CCIT Report #641 October 2007

Constrained Linear Minimum MSE Estimation

Tomer Michaeli and Yonina C. Eldar, *Member, IEEE*

***Abstract*— We address the problem of linear minimum mean-squared error (LMMSE) estimation under constraints on the filter or the estimated signal. We develop a general formula that leads to closed form solutions for a wide class of constrained LMMSE problems. The results are applicable to both finite dimensional problems as well as to the Wiener filtering setup, in which infinitely-many measurements are available. Our approach generalizes previous known results such as the generalized Karhunen-Loeve transform (GKLT), the causal Wiener filter and more. As an application of our framework, we develop Wiener type filters under various restrictions, which allow for practical implementations.**

***Index Terms*— Estimation, Wiener filtering, Constrained Estimation.**

I. INTRODUCTION

A COMMON problem in Bayesian estimation is to obtain an estimate of a random vector (r.v.) $\boldsymbol{x}$ based on a realization of another random vector $\boldsymbol{y}$ such that some error criterion is minimized [1]. The estimator $\widehat{\boldsymbol{x}} = \phi(\boldsymbol{y})$ assigns an estimated vector $\widehat{\boldsymbol{x}}$ to every possible realization of $\boldsymbol{y}$. Thus, constructing a Bayesian estimator amounts to a mapping from the space of measurement vectors to the space of signals based on the joint probability function of $\boldsymbol{x}$ and $\boldsymbol{y}$. One of the most commonly used error criteria is the mean-squared error (MSE), which is given by the expectation of the squared-norm error $E[\|\boldsymbol{x} - \phi(\boldsymbol{y})\|^2]$. It is well known that the estimator minimizing the MSE is the conditional expectation of $\boldsymbol{x}$ given $\boldsymbol{y}$, denoted as $\phi_0(\boldsymbol{y}) = E[\boldsymbol{x}|\boldsymbol{y}]$.

The minimum MSE (MMSE) 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. Second, one often desires to constrain the estimator to belong to a certain class of mappings because of implementation reasons. One way to overcome the difficulties in computation and implementation of the MMSE method is to restrict the estimator to be linear. The linear MMSE estimator (LMMSE) minimizes the MSE among all linear functions. The LMMSE solution has a closed form that depends only on the second order statistics of $\boldsymbol{x}$ and $\boldsymbol{y}$ [2], quantities which may be easily estimated from a set of training data. Moreover, it is easy to implement in practical applications as the estimation procedure involves only matrix multiplication, i.e. $\widehat{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{y}$. The LMMSE approach also has a simple extension to the case where $\boldsymbol{x}$ and $\boldsymbol{y}$ are jointly wide-sense stationary (WSS) random processes, which is known as the Wiener filter [3].

The authors are with the department of electrical engineering, Technion–Israel Institute of Technology, Haifa, Israel (phone: +972-4-8293256, fax: +972-4-8295757, e-mail: tomermic@tx.technion.ac.il, yonina@ee.technion.ac.il).

This work was supported by the Israel Science Foundation under Grant 536/04.

In practical applications, there are situations that require that either the linear estimator itself or the signal at its output possess certain desired properties. These may stem for example from implementation limitations, efficiency of computation or the need to compress the data.

A prime example of constrained linear estimation is the famous work of Wiener [3] on causal LMMSE estimation and prediction of signals. Other restrictions on the Wiener filter include finite impulse response (FIR) [4], finite horizon and general restrictions on the support of the filter in the time domain [5]. Constrained LMMSE estimation arises in array processing applications as well (also termed vector Wiener filtering). In this context, it is often desired to reduce the dimensionality of the measurement vector process. This is analogous to restricting the rank of the estimator. Various approaches were devised in the past for this problem, some directed at minimizing the MSE and some ad hoc (see for example [6],[7] and references therein).

Reduced rank estimators are also at the heart of signal compression. A basic problem encountered in this field is the determination of a small set of vectors that allows the representation (via linear combinations) of a certain class of signals. These vectors are usually chosen to minimize the MSE between the original signal and its compact representation. Like in array processing, this approach can be expressed as a linear estimation problem with a rank constraint. The solution to this problem is known as principal component analysis (PCA) or the Karhunen-Loeve transform (KLT). One extension to this basic concept is the design of a linear compression transform that takes into account additive noise [8]. A more general setting was considered in [9] and [10] in which a compact representation is designed for the task of estimating a different signal (rather than the representation of the original signal itself).

There are applications in which the constraints on the estimator have a stochastic flavor. One example is the MMSE whitening technique which emerged recently [11] and found many applications in the fields of signal processing and communication. In this methodology, a linear transformation is designed such that, when applied to a r.v. $\boldsymbol{y}$, it produces the r.v. $\widehat{\boldsymbol{y}}$ that is as close as possible to $\boldsymbol{y}$ in an MSE sense and whose covariance matrix is diagonal. A generalization of this approach is the covariance shaping technique [12], in which the transform is designed to produce a r.v. with a predefined covariance matrix (not necessarily diagonal). A few of the applications of these approaches are improvement of least squares parameter estimation [13], multiuser detection [14] and matched filtering [15].

Another important class of restrictions emerges in setups where the estimator is only one block in a larger scheme. For example, in a relay communication system it is often desired