MINIMUM MSE ESTIMATION WITH CONVEX CONSTRAINTS

Tomer Michaeli and Yonina C. Eldar

Department of Electrical Engineering
Technion–Israel Institute of Technology, Haifa, Israel
tomermic@tx.technion.ac.il , yonina@ee.technion.ac.il

ABSTRACT

We address the problem of minimum mean-squared error (MMSE) estimation under convex constraints. The familiar orthogonality principle, developed for linear constraints, is generalized to include convex restrictions. Using the extended principle, we study two types of convex constraints: constraints on the estimated vector (e.g. bounded norm), and constraints on the structure of the estimator (e.g. filter with bounded coefficients). It is shown that in both cases there exists a simple closed form expression for the constrained MMSE estimator. As an application of our approach, we develop Wiener type filters under certain restrictions, which allow for practical implementations.

Index Terms— Constrained estimation, Constrained Wiener filtering

1. INTRODUCTION

A common problem in Bayesian estimation is to obtain an estimate of a random vector (r.v.) $x \in \mathcal{X}$ based on a realization of another r.v. $y \in \mathcal{Y}$ such that some error criterion is minimized [3]. The estimator $\phi(\cdot)$ assigns an estimated vector $\widehat{x} \in \mathcal{X}$ to every possible realization of y. Thus, the Bayesian estimation problem is essentially a problem of constructing a mapping from the space of measurement vectors $\mathcal{Y}$ to the space of signals $\mathcal{X}$ based on the joint probability function of x and y. One of the most commonly used error criteria is the mean-squared error (MSE), which is given by the expectation of the l_2-norm of the error $E[\|x - \phi(y)\|_2^2]$. It is well known that the estimator minimizing the MSE is $\phi_0(y) = E[x|y]$, the conditional expectation of x given y. This estimator, although seemingly simple, is not frequently used due to two main reasons. First, in many cases it is very hard to obtain an expression for ϕ_0, and second, one often desires to constrain the estimator to belong to a certain class of mappings because of implementation reasons. For example, when x and y are two discrete-time random processes, it may be desirable to constrain the estimator to be a causal filter with bounded coefficients rather than a general non-linear, non-causal function of the series $\{y[n]\}$. The need to pose restrictions on the estimator arises also when there is poor knowledge of the distribution of x and y, which may lead to an unfeasible estimate. For example,when estimating an image, one may want to restrict the pixel values to be positive.

Solutions to various constrained estimation problems appear in the signal processing literature, starting from the famous work of Wiener [4] on linear minimum MSE (MMSE) estimation and prediction of signals, and ranging to finite impulse response (FIR) filtering [5], finite horizon and more. All these problems possess a common structure - the constraint on the estimator is linear. Linear constraints are well treated using the orthogonality principle, or, more generally, via the concept of conditional expectation in the wide sense [3]. However, there seems to be no unifying approach to solving MMSE estimation problems under general convex constraints.

In this paper we extend the well known orthogonality principle to the case of convex constraints on the estimator. We then study two types of convex restrictions: constraints on the estimated vector $\widehat{x}$, and constraints on the structure of the estimator $\phi(\cdot)$. We show that in the first case, the solution is simply the projection of the MMSE estimate onto the set of constraints. In the later case, the estimator can be obtained from the MMSE estimator using a weighted projection. To demonstrate the approach, we show how certain restrictions can be imposed on the Wiener filter in order to allow for practical implementations. Specifically, we derive the MMSE filter under bounded norm and bounded coefficient constraints.

This work was supported in part by the Israel Science Foundation under Grant no. 536/04 and by the EU 6th framework programme, via the NEWCOM network of excellence.

2. NOTATION AND MATHEMATICAL PRELIMINARIES

Calligraphic letters are used to denote vector spaces, subspaces and sets of vectors. The Moore-Penrose pseudo-inverse of a matrix A is denoted by $A^\dagger$ and the Hermitian conjugate is A^*. The j'th component of a vector v is denoted $(v)_j$ and the (i,j) entry in the matrix A is denoted $(A)_{i,j}$. A positive (semi) definite operator is written as $A \succ (\succeq)\, 0$. Brackets are used for discrete time signals and capital letters for Fourier transforms (e.g. $Z(\omega) = \mathcal{F}\{z[n]\}$). Expectation is denoted by $E[\cdot]$ and conditional expectation is written as $E[\cdot|\cdot]$. An inner product on a vector space is denoted by $\langle \cdot, \cdot \rangle$. The associated norm is defined by $\|x\| = \sqrt{\langle x, x \rangle}$. A G-weighted inner product, where $G \succ 0$, is defined as $\langle x, Gy \rangle$.

A projection operator onto a closed and convex set $\mathcal{A}$ in a Hilbert space $\mathcal{H}$ is denoted by $P_{\mathcal{A}}(\cdot)$ and defined as

$$P_{\mathcal{A}}(h) = \arg\min_{a \in \mathcal{A}} \|h - a\| . \tag{1}$$

It can be shown that $P_{\mathcal{A}}(h) \in \mathcal{A}$ is the projection of h onto $\mathcal{A}$ if and only if [1]

$$\mathrm{Re}\{\langle h - P_{\mathcal{A}}(h), P_{\mathcal{A}}(h) - a \rangle\} \geq 0, \quad \forall a \in \mathcal{A}. \tag{2}$$

Since $P_{\mathcal{A}}(\cdot)$ depends on the inner product on $\mathcal{H}$, we use the notation $P_{\mathcal{A}}^{(G)}(\cdot)$ to denote a projection with respect to a G-weighted inner product, which we refer to as the G-weighted projection onto $\mathcal{A}$. We note that $P_{\mathcal{A}}^{(G)}(h)$ is well defined even when $G \succeq 0$, as long as the vector h is orthogonal to the null-space of G.

Projections onto subspaces can be expressed in terms of frames. A set of vectors $\{h_n\}$ in a Hilbert space $\mathcal{H}$ is called a frame for a subspace $\mathcal{A}$ if there exist constants $0 < A \le B < \infty$ such that

$$A\|a\|^2 \le \sum_n |\langle a, h_n\rangle|^2 \le B\|a\|^2, \quad \forall a \in \mathcal{A}. \tag{3}$$

The projection of $h \in \mathcal{H}$ onto $\mathcal{A}$ can be written as a linear combination of the frame vectors $P_{\mathcal{A}}(h) = \sum_k (a)_k h_k$, where the (possibly infinite) vector of coefficients a can be chosen as [2]:

$$a = G^{\dagger} v \tag{4}$$

with matrix G and vector v defined by

$$(G)_{j,k} = \langle h_k, h_j\rangle, \quad (v)_j = \langle h, h_j\rangle. \tag{5}$$

In our Bayesian estimation setup, the space of measurement vectors is denoted by $\mathcal{Y}$ and the space of signals is denoted by $\mathcal{X}$ (e.g. $\mathbb{R}^m$ and $\mathbb{R}^n$). Throughout the paper, it is assumed that $\mathcal{X}$ is a Hilbert space with inner product $\langle\cdot,\cdot\rangle_{\mathcal{X}}$ and associated norm $\|\cdot\|_{\mathcal{X}}$. The MSE of an estimator $\phi(y)$ is defined as $E\left[\|x - \phi(y)\|_{\mathcal{X}}^2\right]$.

3. EXTENDED ORTHOGONALITY PRINCIPLE

The theory of MMSE estimation under linear constraints is well established via the concept of *conditional expectation in the wide sense* [3]. In this section we give a brief overview of the basic ideas underlying this theory and then extend it to general convex constraints.

Let (Ω, F, P) be a probability space of r.v.'s taking values in $\mathcal{X}$. The set of all finite variance r.v.'s is denoted by $\mathcal{L}^2 = \left\{x : E\left[\|x\|_{\mathcal{X}}^2\right] < \infty\right\}$. An inner product on $\mathcal{L}^2$ is defined by $\langle u, v\rangle_{\mathcal{L}^2} = E\left[\langle u, v\rangle_{\mathcal{X}}\right]$. It can be shown that $\mathcal{L}^2$ is a Hilbert space given that two r.v.'s u,v are considered identical if $u = v$ with probability 1 (w.p.1).

Let $\mathcal{H}_{\mathcal{Y}}$ be the subspace of $\mathcal{L}^2$ generated by applying all (Borel measurable) operators from $\mathcal{Y}$ to $\mathcal{X}$ on the r.v. y:

$$\mathcal{H}_{\mathcal{Y}} = \left\{\phi(y) \mid \phi\colon \mathcal{Y} \to \mathcal{X}, E\left[\|\phi(y)\|_{\mathcal{X}}^2\right] < \infty\right\}. \tag{6}$$

The set $\mathcal{H}_{\mathcal{Y}}$ is the set of r.v.'s in $\mathcal{L}^2$ that can constitute an estimate for x based on y. The conditional expectation of x given y is the projection of x onto the subspace $\mathcal{H}_{\mathcal{Y}}$ and is denoted by $\widehat{x}_0 = E[x|y]$. Clearly, the r.v. $\widehat{x}_0$ is the MMSE estimate of x given y since it minimizes the distance $\|x - \widehat{x}_0\|_{\mathcal{L}^2}^2$, which is equivalent to minimizing the MSE, $E\left[\|x - \widehat{x}_0\|_{\mathcal{X}}^2\right]$.

When one restricts the search for the MMSE estimator to a certain family of operators, the set of candidate r.v.'s is narrowed down to a subset $\mathcal{A} \subseteq \mathcal{H}_{\mathcal{Y}}$. In the special case where $\mathcal{A}$ is a closed *subspace* of $\mathcal{H}_{\mathcal{Y}}$, the MMSE estimate $\widehat{x}_{\mathcal{A}}$ among all r.v.'s in $\mathcal{A}$, is the orthogonal projection of x onto $\mathcal{A}$ and is termed the *conditional expectation in the wide sense* of x given $\mathcal{A}$. Restrictions of this type are referred to herein as linear constraints.

The most famous example of a linear constraint is the restriction that $\phi(\cdot)$ be a linear operator. In this case $\mathcal{A}$ is the subspace of r.v.'s formed by applying all *linear* transformations to the measurement vector y and the conditional expectation in the wide sense of x given $\mathcal{A}$ is the familiar linear MMSE (LMMSE) estimator.

From the properties of projections in Hilbert spaces one immediately obtains the following characterization of the MMSE estimator under a linear constraint, known as the orthogonality principle. Suppose that $\mathcal{A}$ is a closed subspace of $\mathcal{H}_{\mathcal{Y}}$. Then $\phi_{\mathcal{A}}(y)$ is the MMSE estimator of x among all r.v.'s in $\mathcal{A}$ if and only if

$$E\left[\langle x - \phi_{\mathcal{A}}(y), \phi(y)\rangle_{\mathcal{X}}\right] = 0, \quad \forall \phi(y) \in \mathcal{A}. \tag{7}$$

Thus, the error $x - \phi_{\mathcal{A}}(y)$ using the optimal estimator in $\mathcal{A}$ has to be orthogonal to any other estimator in $\mathcal{A}$. For example, if $\mathcal{A}$ is the subspace in $\mathcal{H}_{\mathcal{Y}}$ of all the linear operators from $\mathcal{Y}$ to $\mathcal{X}$, then the error of $\phi_{\mathcal{A}}(y)$, which is the LMMSE estimator, is orthogonal to *every linear transformation* of the measurement vector y.

The orthogonality principle can be employed to solve a variety of constrained estimation problems in which the constraint is linear. For example, in the Wiener filtering setting, restricting the filter to be causal, FIR of order N, or any other restriction on the support of the filter are all linear constraints as they all correspond to subspaces of $\mathcal{H}_{\mathcal{Y}}$. However, an amplitude constraint such as $|h[n]| \le \varepsilon$ is nonlinear and cannot be treated by the same procedure

In the following theorem, we extend the orthogonality principle to nonlinear constraints that are closed convex sets in $\mathcal{H}_{\mathcal{Y}}$.

Theorem 1 *Let $\mathcal{A}$ be a closed convex set in the space of r.v.'s $\mathcal{H}_{\mathcal{Y}}$. Then $\phi_{\mathcal{A}}(y) \in \mathcal{A}$ is the MMSE estimator in $\mathcal{A}$ if and only if*

$$\operatorname{Re}\left\{E\left[\langle x - \phi_{\mathcal{A}}(y), \phi_{\mathcal{A}}(y) - \phi(y)\rangle_{\mathcal{X}}\right]\right\} \ge 0, \quad \forall \phi(y) \in \mathcal{A}. \tag{8}$$

Proof. Follows immediately from (2). ■

Note that in contrast to the orthogonality principle, condition (8) is an inequality and therefore does not lead to an equation, whose solution is the MMSE estimator in $\mathcal{A}$. Nevertheless, there are cases in which (8) leads to a simple scheme for obtaining the constrained MMSE estimator, as discussed in the following sections.

4. CONSTRAINTS ON THE ESTIMATED VECTOR

Consider the problem of MMSE estimation under a constraint on the estimated vector $\widehat{x}$. Specifically, we are interested in the case where $\widehat{x}$ is constrained to lie in a given closed convex set $\mathcal{W} \subseteq \mathcal{X}$. This type of constraints includes, for example, restricting the norm of the vector $\widehat{x}$ to be bounded by a given value ε or imposing that $(\widehat{x})_i \ge 0$. From the viewpoint of the estimator $\phi(\cdot)$, we confine ourselves to operators whose image is contained in $\mathcal{W}$, i.e. we consider only r.v.'s in the set $\mathcal{A} = \left\{\phi(y) \mid \phi\colon \mathcal{Y} \to \mathcal{W}, E\left[\|\phi(y)\|_{\mathcal{X}}^2\right] < \infty\right\}$. Clearly, $\mathcal{A}$ is a closed convex set in $\mathcal{H}_{\mathcal{Y}}$. Therefore, using (8) we obtain the following.

Theorem 2 *Let $\mathcal{W}$ be a closed convex set in the space of signals $\mathcal{X}$, and let $\mathcal{A}$ be the set of r.v.'s in $\mathcal{H}_{\mathcal{Y}}$ that take values only in $\mathcal{W}$. Then the MMSE estimator in $\mathcal{A}$ is*

$$\phi_{\mathcal{A}}(y) = P_{\mathcal{W}}(E[x|y]). \tag{9}$$

Proof. Plugging (9) in (8) and using the smoothing property, we get

$$\begin{aligned}
&\operatorname{Re}\left\{E\left[\langle x - P_{\mathcal{W}}(E[x|y]), P_{\mathcal{W}}(E[x|y]) - \phi(y)\rangle_{\mathcal{X}}\right]\right\} \\
&= \operatorname{Re}\left\{E\left[E\left[\langle x - P_{\mathcal{W}}(E[x|y]), P_{\mathcal{W}}(E[x|y]) - \phi(y)\rangle_{\mathcal{X}}\right]|y\right]\right\} \\
&= \operatorname{Re}\left\{E\left[\langle E[x|y] - P_{\mathcal{W}}(E[x|y]), P_{\mathcal{W}}(E[x|y]) - \phi(y)\rangle_{\mathcal{X}}\right]\right\}.
\end{aligned}$$

Now, since $\phi(y) \in \mathcal{W}$ for every y, we conclude from (2) that the real part of the inner product is non-negative for every y. After taking the expectation over y, the real-part remains positive and hence (8) is satisfied. ■

Theorem 2 states that the estimation scheme consists of projecting the optimal estimate $\widehat{x}_0 = E[x|y]$ onto the set $\mathcal{W}$. This intuitive result can be used to obtain the optimal estimate of x with bounded energy, bounded components, or in a subspace of $\mathcal{X}$. For example, let $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$. Then using (9), the MMSE estimate of x with l_2-norm bounded by a given value ε, is simply

$$\widehat{x} = \begin{cases} \widehat{x}_0 & \|\widehat{x}_0\|_2 \leq \varepsilon \\ \frac{\varepsilon \widehat{x}_0}{\|\widehat{x}_0\|_2} & \|\widehat{x}_0\|_2 > \varepsilon. \end{cases} \tag{10}$$

Similarly, the optimal estimate whose components satisfy $\left|(\widehat{x})_i\right| \leq \varepsilon_i$, for some given set of non-negative scalars $\{\varepsilon_i\}$, is $\widehat{x} = f_\varepsilon(\widehat{x}_0)$, where $f_\varepsilon(\cdot)$ is the component-wise clipping function given by

$$(f_\varepsilon(x))_i = \begin{cases} (x)_i & \left|(x)_i\right| \leq \varepsilon_i \\ \varepsilon_i & (x)_i > \varepsilon_i \\ -\varepsilon_i & (x)_i < -\varepsilon_i. \end{cases} \tag{11}$$

Finally, the optimal estimate in a subspace $\mathcal{W} \subseteq \mathbb{R}^m$ is

$$\widehat{x} = P_{\mathcal{W}} \widehat{x}_0 \tag{12}$$

where $P_{\mathcal{W}}$ is the $m \times m$ projection matrix onto $\mathcal{W}$.

5. STRUCTURAL CONSTRAINTS ON THE ESTIMATOR

We now consider the case in which the *structure* of the estimator is constrained to be of a certain form. Specifically, we design an estimator possessing the form of a linear combination of operators with certain convex constraints on the coefficients of the combination.

Suppose that one desires an estimator of the type

$$\phi(y) = \sum_k (a)_k \phi_k(y) \tag{13}$$

where $\{\phi_k\}$ are a set of given operators from $\mathcal{Y}$ to $\mathcal{X}$. If the r.v.'s $\{\phi_k(y)\}$ form a frame for a subspace $\mathcal{A} \subseteq \mathcal{H}_\mathcal{Y}$, and no further restriction is posed on the coefficient vector a, then the MMSE estimator in $\mathcal{A}$ is obtained by using the vector (4):

$$a_0 = G^\dagger v \tag{14}$$

where the matrix G and vector v in (5) are given, in our case, by

$$(G)_{j,k} = E\left[\left\langle \phi_k(y), \phi_j(y) \right\rangle_\mathcal{X}\right], \quad (v)_j = E\left[\left\langle x, \phi_j(y) \right\rangle_\mathcal{X}\right]. \tag{15}$$

This is the familiar LMMSE estimator of x given the r.v.'s $\{\phi_k(y)\}$.

In many cases, the matrix G is close to singular, causing the coefficients of the linear combination to attain very large values. Such an effect is usually undesired in practical implementations. Hence, we are interested in the case where the vector a is constrained to lie in a closed convex set $\mathcal{W} \subseteq l_2$. This, for example, includes seeking a representation with bounded coefficients. Interestingly, the solution to this constrained estimation problem, as in (9), amounts to applying a certain transformation to the unconstrained coefficient vector a_0 in (14), as described in the following theorem.

Theorem 3 *Let $\{\phi_k(y)\}$ be a set of r.v.'s that form a frame for a subspace in $\mathcal{H}_\mathcal{Y}$, and let $\mathcal{W}$ be a closed convex set in l_2. Then among all the estimators of the form $\phi(y) = \sum_k (a)_k \phi_k(y)$ with $a \in \mathcal{W}$, the coefficient vector that minimizes the MSE is $a_\mathcal{W} = P_\mathcal{W}^{(G)}(a_0)$, with G and a_0 given in (15) and (14) respectively.*

The proof of the theorem relies on condition (8) and is omitted due to lack of space.

Note that in contrast to the case of constraints on the estimated vector (9), here the relation between the unconstrained and constrained coefficients is not a simple projection but rather a G-weighted projection. The methods coincide only if $\{\phi_k(y)\}$ are uncorrelated r.v.'s with equal variance, in which case $G = \sigma^2 I$.

We now present some examples of Theorem 3.

5.1. A Quadratic constraint

Consider the family of estimators of the form (13) where the coefficients are constrained to have bounded weighted norm:

$$a^* W a \leq \varepsilon \tag{16}$$

where $\varepsilon > 0$ and W is a positive definite matrix such that W^{-1} is bounded. From Theorem 3, the optimal set of coefficients under this constraint is

$$a = (G + \lambda W)^\dagger G a_0 \tag{17}$$

where $\lambda \geq 0$ is the minimal value for which (16) holds.

We see that imposing a quadratic constraint amounts to adding a regularization term to the matrix G in (14) (as Ga_0 can be replaced by v). An example of the use of (17) is given in Section 6 in the context of Wiener filtering.

5.2. An l_∞ Constraint

Suppose that the matrix G is not singular and that it can be factored as $G = UU^*$. We wish to obtain an estimator of the form (13) with a coefficient vector a, whose coordinates along the columns $\{U_i\}$ of U are bounded by a given set of non-negative scalars $\{\varepsilon_i\}$:

$$\left|\langle U_i, a\rangle\right| \leq \varepsilon_i \tag{18}$$

From Theorem 3, the optimal coefficient vector is

$$a = (U^*)^{-1} f_\varepsilon(U^* a_0) \tag{19}$$

where $f_\varepsilon(\cdot)$ is the component-wise clipping function (11).

It can be seen that in contrast to (17), this constraint causes the coefficient vector a to be a non-linear function of a_0. In Section 6 we use (19) to impose certain restrictions on the coefficients of the Wiener filter.

6. CONSTRAINED WIENER FILTER

One of the most widely used applications of MSE estimation in the field of signal processing is the Wiener filter. The discrete version of the optimal filtering problem is the following. Given a wide sense stationary (WSS) signal $y[n]$, produce an estimate $\widehat{x}[n]$ of the signal $x[n]$ such that the MSE $E[|x[n] - \widehat{x}[n]|^2]$ is minimized. The optimal *linear* estimator happens to be also time invariant and is known as the Wiener filter [4]. In this section we point out some possible drawbacks of the Wiener solution, and derive an expression for the MMSE filter under certain constraints which allow for practical implementations.

In order to derive an expression for a constrained Wiener filter, we first develop the unconstrained Wiener solution within our framework. Our problem is to design a filter $h_{m_0}[n]$ that minimizes the MSE at a certain time instance m_0, thus

$$\widehat{x}[m_0] = \sum_{k=-\infty}^{\infty} h_{m_0}[k]\, y[m_0 - k]. \tag{20}$$

Comparing (20) with (13) we see that they are of the same form with the random variables $\{y[m_0 - k]\}$ playing the role of the random vectors $\{\phi_k(y)\}$. Using (3) it can be shown that $\{y[m_0 - k]\}$ form a frame for a subspace in $\mathcal{H}_\mathcal{Y}$ if and only if there exist constants $0 < A \leq B < \infty$ such that the spectrum of $y[n]$ satisfies

$$A \leq S_{yy}(\omega) \leq B, \quad \omega \in \Omega_u \tag{21}$$

where $\Omega_u \triangleq \{\omega | S_{yy}(\omega) \neq 0\}$.

To find $h_{m_0}[n]$, the matrix form in (14) is used, with v and G given by (15):

$$(v)_j = E[x[m_0] y[m_0 - j]] = R_{xy}[j], \quad (22)$$

$$(G)_{j,k} = E[y[m_0 - k] y[m_0 - j]] = R_{yy}[j-k]. \quad (23)$$

Evidently, both v and G are independent of m_0. The optimal filter is thus also independent of m_0 and is given from (4) by $h = G^\dagger v$, where the infinite vector h corresponds to the series $\{h[n]\}$.

The matrix G is an infinite Toeplitz matrix that corresponds to convolution with $R_{yy}[n]$ and v is an infinite vector corresponding to the series $R_{xy}[n]$. Hence, it is easily verified that the frequency response of the optimal filter $H(\omega)$ can be chosen as

$$H(\omega) = \begin{cases} \frac{S_{xy}(\omega)}{S_{yy}(\omega)} & \omega \in \Omega_u \\ 0 & \omega \notin \Omega_u. \end{cases} \quad (24)$$

If the lower bound in (21) is very small, then there may be frequencies at which $S_{yy}(\omega)$ is close to zero. This typically causes the impulse response of the filter to attain large values and have a slow decay, properties which may be undesirable in practical implementations. To overcome these difficulties, we may impose certain constraints on the filter that suppress the magnitude of $H(\omega)$ or, alternatively, bound the coefficients of the filter $h[n]$ in the time domain. Two such constraints are considered in the following sections.

6.1. Wiener filter with a quadratic constraint

We first restrict the norm of the filter by considering only those filters for which

$$\int_{-\pi}^{\pi} |H(\omega)|^2 L(\omega) \, d\omega \leq \varepsilon \quad (25)$$

for some $\varepsilon > 0$ and weighting $L(\omega)$ that satisfies $L(\omega) \geq C$ for some $C > 0$. Using Parseval's theorem, (25) can be written as $h^* W h \leq \varepsilon / 2\pi$, where W is the infinite Toeplitz matrix defined by $(W)_{m,n} \triangleq l[m-n]$. As can be seen, this is a convex constraint of the type (16). Thus, the solution of the optimal filtering problem under constraint (25) is given by (17):

$$h = (G + \lambda W)^\dagger v. \quad (26)$$

The matrices G and W are both infinite Toeplitz matrices that correspond to convolution with $R_{yy}[n]$ and $l[n]$ respectively. Therefore, the quadratically-constrained Wiener filter $H(\omega)$ is given by

$$H(\omega) = \begin{cases} \frac{S_{xy}(\omega)}{S_{yy}(\omega) + \lambda L(\omega)} & \omega \in \Omega_c \\ 0 & \omega \notin \Omega_c \end{cases} \quad (27)$$

where $\lambda \geq 0$ is the minimal value for which (25) holds and $\Omega_c \triangleq \{\omega | S_{yy}(\omega) + \lambda L(\omega) \neq 0\}$. Note that now, even if the lower bound of $S_{yy}(\omega)$ is small, we can still ensure that $H(\omega)$ does not explode by choosing a weighting function $L(\omega)$ that is bounded below by a larger positive number.

The quadratically-constrained Wiener filter does not have an explicit solution in the sense that there is no closed form for λ. However, in practical applications the exact value of ε in (25) is usually not very important as long as the filter is realizable. Hence the exact value of λ is not crucial but rather the form of the frequency response is what matters. In cases where the value of ε is important, one can use the bisection algorithm to obtain the optimal λ.

An interesting fact is that the quadratically-constrained Wiener filter is identical to the unconstrained Wiener solution designed to estimate $x[n]$ from the measurements $\tilde{y}[n] = y[n] + z[n]$, where $z[n]$ is a WSS process with spectrum $S_{zz}(\omega) = \lambda L(\omega)$. Hence, the effect of constraining the filter to belong to the class defined in (25), can be understood as designing a filter to estimate $x[n]$ from a noisy version of the measurements instead of using the measurements themselves.

6.2. Wiener filter with an l_∞ constraint

A more intuitive approach for specifying a restriction on the Wiener filter is to do so in the time domain. Specifically, we seek the MMSE filter subject to the constraint:

$$\left| \mathcal{F}^{-1} \{H(\omega) L^*(\omega)\}[n] \right| \leq \varepsilon[n], \quad n \in \mathbb{Z} \quad (28)$$

where $\{\varepsilon[n]\}_{n \in \mathbb{Z}}$ is a series of non-negative numbers. It turns out that this problem has a simple solution given that $S_{yy}(\omega) > 0$ and $L(\omega)$ is a factorization of $S_{yy}(\omega)$, i.e. the relation $L(\omega) L^*(\omega) = S_{yy}(\omega)$ holds. Note that we do not require spectral factorization in the sense that $l[n]$ be a causal series, but any decomposition of $S_{yy}(\omega)$ into a multiplication of two conjugate functions. For example $L(\omega) = \sqrt{S_{yy}(\omega)}$ is a valid choice. Under these restrictions, constraint (28) is of the form (18) and thus the solution is given by (19), which, in our case, reduces to

$$H(\omega) = \frac{1}{L^*(\omega)} \mathcal{F} \left\{ f_\varepsilon \left(\mathcal{F}^{-1} \left\{ \frac{S_{xy}(\omega)}{L(\omega)} \right\} [n] \right) \right\} (\omega) \quad (29)$$

where $f_\varepsilon(\cdot)$ is the component-wise clipping function (11).

As a special case of (28), we may obtain the causal Wiener filter. This is done by setting $\varepsilon[n] = 0$ for $n < 0$ and $\varepsilon[n] \to \infty$ for $n \geq 0$ and using the spectral factorization $L(\omega) = S^-(\omega)$ and $L^*(\omega) = S^+(\omega)$, where $S^+(\omega)$ is the Fourier transform of a causal series. Constraint (28), then, causes $H(\omega) S^+(\omega)$ to be causal, and since $S^+(\omega)$ is causal, this implies that $H(\omega)$ is causal as well. Substituting this specific choice of $\varepsilon[n]$ and $L(\omega)$ in (29), we get the known expression for the causal Wiener filter [4].

Another interesting special case of (28) is the condition $\left| \int_{-\pi}^{\pi} H(\omega) \sqrt{S_{yy}(\omega)} d\omega \right| \leq \varepsilon_0$. This constraint can be brought to the form of (28) by setting $\varepsilon[n] \to \infty$ for $n \neq 0$ and $\varepsilon[0] = \varepsilon_0 / 2\pi$ and using $L(\omega) = \sqrt{S_{yy}(\omega)}$. Substituting these expressions into (29), the solution is

$$H(\omega) = \frac{1}{\sqrt{S_{yy}(\omega)}} \left(\frac{S_{xy}(\omega)}{\sqrt{S_{yy}(\omega)}} - c \right) \quad (30)$$

where $c = \max \left\{ \frac{1}{2\pi} \left(\left| \int_{-\pi}^{\pi} S_{xy}(\omega) \Big/ \sqrt{S_{yy}(\omega)} \, d\omega \right| - \varepsilon_0 \right), 0 \right\}$.

7. REFERENCES

[1] J. F. Bonnans, C. Lemarechal, C. A. Sagastizabal, and J. C. Gilbert, *Numerical Optimization: Theoretical and Practical Aspects.* Springer, 2003.

[2] Y. C. Eldar and O. Christansen, "Characterization of oblique dual frame pairs," *J. Applied Signal Processing*, pp. 1–11, 2006.

[3] A. N. Shiryaev, *Probability.* Springer, 1995.

[4] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series.* MIT Press, Cambridge, Mass. and Wiley, New York, 1949.

[5] A. Zadeh and J. R. Ragazzini, "An extension of Wiener's theory of prediction," *Journal of Applied Physics*, vol. 21, pp. 654–655, 1950.