ECE 516: Adaptive Digital Filters
Lecture 12 (Least-Squares Methods)
Algorithm 14.1 (RLS algorithm) Consider a zero-mean random variable $d$ with realizations $\{d(0), d(1), \ldots\}$, and a zero-mean random row vector $u$ with realizations $\{u_0, u_1, \ldots\}$. The weight vector $w_o$ that solves

$$\min_w E |d - uw|^2$$

can be approximated iteratively via the recursion

$$P_i = \lambda^{-1} \left[ P_{i-1} - \frac{\lambda^{-1} P_{i-1} u_i^* u_i P_{i-1}}{1 + \lambda^{-1} u_i P_{i-1} u_i^*} \right]$$

$$w_i = w_{i-1} + P_i u_i^* [d(i) - u_i w_{i-1}], \quad i \geq 0$$

with initial condition $P_{-1} = \epsilon^{-1} I$ and where $0 \ll \lambda \leq 1$.

RLS was formulated and studied as an stochastic gradient algorithm.  
--- we will look at the RLS through the well-established theory of least squares.
LEAST-SQUARES PROBLEM

Assume we have available $N$ realizations of the random variables $d$ and $u$, say,

$$\{d(0), d(1), \ldots, d(N-1)\} \text{ and } \{u_0, u_1, \ldots, u_{N-1}\}$$

respectively, where the $\{d(i)\}$ are scalars and the $\{u_i\}$ are $1 \times M$. Given the $\{d(i), u_i\}$, and assuming ergodicity, we can approximate the mean-square-error cost in (29.1) by its sample average as

$$E |d - uw|^2 \approx \frac{1}{N} \sum_{i=0}^{N-1} |d(i) - u_i w|^2 \quad (29.3)$$

In this way, the optimization problem (29.1) can be replaced by the related problem:

$$\min_w \left( \sum_{i=0}^{N-1} |d(i) - u_i w|^2 \right) \quad (29.4)$$

where we have removed the scaling factor $1/N$.

Let’s look at the vector form of (29.4)

– it’s more interesting, and also easier to deal with.
LEAST-SQUARES PROBLEM

The cost function (29.4) can be reformulated in vector notation as follows. We collect the observations \{d(i)\} into an \(N \times 1\) vector \(y\) and the row vectors \(\{u_i\}\) into an \(N \times M\) data matrix \(H\):

\[
y \triangleq \begin{bmatrix} d(0) \\ d(1) \\ d(2) \\ \vdots \\ d(N - 1) \end{bmatrix}, \quad H \triangleq \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \\ u_{N-1} \end{bmatrix}
\]

Then (29.4) can be rewritten as

\[
\min_w \| y - Hw \|^2
\]

(29.5)

where the notation \(\| \cdot \|^2\) denotes the squared Euclidean norm of its argument, namely, \(\|a\|^2 = a^*a\) for any column vector \(a\). Problem (29.5) is known as the standard least-squares problem.
LEAST-SQUARES PROBLEM

Definition 29.1 (Least-squares problem) Given an $N \times 1$ vector $y$ and an $N \times M$ data matrix $H$, the least-squares problem seeks an $M \times 1$ vector $w$ that solves $\min_{w} \|y - Hw\|^2$.

Two cases can occur depending on the relation between the dimensions \{N, M\}:

1. **Over-determined least-squares** ($N \geq M$): In this case, the data matrix $H$ has at least as many rows as columns, so that the number of measurements (i.e., the number of entries in $y$) is at least equal to the number of unknowns (i.e., the number of entries in $w$). This situation corresponds to an *over-determined* least-squares problem and, as we shall see, (29.5) will either have a unique solution or an infinite number of solutions.

2. **Under-determined least-squares** ($N < M$): In this case, the data matrix $H$ has fewer rows than columns, so that the number of measurements is less than the number of unknowns. This situation corresponds to an *under-determined* least-squares problem for which (29.5) will have an infinite number of solutions.
Definition 29.1 (Least-squares problem) Given an $N \times 1$ vector $y$ and an $N \times M$ data matrix $H$, the least-squares problem seeks an $M \times 1$ vector $w$ that solves $\min_w \| y - Hw \|^2$.

The purpose of the discussion that follows is to show that all solutions $\hat{w}$ to the least-squares problem (29.5) are characterized as solutions to the linear system of equations

$$H^*H\hat{w} = H^*y$$

which are known as the normal equations.
Differentiation Argument

Let $J(w)$ denote the cost function in (29.5), i.e.,

$$J(w) \triangleq \|y - Hw\|^2 = \|y\|^2 - y^*Hw - w^*H^*y + w^*H^*Hw$$  \hspace{1cm} (29.11)

Differentiating $J(w)$ with respect to $w$ we find that its gradient vector evaluates to zero at all $\hat{w}$ that satisfy

$$-y^*H + \hat{w}^*H^*H = 0$$

which are again the normal equations (29.7). The solution(s) $\hat{w}$ so obtained correspond to minima of $J(w)$ since its Hessian matrix is nonnegative-definite, i.e.,

$$\nabla^2_w [J(w)] = H^*H \geq 0$$
LEAST-SQUARES PROBLEM

GEOMETRIC ARGUMENT

\( y - H \hat{w}, \) is orthogonal to all vectors in \( \mathcal{R}(H) \)

A least-squares solution is obtained when \( y - H \hat{w} \) is orthogonal to \( \mathcal{R}(H) \).

Therefore, it must hold that any candidate solution \( \hat{w} \) should result in a residual vector, \( y - H \hat{w} \), that is orthogonal to \( Hp \), for any vector \( p \) or, equivalently, \( p^* H^* (y - H \hat{w}) = 0 \). Clearly, the only vector that is orthogonal to any vector \( p \) is the zero vector, so that we must have

\[ H^* (y - H \hat{w}) = 0 \]  \hspace{1cm} (29.6)

and we conclude that any solution \( \hat{w} \) of the least-squares problem (29.5) must satisfy the so-called normal equations:

\[ H^* H \hat{w} = H^* y \]  \hspace{1cm} (29.7)
LEAST-SQUARES PROBLEM

\[ \hat{y} = H \hat{\omega} \triangleq \text{projection of } y \text{ onto } \mathcal{R}(H) \]

\[ \xi = ||y - H \hat{\omega}||^2 \]

\[ = (y - H \hat{\omega})^* (y - H \hat{\omega}) \]

\[ = y^* (y - H \hat{\omega}), \text{ since } \hat{\omega}^* H^* \bar{y} = 0 \text{ by (29.9)} \]

\[ = y^* y - y^* H \hat{\omega} \]

\[ = y^* y - \hat{\omega}^* H^* H \hat{\omega}, \text{ since } y^* H = \hat{\omega}^* H^* H \text{ by (29.7)} \]

\[ = y^* y - \bar{y}^* \bar{y} \]

**Theorem 29.1 (The normal equations)** A vector \( \hat{\omega} \) solves the least-squares problem (29.5) if, and only if, it satisfies the normal equations

\[ H^* H \hat{\omega} = H^* y \]

or, equivalently, if and only if, it satisfies the orthogonality condition

\[ y - H \hat{\omega} \perp \mathcal{R}(H) \]

The normal equations are always consistent, i.e., a solution \( \hat{\omega} \) always exists and the resulting minimum cost is given by either expression:

\[ \xi = ||y||^2 - ||\bar{y}||^2 = y^* \bar{y} \]

where \( \bar{y} = H \hat{\omega} \) is the projection of \( y \) onto \( \mathcal{R}(H) \) and \( \bar{y} = y - \bar{y} \) is the residual vector.
LEAST-SQUARES PROBLEM

When many solutions \( \hat{w} \) exist, the one that has the smallest Euclidean norm, namely, the one that solves

\[
\min_{\hat{w}} \| \hat{w} \|^2 \quad \text{subject to} \quad H^* H \hat{w} = H^* y
\]

is given by \( \hat{w} = H^\dagger y \), where \( H^\dagger \) denotes the pseudo-inverse of \( H \).

**Note:** We first remark that, for a general matrix \( H \), the pseudo-inverse is defined in Sec. B.6, where the fourth statement in the theorem is also proven (see Lemma B.7). Here we note that when \( H \) has full rank, its pseudo-inverse is given by the following expressions:

\[
H^\dagger = \begin{cases} 
(H^* H)^{-1} H^* & \text{when } N > M \text{ (a “tall” matrix)} \\
H^* (H H^*)^{-1} & \text{when } N < M \text{ (a “fat” matrix)} \\
H^{-1} & \text{when } N = M \text{ (a square matrix)}
\end{cases}
\]

When \( H \) is rank-deficient, it is more convenient to define its pseudo-inverse in terms of its singular value decomposition, as explained in Sec. B.6. [See also Prob. VII.6 for a proof, from first principles, of the fourth statement of the theorem in the under-determined case.]
We restrict ourselves in this section to the case of over-determined least-squares problems with a full-rank data matrix $H$ (and, hence, $N \geq M$). In this case, the coefficient matrix $H^*H$ is invertible (actually positive-definite) and the least-squares problem (29.5) will have a unique solution that is given by

$$\hat{w} = (H^*H)^{-1}H^*y$$

with the corresponding projection vector

$$\hat{y} = H\hat{w} = H(H^*H)^{-1}H^*y$$

The matrix multiplying $y$ in the above expression is called the projection matrix and we denote it by

$$\mathcal{P}_H \triangleq H(H^*H)^{-1}H^*, \text{ when } H \text{ has full column rank}$$

(29.19)

useful properties $\mathcal{P}_H^* = \mathcal{P}_H$, $\mathcal{P}_H^2 = \mathcal{P}_H$
Lemma 29.1 (Unique solution) When the matrix $H$ has full-column rank (and, hence, $N \geq M$), the least-squares problem (29.5) will have a unique solution that is given by $\hat{w} = (H^* H)^{-1} H^* y$ Moreover, the projection of $y$ onto $\mathcal{R}(H)$, and the corresponding residual vector, are given by $\hat{y} = P_H y$ and $\tilde{y} = P_H^\perp y$ so that $y$ can be decomposed as

$$y = \hat{y} + \tilde{y} = P_H y + P_H^\perp y$$

with $\|y\|^2 = \|\hat{y}\|^2 + \|\tilde{y}\|^2$. The resulting minimum cost is $\xi = y^* P_H^\perp y$. 
WEIGHTED LEAST-SQUARES

It is often the case that weighting is incorporated into the cost function of the least-squares problem, so that (29.5) is replaced by

$$\min_w (y - Hw)^* W (y - Hw) \quad W > 0$$

(29.21)

where $W$ is a Hermitian positive-definite matrix. For example, when $W$ is diagonal, its elements assign different weights to the entries of the error vector $y - Hw$.

We shall often rewrite the cost function in (29.21) more compactly as

$$\min_w \| y - Hw \|^2_W$$

(29.22)

where, for any column vector $x$, the notation $\| x \|^2_W$ refers to the weighted Euclidean norm of $x$, i.e., $\| x \|^2_W = x^* W x$.

**Theorem 29.3 (Weighted least-squares)** A vector $\hat{w}$ is a solution of the weighted least-squares problem (29.21) if, and only if, it satisfies the normal equations $H^* W H \hat{w} = H^* W y$. 
A second variation of the standard least-squares problem (29.5) is regularized least-squares. In this formulation, we seek a vector \( \hat{w} \) that solves

\[
\min_w \left[ (w - \bar{w})^T \Pi (w - \bar{w}) + \|y - Hw\|^2 \right]
\]

(29.28)

where, compared with (29.5), we are now incorporating the so-called regularization term \( \|w - \bar{w}\|^2_\Pi \). Here, \( \Pi \) is a positive-definite matrix, usually a multiple of the identity, and \( \bar{w} \) is a given column vector, usually \( \bar{w} = 0 \).

One motivation for using regularization is that it allows us to incorporate some \textit{a priori} information about the solution into the problem statement. Assume, for instance, that we set \( \Pi = \delta I \) and choose \( \delta \) as a large positive number. Then, the first term in the cost function (29.28) becomes dominant and it is not hard to imagine that the cost will be minimized by a vector \( \hat{w} \) that is close to \( \bar{w} \) in order to offset the dominant effect of this first term. For this reason, we say that a “large” \( \Pi \) reflects high confidence that \( \bar{w} \) is a good guess for the solution \( \hat{w} \). On the other hand, a “small” \( \Pi \) indicates a high degree of uncertainty in the initial guess \( \bar{w} \).
Theorem 29.4 (Regularized least-squares) The solution of the regularized least-squares problem (29.28) is always unique and given by

\[
\hat{w} = \bar{w} + [\Pi + H^*H]^{-1} H^*(y - H\bar{w})
\]

The resulting minimum cost is given by either expression:

\[
\xi = (y - H\bar{w})^*\tilde{y} = (y - H\bar{w})^* [I + H\Pi^{-1}H^*]^{-1} (y - H\bar{w})
\]

where \(\tilde{y} = y - \hat{y}\) and \(\hat{y} = H\hat{w}\). Moreover, \(\hat{w}\) satisfies the orthogonality condition \(H^*\tilde{y} = \Pi(\hat{w} - \bar{w})\).
WEIGHTED REGULARIZED LEAST-SQUARES

We can combine the formulations of Secs. 29.6 and 29.7 and introduce a weighted regularized least-squares problem. The weighted version of (29.28) would have the form

$$\min_{\hat{w}} \left[ (w - \bar{w})^* \Pi (w - \bar{w}) + (y - Hw)^*W(y - Hw) \right]$$  \hspace{1cm} (29.37)$$

where, as before, $W$ is positive-definite. Actually, with $\Pi > 0$, the weighting matrix $W$ can be allowed to be nonnegative-definite. It is easy to verify that all the expressions in Thm. 29.5 further ahead that do not involve an inverse of $W$ will still hold.

**Theorem 29.5 (Weighted regularized least-squares)** The solution of the weighted regularized least-squares problem (29.37) is always unique and given by

$$\hat{w} = \bar{w} + \left[ \Pi + H^*WH \right]^{-1} H^*W(y - H\bar{w})$$

and the resulting minimum cost is given by

$$\xi = (y - H\bar{w})^*W\tilde{y} = (y - H\bar{w})^* \left[ W^{-1} + H\Pi^{-1}H^* \right]^{-1} (y - H\bar{w})$$

where $\tilde{y} = y - \hat{y}$ and $\hat{y} = H\hat{w}$. Moreover, $\hat{w}$ satisfies the orthogonality condition $H^*W\tilde{y} = \Pi(\hat{w} - \bar{w})$. 
### Summary of the Studied Least Squares Problems

**TABLE 29.1** Normal equations associated with several least-squares problems.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Cost function</th>
<th>Normal equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard least-squares</td>
<td>$\min_{\omega} |y - H\omega|^2$</td>
<td>$H^*H\hat{\omega} = H^*y$</td>
</tr>
<tr>
<td>Weighted least-squares</td>
<td>$\min_{\omega} |y - H\omega|^2_W, W &gt; 0$</td>
<td>$H^*WH\hat{\omega} = H^*Wy$</td>
</tr>
<tr>
<td>Regularized least-squares</td>
<td>$\min_{\omega} |\omega - \bar{\omega}|^2_\Pi + |y - H\omega|^2$</td>
<td>$(\Pi + H^<em>H)(\hat{\omega} - \bar{\omega}) = H^</em>(y - H\bar{\omega})$</td>
</tr>
<tr>
<td>Weighted regularized least-squares</td>
<td>$\min_{\omega} |\omega - \bar{\omega}|^2_\Pi + |y - H\omega|^2_W, \Pi &gt; 0, W \geq 0$</td>
<td>$(\Pi + H^*WH)(\hat{\omega} - \bar{\omega}) = H^*W(y - H\bar{\omega})$</td>
</tr>
</tbody>
</table>
Summary of the Studied Least Squares Problems

**TABLE 29.2** Orthogonality conditions associated with several least-squares problems. In the statements below, $\tilde{y} = y - \hat{y}$ where $\hat{y} = H\hat{w}$.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Cost function</th>
<th>Orthogonality condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard least-squares</td>
<td>$\min_{\hat{w}} |y - H\hat{w}|^2$</td>
<td>$H^*\tilde{y} = 0$</td>
</tr>
<tr>
<td>Weighted least-squares</td>
<td>$\min_{\hat{w}} |y - H\hat{w}|_W^2, \ W &gt; 0$</td>
<td>$H^*W\tilde{y} = 0$</td>
</tr>
<tr>
<td>Regularized least-squares</td>
<td>$\min_{\hat{w}} |w - \hat{w}|_\Pi^2 + |y - H\hat{w}|^2$ \ \ $\Pi &gt; 0$</td>
<td>$H^*\tilde{y} = \Pi(\hat{w} - \hat{w})$</td>
</tr>
<tr>
<td>Weighted regularized least-squares</td>
<td>$\min_{\hat{w}} |w - \hat{w}|_W^2 + |y - H\hat{w}|_W^2$ \ \ $\Pi &gt; 0, \ W \geq 0$</td>
<td>$H^*W\tilde{y} = \Pi(\hat{w} - \hat{w})$</td>
</tr>
</tbody>
</table>
### Summary of the Studied Least Squares Problems

**TABLE 29.3** Minimum costs associated with several least-squares problems. In the statements below, $\tilde{y} = y - \hat{y}$ where $\hat{y} = H\hat{w}$.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Cost function</th>
<th>Minimum cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard least-squares</td>
<td>$\min_w |y - Hw|^2$</td>
<td>$y^*\tilde{y}$</td>
</tr>
<tr>
<td>Weighted least-squares</td>
<td>$\min_w |y - Hw|_W^2, \quad W &gt; 0$</td>
<td>$y^*W\tilde{y}$</td>
</tr>
</tbody>
</table>
| Regularized least-squares| $\min_w \|w - \overline{w}\|_\Pi^2 + \|y - Hw\|^2$  
\quad $\Pi > 0$ | $(y - H\overline{w})^*\tilde{y}$                                           |
| Weighted regularized least-squares | $\min_w \|w - \overline{w}\|_\Pi^2 + \|y - Hw\|_W^2$  
\quad $\Pi > 0, \quad W \geq 0$ | $(y - H\overline{w})^*W\tilde{y}$                                        |