu)$ is the power spectrum of $x(t)$ and $\Gamma_{yx}(\nu)$ the Fourier transform of $\gamma_{yx}(\tau)$. The mean square error is then

$$\varepsilon_L^2 = \int [\Gamma_x(\nu)]^{-1}\, [\,\Gamma_x(\nu)\Gamma_y(\nu) - |\Gamma_{yx}(\nu)|^2]\, d\nu, \tag{3.13}$$

On the other hand the wide sense LMSE of $y(t)$ is, as in (3.7),

$$\hat{y}_{WL}(t) = 2y_1(t), \tag{3.14}$$

where $y_1(t)$ is the real part of $\hat{y}_L(t)$ given by (3.10). However, as $x(t)$ is circular we have $E[\hat{y}_L^2(t)] = 0$, and then

$$E[\,|\hat{y}_L(t)|^2] = E[y_1^2(t)] + E[y_2^2(t)] = 2E[y_1^2(t)], \tag{3.15}$$

where $y_2(t)$ is the imaginary part of $\hat{y}_L(t)$. As a result we have $E[\hat{y}_{WL}^2(t)] = 2E[\,|\hat{y}_L(t)|^2]$ and (3.13) becomes

$$\varepsilon_{WL}^2 = \int [\Gamma_x(\nu)]^{-1}\, [\,\Gamma_x(\nu)\Gamma_y(\nu) - 2\,|\Gamma_{yx}(\nu)|^2]\, d\nu. \tag{3.16}$$

This shows the advantage of the wide sense linear procedure over the strictly linear one.

The conclusion is quite general: when the complex data are not jointly circular, the classical LMSE is not the best procedure of estimation that only uses the second order properties of the signals. This conclusion confirms and extends the results of [5]

3.4. Singular Estimation

The estimation is singular when the mean square error is zero. If the wide sense linear mean square error (2.12) is zero, the estimandum y belongs to the Hilbert subspace defined by (2.1) and can then be written as

$$y = \mathbf{a}^H \mathbf{x} + \mathbf{b}^H \mathbf{x}^*. \tag{3.17}$$

It is now interesting to study the behavior of the strictly LMSE when (3.17) holds, or in the case of singular WLMSE. Note first that if $\mathbf{b} = \mathbf{0}$, (3.17) becomes strictly linear. In this case singular estimation appears equally well with the two forms of linear procedures.

Let us now investigate the completely opposite situation. It corresponds to the case where the strictly linear procedure provides a zero estimation. This means that the mean square error ε_L^2 is equal to $E[\,|y|^2]$. This situation appears when $\mathbf{r}$ defined by (2.8) is zero, which means that the estimandum y and the observation $\mathbf{x}$ are uncorrelated. By replacing y given by (3.17) in (2.8) we obtain

$$\mathbf{r} = \Gamma\mathbf{a} + \mathbb{C}\mathbf{b}. \tag{3.18}$$

As Γ is positive definite, the condition $\mathbf{r} = \mathbf{0}$ is equivalent to

$$\mathbf{a} = -\Gamma^{-1}\mathbb{C}\,\mathbf{b}. \tag{3.19}$$

Then if $\mathbb{C} \neq \mathbb{0}$, or if $\mathbf{x}$ is not circular, it is possible to associate to any non-zero vector $\mathbf{b}$ another non-zero vector $\mathbf{a}$ given by (3.19) and such that y given by (3.17) is uncorrelated with $\mathbf{x}$. This implies a zero LMSE. On the other hand, because of (3.17), $\hat{y}$ given by (1.1) is equal to y and the mean square error is zero, which means singular WLMSE. We then have zero estimation with the strictly linear procedure and perfect estimation with the wide sense linear procedure.

IV. AUTOREGRESSIVE SIGNALS AND PREDICTION

A complex autoregressive (AR) signal is defined by the difference equation

$$z[k] = \mathbf{a}^H \mathbf{Z}[k] + w[k], \tag{4.1}$$

where **a** is the (generally complex) regression vector, $\mathbf{Z}[k]$ the past vector with components $z[k-i]$, $1 \le i \le p$, and $w[k]$ a complex white noise. The second order properties of this signal are specified by its correlation and relation functions defined by

$$\gamma_z[n] = E\{z[k]z^*[k-n]\} \qquad (4.2)$$

$$c_z[n] = E\{z[k]z[k-n]\}. \qquad (4.3)$$

The whiteness assumption means that the spectrum is flat, which gives $\gamma_w[n] = \sigma_w^2\,\delta[n]$. On the other hand no specific property is assigned to $c_w[n]$ by the whiteness. However we shall say, by extension, that $w[k]$ is doubly white if $c_w[n] = c_w\,\delta[n]$, where c_w is a complex parameter. Finally $w[k]$ is circular if $c_w[n] = 0$. As $z[k]$ is obtained from $w[k]$ by a linear filter defined by (4.1), it is easy, a least in principle, to calculate its correlation and relation functions in terms of **a**, σ_w^2 and $c[n]$.

However the most interesting problem is the identification problem, or the determination of **a**, σ_w^2 and $c_w[n]$ in terms of $\gamma_z[n]$ and $c_z[n]$. For classical AR signals, this is realized by using the normal equations.

4.1. Extended Normal Equations

It results from the causality of the filter that $\mathbf{Z}[k]$ and $w[k]$ are uncorrelated. As a consequence we deduce from (4.1) that

$$\Gamma\mathbf{a} = \mathbf{c} \;\; ; \;\; \sigma_w^2 = \sigma_z^2 - \mathbf{a}^H\Gamma\mathbf{a} \qquad (4.4)$$

where Γ is the covariance matrix of $\mathbf{Z}_k$ defined as in (2.7) and **c** the vector $E\{z^*[k]\mathbf{Z}[k]\}$. This means that the regression vector **a** and the variance σ_w^2 are determined by standard equations and without taking into consideration the relation function $c_z[k]$. In other words it is not necessary at all to assume that $c_w[n] = 0$ to calculate **a** and σ_w^2 from the standard normal equations.

In order to calculate $c_w[n]$, we start from (4.1) which gives $w[k]$ in terms of $x[k]$ and its past. Introducing the quantities

$$\mathbf{c}_n \triangleq E\{z[k]\mathbf{Z}[k-n]\} \; ; \mathbb{C}_n \triangleq E\{\mathbf{Z}[k]\mathbf{Z}^T[k-n]\} \qquad (4.5)$$

we deduce

$$c_w[n] = c_z[n] - \mathbf{a}^H(\mathbf{c}_n + \mathbf{c}_{-n}) + \mathbf{a}^H\mathbb{C}_n\mathbf{a}^* \qquad (4.6)$$

This shows that the relation function of $w[k]$ can be deduced from that of $z[k]$. In the case of a doubly white noise, we obtains simply

$$c_w = c_z[0] - \mathbf{a}^H\mathbb{C}\mathbf{a}^* \qquad (4.7)$$

where $\mathbb{C}$ is the symmetric matrix $\mathbb{C}_0$.

4.2. Prediction

In many problems it is assumed that $w[k]$ is circular, or that $c_w[n] = 0$ (see [4], p. 55). In this case $z[k]$ and $\mathbf{Z}[k]$ are also jointly circular and we can apply the results of Section 3.1. As a consequence the best MS prediction of $z[k]$ in terms of all its past is $\mathbf{a}^H\mathbf{Z}[k]$. This is the classical property of AR signals, valid in the real case and also in the complex circular case.

The converse is not true, and linear prediction does not imply circularity. In fact the result remains valid if, for example, $w[k]$ is no longer circular but doubly white. This results from the fact that the innovation $\tilde{z}[k] = z[k] - \mathbf{a}^H\mathbf{Z}[k]$ is orthogonal to $Z[k]$ and to $Z^*[k]$. We have then an example of non circular signal for which strictly LMSE is equivalent to WLSME. This means that the vector appearing in the quadratic form (2.13) is zero, which can easily be verified for an AR signal obtained from a doubly white signal.

If now the driving noise $w[k]$ is no longer doubly white, the best prediction of $z[k]$ in terms of its finite past is no longer $\mathbf{a}^H\mathbf{Z}[k]$ but takes the form (1.1), and is then widely linear. Some examples of such a situation will be analyzed elsewhere.

V. Conclusion

The results of this paper can be summarized as follows. The estimation of a complex quantity in terms of a complex observation by using only second order statistics requires in general the use of WL systems. The structure of such systems has been calculated and also the gain in performance with respect to linear systems. Some examples of applications have been presented and it especially appears that, when the circularity assumption is introduced, WLMSE becomes linear. Finally all these results have been applied to some prediction problems.

References

[1] L.L. Scharf, *Statistical Signal Processing*, Addison-Wesley, 1990.

[2] B. Picinbono, *Random Signals and Systems*. Englewood Cliffs, N.J: Prentice-Hall, 1993.

[3] B. Picinbono, "Estimation with complex data", submitted to *IEEE Trans. Sign. Proc.*

[4] S. Kay, *Modern Spectral Analysis, Theory and Applications*. Englewood Cliffs, N.J: Prentice-Hall, 1988.

[5] W.M. Brown. and R. B. Crawe, "Conjugate linear filtering," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 462-465, 1969.