# A General Formulation for Least-Squares Problems

**Mark Flanagan**† **and A. D. Fagan**∗

*Department of Electronic and Electrical Engineering*
*University College Dublin*
*IRELAND*

E-mail: †mark.flanagan@ee.ucd.ie ∗tony.fagan@ucd.ie

*Abstract* — **We show in this paper that many different least-squares problems which have applications in signal processing may be seen as special cases of a more general vector space minimization problem called the Minimum Norm problem. We show that special cases of the Minimum Norm problem include: least squares fitting of a finite set of points to a linear equation and to a quadratic equation; the infinite length MMSE-optimum linear equalizer; the finite length MMSE-optimum linear equalizer; the steepest descent algorithm and the more practical LMS algorithm for iterative estimation of the finite-length MMSE-optimum linear equalizer for an unknown channel; and the finite-length least-squares-optimum linear equalizer with a forgetting factor. These examples are not exhaustive but are chosen to illustrate the scope of this framework.**

*Keywords* — **vector space theory, least squares fitting, MMSE filtering, LMS algorithm, least-squares filtering**

## I Introduction

Many different least-squares problems in communications and signal processing have a unity which is not emphasized in the current literature. In [1] the Wiener filter and least-squares filter are both derived in the same text with little reference to the unity of the two problems. Moreover, when deriving the MMSE-optimum linear filter, [1] explicitly states that the vector space viewpoint for the principle of orthogonality is merely an analogy. We show in this paper that this is not the case, and that these problems, among many others, are merely special cases of the more general Minimum Norm problem for which the mathematics is simpler. This problem is usually solved using the projection theorem (e.g. [2]); we take a different approach using the complex gradient because we are interested not only in the solution but in its estimation by steepest descent and stochastic gradient algorithms. We then show that many problems in signal processing and their solutions may be deduced from this analysis simply by substituting for the vector space and inner product. Also, certain properties which are well known for the correlation matrix of observed data are shown to hold for our "inner product matrix" $\mathbf{R}$, of which the correlation matrix is a special case. A generalization of the famous Cauchy-Schwartz inequality also arises naturally from this work.

The examples chosen here are for tutorial purposes; the reader should recognize that from this formulation follow the solutions to a general class of problems of unconstrained least-squares fitting of a finite set of points to an equation. Also as special cases follow the MMSE-optimum linear equalizer (LE), decision-feedback equalizer (DFE), interference canceller (IC), fractionally-spaced equalizer (FSE) and more general structures, together with all LMS-type algorithms which are used to obtain these equalizers in practice for an unknown channel. Also, a simpler but similar formulation is possible with real instead of complex scalars. Finally, we note that many of the equalization results derived here may be found in [3].

## II Vector Space Preliminaries

A complex vector space is a set $V$, closed under two operations called addition (so that there exists $v+w \in V$ for all $v, w \in V$) and scalar multiplication (so that there exists $\alpha v \in V$ for all $v \in V, \alpha \in \mathbf{C}$), with the following properties:

1. $u + (v + w) = (u + v) + w \quad \forall u, v, w \in V$
2. $v + w = w + v \quad \forall v, w \in V$
3. $\exists\, 0 \in V,$ such that $v + 0 = 0 + v = v, \ \forall v \in V$
4. $\forall v \in V\, \exists\, (-v) \in V,$ such that $v + (-v) = 0$
5. $(x + y)\, v = xv + yv \quad \forall x, y \in \mathbf{C}$ and $\forall v \in V$
6. $z\,(v + w) = zv + zw \quad \forall v, w \in V$ and $\forall z \in \mathbf{C}$
7. $(xy)\, v = x\,(yv) \quad \forall x, y \in \mathbf{C}$ and $\forall v \in V$
8. $1v = v \quad \forall v \in V$

An inner product on a vector space $V$ is a mapping $\langle\rangle : V \times V \to \mathbf{C}$ with the following properties:

1. $\langle x, x\rangle \geq 0 \quad \forall x \in V,$ equality iff $x = 0$
2. $\langle x, y\rangle = \langle y, x\rangle^* \quad \forall x, y \in V$
3. $\langle \alpha x_1 + \beta x_2, y\rangle = \alpha\, \langle x_1, y\rangle + \beta\, \langle x_2, y\rangle$

   $\forall x_1, x_2, y \in V$ and $\forall \alpha, \beta \in \mathbf{C}$

A norm for the vector space $V$ is a mapping $\|\,\| : V \to \mathbf{R}$ with the following properties:

1. $\|x\| \geq 0\ \forall x \in V,$ equality iff $x = 0$
2. $\|\alpha x\| = |\alpha|\, \|x\|\ \forall \alpha \in \mathbf{C}$ and $\forall x \in V$
3. $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in V$

If $V$ is a vector space with inner product $\langle\rangle$, $\|x\| = \sqrt{\langle x, x\rangle}$ is a norm for $V$ and is called the *associated norm.* For this particular norm, we also have *Parseval's Theorem*

$$\|a\|^2 + \|b\|^2 = \|c\|^2 \ \text{ if } a + b = c \text{ and } \langle a, b\rangle = 0$$

We shall assume the use of this norm throughout the paper. Finally, a set of vectors $\{x_i | i \in S\}$ in $V$ is said to be *linearly independent* if and only if $\sum_{i \in S} w_i^* x_i = 0$ only when $w_i = 0\, \forall i \in S$. Otherwise it is said to be *linearly dependent.*

## III The Minimum Norm Problem

### a) General Vector Space

Suppose we are given a vector space $V$ with inner product $\langle\rangle$, and associated norm $\|\,\|$, a vector $d \in V$ and a linearly independent set of vectors $\{x_i | i \in S\}$ in $V$, where the index set $S$ is either finite or countably infinite. The minimum norm problem is that of minimizing $J(\{w_i\}) = \|e\|^2$ where

$$e = \sum_{i \in S} w_i^* x_i - d$$

with respect to the scalars $\{w_i\}, w_i \in \mathbf{C}$. We call $e$ the *error vector*, $z = \sum_{i \in S} w_i^* x_i$ the *attempt vector* and $d$ the *desired vector.* Then

$$J(\{w_i\}) = \|e\|^2 = \left\langle \sum_{i \in S} w_i^* x_i - d, \sum_{j \in S} w_j^* x_j - d \right\rangle$$

$$= \sum_{i \in S} \sum_{j \in S} w_i^* w_j R_{ij} - 2\text{Re}\left( \sum_{i \in S} w_i^* c_i \right) + \|d\|^2$$

where $R_{ij} = \langle x_i, x_j\rangle$ and $c_i = \langle x_i, d\rangle$. If we let $w_i = a_i + jb_i$ for each $i \in S$, then $J(\{w_i\})$ is a quadratic expression in $\{a_i\} \cup \{b_i\}$ with positive coefficients $\left(\left\{\|x_i\|^2\right\}\right)$ on the squared terms. Define the complex gradient as

$$\frac{\partial J}{\partial w_i} = \frac{\partial J}{\partial a_i} + j\frac{\partial J}{\partial b_i}$$

Two properties of the complex gradient are of interest:

1. $\frac{\partial J}{\partial w_i} = 0$ if and only if $\frac{\partial J}{\partial a_i} = 0$ and $\frac{\partial J}{\partial b_i} = 0$
2. The mapping

   $$w_i^{(k+1)} = w_i^{(k)} - \frac{1}{2}\mu\frac{\partial J}{\partial w_i}$$

   is equivalent to the pair of mappings

   $$a_i^{(k+1)} = a_i^{(k)} - \frac{1}{2}\mu\frac{\partial J}{\partial a_i}$$

   $$b_i^{(k+1)} = b_i^{(k)} - \frac{1}{2}\mu\frac{\partial J}{\partial b_i}$$

Now

$$\frac{\partial}{\partial w_k}\left(\sum_{i\in S} w_i^* c_i\right) = (c_k) + j\,(-jc_k) = 2c_k$$

$$\frac{\partial}{\partial w_k}\left(\sum_{i\in S} w_i c_i^*\right) = (c_k^*) + j\,(jc_k^*) = 0$$

$$\frac{\partial}{\partial w_k}\left(\sum_{i\in S}\sum_{j\in S} w_i^* w_j R_{ij}\right)$$

$$\begin{aligned}
&= \left(\sum_{i\in S}\sum_{j\in S}\left[\delta_{ik} w_j R_{ij} + \delta_{jk} w_i^* R_{ij}\right]\right) \\
&+ \; j\left(\sum_{i\in S}\sum_{j\in S}\left[-j\delta_{ik} w_j R_{ij} + j\delta_{jk} w_i^* R_{ij}\right]\right) \\
&= \left(\sum_{j\in S} w_j R_{kj} + \sum_{i\in S} w_i^* R_{ik}\right) + \\
&+ \; j\left(\sum_{j\in S} -jw_j R_{kj} + \sum_{i\in S} jw_i^* R_{ik}\right) \\
&= 2\sum_{j\in S} w_j R_{kj}
\end{aligned}$$

Thus, for all $k \in S$,

$$\begin{aligned}
\frac{\partial J}{\partial w_k} &= 2\left(\sum_{j\in S} w_j R_{kj} - c_k\right) \\
&= 2\left(\sum_{j\in S} w_j \langle x_k, x_j\rangle - \langle x_k, d\rangle\right) \\
&= 2\left(\sum_{j\in S} w_j^* \langle x_j, x_k\rangle - \langle d, x_k\rangle\right)^* \\
&= 2\left(\langle e, x_k\rangle\right)^*
\end{aligned}$$

so the solution to the Minimum Norm problem satisfies

$$\langle e, x_k\rangle = 0 \qquad \forall k \in S \tag{1}$$

This is called the *principle of orthogonality:* at optimum, the error vector is orthogonal to the "used" vectors. A corollary to this is that the attempt vector $z$ is orthogonal to $e$ at optimum, i.e. $\langle e, z\rangle = 0$.

From Parseval's theorem it then follows that for the optimum solution

$$\begin{aligned}
J(\{w_i\}) = J_{min} &= \|d\|^2 - \|z\|^2 \qquad (2) \\
&= \|d\|^2 - \sum_{i\in S}\sum_{j\in S} w_i^* w_j R_{ij}
\end{aligned}$$

*b) Finite-dimensional Space*

In the case where $S$ is finite, we can express the problem and solution in a simpler matrix form. Suppose $S = \{1, 2, \ldots n\}$. Then

$$J(\mathbf{w}) = \mathbf{w}^H\mathbf{R}\mathbf{w} - 2\mathrm{Re}\left(\mathbf{w}^H\mathbf{c}\right) + \|d\|^2$$

where $\mathbf{R} = (R_{ij}) = (\langle x_i, x_j\rangle)$ and $\mathbf{c} = (c_i) = (\langle x_i, d\rangle)$. $\mathbf{R}$ is Hermitian from inner product property 2. Now we make the observation that

$$\begin{aligned}
\mathbf{w}^H\mathbf{R}\mathbf{w} &= \left\langle \sum_{i=1}^{n} w_i^* x_i, \sum_{j=1}^{n} w_j^* x_j \right\rangle \\
&= \|z\|^2 > 0
\end{aligned}$$

assuming not all of the $\{w_i\}$ are equal to zero. Therefore $\mathbf{R}$ is positive definite (and thus invertible). The principal minors criterion then gives

$$\det \mathbf{R} > 0$$

The invertibility of $\mathbf{R}$ allows us to write

$$\begin{aligned}
J(\mathbf{w}) &= \left(\|d\|^2 - \mathbf{c}^H\mathbf{R}^{-1}\mathbf{c}\right) \\
&+ \left(\mathbf{w} - \mathbf{R}^{-1}\mathbf{c}\right)^H \mathbf{R}\left(\mathbf{w} - \mathbf{R}^{-1}\mathbf{c}\right)
\end{aligned}$$

Thus

$$\begin{aligned}
J(\mathbf{w}) \geq J_{min} &= \left(\|d\|^2 - \mathbf{c}^H\mathbf{R}^{-1}\mathbf{c}\right) \\
&= \left(\|d\|^2 - \mathbf{w}_{opt}^H\mathbf{R}\mathbf{w}_{opt}\right)
\end{aligned}$$

with equality iff $\mathbf{w} = \mathbf{w}_{opt} = \mathbf{R}^{-1}\mathbf{c}$ and then $\|z\|^2 = \mathbf{c}^H\mathbf{R}^{-1}\mathbf{c} = \mathbf{w}_{opt}^H\mathbf{R}\mathbf{w}_{opt}$. It is easy to check that if the $\{x_i\}$ is linearly dependent, $\mathbf{R}$ has determinant zero. Thus we obtain the general result

$$\det\begin{pmatrix}
\langle x_1, x_1\rangle & \langle x_1, x_2\rangle & \cdots & \langle x_1, x_n\rangle \\
\langle x_2, x_1\rangle & \langle x_2, x_2\rangle & & \vdots \\
\vdots & & \ddots & \vdots \\
\langle x_n, x_1\rangle & \cdots & \cdots & \langle x_n, x_n\rangle
\end{pmatrix} \geq 0 \tag{3}$$

with equality iff the $\{x_i\}$ is a linearly dependent set. For $n = 2$ this inequality reduces to

$$\|x_1\|^2 \|x_2\|^2 \geq |\langle x_1, x_2 \rangle|^2$$

with equality iff the vectors $\{x_1, x_2\}$ are linearly dependent: this is the famous *Cauchy-Schwartz inequality.* Equation (3) for $n > 2$ gives a generalization of the Cauchy-Schwartz inequality.

## IV Special Cases of the Minimum Norm Problem

### a) *Least Squares Fitting of Points to a Linear Equation (complex linear regression)*

Suppose we have $n$ points $(x_i, y_i)$, $i = 1, 2, \ldots n$. Consider the vector space of all complex sequences of length $n$; this has inner product $\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i^*$, where $x = \{x_1, x_2, \ldots x_n\}$. Let

$$e = m^* x + c^* 1 - y$$

where 1 denotes the all-ones complex sequence of length $n$. Here $\mathbf{w} = \begin{pmatrix} m & c \end{pmatrix}^T$ and we wish to minimize

$$J(\mathbf{w}) = \|e\|^2 = \sum_{i=1}^{n} |m^* x_i + c^* - y_i|^2$$

The solution is given by $\mathbf{w} = \mathbf{R}^{-1}\mathbf{c}$, i.e.

$$\begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} \langle x, x \rangle & \langle x, 1 \rangle \\ \langle 1, x \rangle & \langle 1, 1 \rangle \end{pmatrix}^{-1} \begin{pmatrix} \langle x, y \rangle \\ \langle 1, y \rangle \end{pmatrix}$$

i.e.

$$\begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} \sum |x_i|^2 & \sum x_i \\ \sum x_i^* & n \end{pmatrix}^{-1} \begin{pmatrix} \sum x_i y_i^* \\ \sum y_i^* \end{pmatrix}$$

Inverting this matrix gives the well-known expressions for complex linear regression.

### b) *Least Squares Fitting of Points to a Quadratic Equation*

Again suppose we have $n$ points $(x_i, y_i)$, $i = 1, 2, \ldots n$. We again use the vector space of all complex sequences of length $n$, with inner product as before. Imagine this time that we wish to fit our set of points to a second-order equation in a least-squares manner, i.e.

$$e = p^* x^2 + q^* xy + r^* y^2 + s^* x + t^* y - 1$$

here $xy$ is to be interpreted as $\{x_1 y_1, x_2 y_2, \ldots x_n y_n\}$ for any $x, y \in V$ and 1 denotes the all-ones complex sequence as before. Here $\mathbf{w} = \begin{pmatrix} p & q & r & s & t \end{pmatrix}^T$ and $J(\mathbf{w}) = \|e\|^2 = \sum_{i=1}^{n} \left| p^* x_i^2 + q^* x_i y_i + r^* y_i^2 + s^* x_i + t^* y_i - 1 \right|^2$. The solution is then $\mathbf{w} = \mathbf{R}^{-1}\mathbf{c}$, where

$$\mathbf{R} = \begin{pmatrix} \langle x^2,x^2 \rangle & \langle x^2,xy \rangle & \langle x^2,y^2 \rangle & \langle x^2,x \rangle & \langle x^2,y \rangle \\ \langle xy,x^2 \rangle & \langle xy,xy \rangle & \langle xy,y^2 \rangle & \langle xy,x \rangle & \langle xy,y \rangle \\ \langle y^2,x^2 \rangle & \langle y^2,xy \rangle & \langle y^2,y^2 \rangle & \langle y^2,x \rangle & \langle y^2,y \rangle \\ \langle x,x^2 \rangle & \langle x,xy \rangle & \langle x,y^2 \rangle & \langle x,x \rangle & \langle x,y \rangle \\ \langle y,x^2 \rangle & \langle y,xy \rangle & \langle y,y^2 \rangle & \langle y,x \rangle & \langle y,y \rangle \end{pmatrix}$$

and

$$\mathbf{c} = \left( \left( \langle x^2, 1 \rangle \quad \langle xy, 1 \rangle \quad \langle y^2, 1 \rangle \quad \langle x, 1 \rangle \quad \langle y, 1 \rangle \right) \right)^T$$

This can be easily computed, e.g. $\langle y^2, xy \rangle = \sum_{i=1}^{n} y_i^2 x_i^* y_i^* = \sum_{i=1}^{n} |y_i|^2 x_i^* y_i$.

### c) *Infinite-length MMSE-optimum Linear Equalizer*

Consider the equalization problem illustrated in Figure 1. $H(f)$ is the transfer function of a channel, $\{d_k\}$ is a white sequence and $\{r_k\}$ is additive white noise. We wish to derive the frequency response $W(f)$ of the optimum linear equalizer, i.e. that which minimizes $E\left\{|e_k|^2\right\}$. Here we take as vector space the space of all random variables, which has inner product $\langle X, Y \rangle = E\{XY^*\}$ (correlation). The error vector here is the random variable $e_k = \sum_{j \in \mathbf{Z}} w_j^* u_{k-j} - d_k$. This gives $J(\{w_i\}) = \|e\|^2 = E\left\{|e_k|^2\right\}$, as required, and here the set of "used" vectors $\{x_i\}$ is the set of random variables $\{u_{k-j} | j \in \mathbf{Z}\}$. The condition for optimum is then $\langle e_k, u_{k-j} \rangle = 0$ for all $j \in \mathbf{Z}$, i.e. the cross-correlation function $R_{eu}(j) = 0$ for all $j \in \mathbf{Z}$. Taking the Fourier transform, the cross-spectral density $S_{eu}(f) = 0$ or

$$S_{zu}(f) = S_{du}(f)$$

$$\Longrightarrow W(f) \left[ |H(f)|^2 \sigma_d^2 + \sigma_r^2 \right] = H^*(f) \sigma_d^2$$

$$\Longrightarrow W(f) = \frac{H^*(f) \sigma_d^2}{|H(f)|^2 \sigma_d^2 + \sigma_r^2}$$

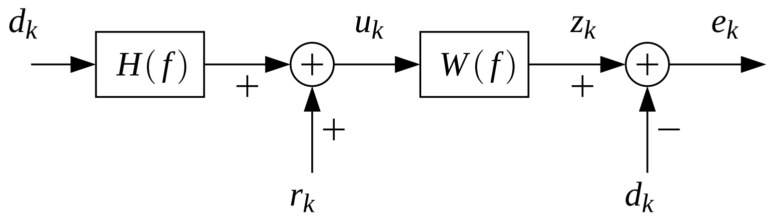


Fig. 1: General Linear Equalization Problem

The MMSE may be found from Equation (2) as

$$\begin{aligned} J_{min} &= \|d\|^2 - \|z\|^2 \\ &= E\left\{|d_k|^2\right\} - E\left\{|z_k|^2\right\} \\ &= \sigma_d^2 - T\int_{-\frac{1}{2T}}^{\frac{1}{2T}} S_{zz}(f)\ df \end{aligned}$$

$$= T\int_{-\frac{1}{2T}}^{\frac{1}{2T}} \sigma_d^2 - |W(f)|^2\left(|H(f)|^2\sigma_d^2 + \sigma_r^2\right)\ df$$

substituting for the optimum $W(f)$ gives

$$MMSE_{LE} = T\int_{-\frac{1}{2T}}^{\frac{1}{2T}} \left(\frac{\sigma_d^2\sigma_r^2}{|H(f)|^2\sigma_d^2 + \sigma_r^2}\right)\ df$$

*d) Finite-length Linear MMSE-optimum Linear Equalizer*

Unlike the equalizer derived in the previous subsection, a practical solution has to be causal and finite-length. We reformulate the problem with $w_i$ nonzero only for $i \in \{0, 1, \ldots n-1\}$. $V$ is again the space of all random variables with inner product as before. The error vector is $\mathbf{e} = e_k = \sum_{i=0}^{n-1} w_i^* u_{k-i} - d_k$ and $J(\mathbf{w}) = E\left\{|e_k|^2\right\}$ as before. The inner product matrix $\mathbf{R}$ is the correlation matrix of the sequence $\{u_k\}$, $\mathbf{c}$ is the cross-correlation vector between the sequences $\{u_k\}$ and $\{d_k\}$ and we obtain the familiar result for the optimum linear equalizer

$$\mathbf{w} = \left(E\left\{\mathbf{u}_k\mathbf{u}_k^H\right\}\right)^{-1} E\left\{\mathbf{u}_k d_k^*\right\}$$

where $\mathbf{u}_k = \begin{pmatrix} u_k & u_{k-1} & \cdots & u_{k-n+1} \end{pmatrix}^T$.

*e) Steepest Descent Algorithm*

For the general vector space $V$ , it is possible to find the minimum of $J(\{w_i\})$ via an iterative formula which moves the function arguments (i.e.$\{a_i\} \cup \{b_i\}$) along the direction of maximum decrease of the function $J$, i.e. $-\text{grad}\ J$. This is acheived via the pair of mappings

$$a_i^{(k+1)} = a_i^{(k)} - \frac{1}{2}\mu\ \frac{\partial J}{\partial a_i}\bigg|^{(k)}$$

$$b_i^{(k+1)} = b_i^{(k)} - \frac{1}{2}\mu\ \frac{\partial J}{\partial b_i}\bigg|^{(k)}$$

or their equivalent

$$w_i^{(k+1)} = w_i^{(k)} - \frac{1}{2}\mu\ \frac{\partial J}{\partial w_i}\bigg|^{(k)}$$

where $\mu$ is a step size parameter. Therefore, using (1)

$$w_i^{(k+1)} = w_i^{(k)} - \mu\langle e, x_i\rangle^*$$

This provides a steepest descent algorithm which is a generalization of that on which the LMS algorithm is based.

*f) LMS Algorithm*

For the equalizer of Figure 1, the steepest descent algorithm becomes

$$w_i^{(k+1)} = w_i^{(k)} - \mu E\left\{e_k^* u_{k-i}\right\}.$$

We can replace true correlation by sample correlation to obtain the LMS algorithm

$$w_i^{(k+1)} = w_i^{(k)} - \mu e_k^* u_{k-i}$$

*g) Finite-length Least-squares-optimum Linear Equalizer*

Consider now the problem of finding the optimum linear equalizer for Figure 1 based on the *minimum sum of error squares* criterion with a forgetting factor $\lambda \in \mathbf{R}$; i.e. we want to find $\mathbf{w}$ to minimize $\sum_{k=1}^{N} \lambda^{N-k} |e_k|^2$ based on filtering data $u_k$, $k = 1, 2, \ldots N + n - 1$. For each $k$ in $1, 2, \ldots N$ we have

$$e_k = w_1^* u_{n+k-1} + w_2^* u_{n+k-2} + \cdots + w_n^* u_k - d_k$$

This is referred to in the literature as the *Covariance Method*. If we choose for $V$ the vector space of all complex sequences of length $N$, this time with inner product $\langle x,y \rangle = \sum_{k=1}^{N} \lambda^{N-k} x_k y_k^*$, then we can express the error vector $\mathbf{e} = \begin{pmatrix} e_1 & e_2 & \cdots & e_N \end{pmatrix}^T$ in the form

$$\mathbf{e} = \sum_{i=1}^{n} w_i^* \mathbf{x}_i - \mathbf{d}$$

where $\mathbf{x}_i = \begin{pmatrix} u_{n-i+1} & u_{n-i+2} & \cdots & u_{n-i+N} \end{pmatrix}^T$. We obtain $J(\mathbf{w}) = \|\mathbf{e}\|^2 = \sum_{k=1}^{N} \lambda^{N-k} |e_k|^2$ as required. The relevant inner products in the solution for $\mathbf{w}$ are

$$c_i = \langle x_i, d \rangle = \sum_{k=1}^{N} \lambda^{N-k} u_{n-i+k} d_k^*$$

and

$$R_{ij} = \langle x_i, x_j \rangle = \sum_{k=1}^{N} \lambda^{N-k} u_{n-i+k} u_{n-j+k}^*$$

If we define $\mathbf{u}_k = \begin{pmatrix} u_{k+n-1} & u_{k+n-2} & \cdots & u_k \end{pmatrix}^T$ for each $k = 1, 2, \ldots N$, this gives a solution $\mathbf{w} = \mathbf{R}^{-1}\mathbf{c}$ with

$$\mathbf{R} = \sum_{k=1}^{N} \lambda^{N-k} \mathbf{u}_k \mathbf{u}_k^H$$

and

$$\mathbf{c} = \sum_{k=1}^{N} \lambda^{N-k} \mathbf{u}_k d_k^*$$

These are the famous Normal Equations for the optimum least squares equalizer.

## V Conclusion

We have solved the Minimum Norm problem in a general inner product space for both finite-dimensional and infinite-dimensional spaces, and illustrated that many optimization problems in signal processing are special cases of this general problem. By deriving the expressions for and properties of the general solution to the Minimum Norm problem, we illuminate the source of these similar expressions and properties and also circumvent the duplication of these results in the environment of each specific problem.

## References


[1] S. Haykin. “Adaptive Filter Theory”. Fourth Edition (Prentice Hall), 2002.

[2] D. Luenberger. “Optimization by Vector Space Methods”. Wiley, 1969.

[3] S. U. H. Qureshi. “Adaptive Equalization”. *Proc. IEEE*, Vol. 73, no. 9, pp.1349–1387, Sep. 1985.