LEAST-SQUARES PROBLEM

$y - H \hat{w}$, 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

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

(29.6)

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

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

(29.7)
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:

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

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.]
### 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}</td>
<td></td>
</tr>
<tr>
<td>Weighted least-squares</td>
<td>( \min_{\omega}</td>
<td></td>
</tr>
<tr>
<td>Regularized least-squares</td>
<td>( \min_{\omega}</td>
<td></td>
</tr>
<tr>
<td>Weighted regularized least-squares</td>
<td>( \min_{\omega}</td>
<td></td>
</tr>
</tbody>
</table>
Summary of the Studied 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>$\min_{\omega}</td>
<td></td>
</tr>
<tr>
<td>Weighted least-squares</td>
<td>$\min_{\omega}</td>
<td></td>
</tr>
<tr>
<td>Regularized least-squares</td>
<td>$\min_{\omega}</td>
<td></td>
</tr>
<tr>
<td>Weighted regularized least-squares</td>
<td>$\min_{\omega}</td>
<td></td>
</tr>
</tbody>
</table>

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

*From the Weighted Regularized Least-Squares to RLS*

Given an $N \times 1$ measurement vector $y$, an $N \times M$ data matrix $H$ and an $M \times M$ positive-definite matrix $\Pi$, we saw in Sec. 29.7 that the $M \times 1$ solution to the following regularized least-squares problem:

$$\min_w \left[ w^* \Pi w + \| y - H w \|^2 \right]$$  \hspace{1cm} (30.1)

is given by

$$\hat{w} = (\Pi + H^* H)^{-1} H^* y$$  \hspace{1cm} (30.2)

where, in comparison with (29.28), we are assuming $\bar{w} = 0$ for simplicity of presentation. The arguments would apply equally well to the case $\bar{w} \neq 0$. 
**RLS Algorithm**

From the Weighted Regularized Least-Squares to RLS

We denote the individual entries of \( y \) by \( \{d(i)\} \), and the individual rows of \( H \) by \( \{u_i\} \), say,

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

so that the solution \( \hat{w} \) in (30.2) is determined by data \( \{d(i), u_i\} \) defined up to time \( N - 1 \). In order to indicate this fact explicitly, we shall write \( w_{N-1} \) instead of \( \hat{w} \) from now on, with a time subscript \( (N - 1) \). We shall also write \( y_{N-1} \) and \( H_{N-1} \) instead of \( y \) and \( H \) since these quantities are defined in terms of data up to time \( N - 1 \) as well. With this notation, we replace problem (30.1) by

\[
\min_w \left[ w^* \Pi w + \|y_{N-1} - H_{N-1} w\|^2 \right] \quad (30.3)
\]

\[
\overset{\longrightarrow}{w_{N-1}} = (\Pi + H_{N-1}^* H_{N-1})^{-1} H_{N-1}^* y_{N-1}
\]
**RLS Algorithm**

From the Weighted Regularized Least-Squares to RLS

The time-update: \[ y_N = \begin{bmatrix} y_N^{-1} \\ d(N) \end{bmatrix}, \quad H_N = \begin{bmatrix} H_N^{-1} \\ u_N \end{bmatrix} \]

\[ \text{time-updated least-squares problem} \]

\[ \min_w \left[ w^* \Pi w + \| y_N - H_N w \|^2 \right] \]

\[ w_N = (\Pi + H_N^* H_N)^{-1} H_N^* y_N \]

Remember: \[ w_{N-1} = (\Pi + H_{N-1}^* H_{N-1})^{-1} H_{N-1}^* y_{N-1} \]
**RLS Algorithm**

*From the Weighted Regularized Least-Squares to RLS*

*The time-update:*

Introduce the matrices

\[
P_N \triangleq (\Pi + H_N^* H_N)^{-1}, \quad P_{N-1} \triangleq (\Pi + H_{N-1}^* H_{N-1})^{-1}
\]  

(30.8)

with initial condition \( P_{-1} = \Pi^{-1} \). Then (30.4) and (30.7) can be written more compactly as

\[
\begin{align*}
\omega_{N-1} &= P_{N-1} H_{N-1}^* y_{N-1}, \\
\omega_N &= P_N H_N^* y_N
\end{align*}
\]  

(30.9)

The time-update relation (30.5) between \( \{y_N, H_N\} \) and \( \{y_{N-1}, H_{N-1}\} \) can be used to relate \( P_N \) to \( P_{N-1} \) and \( \omega_N \) to \( \omega_{N-1} \).
**RLS Algorithm**

*From the Weighted Regularized Least-Squares to RLS*

**The time-update:**

\[
P_N^{-1} = \Pi + H^*_N H_N \\
= \Pi + H^*_{N-1} H_{N-1} + u^*_N u_N \\
= P_{N-1}^{-1} + u^*_N u_N
\]

\[
(A + BCD)^{-1} = A^{-1} - A^{-1} B (C^{-1} + DA^{-1} B)^{-1} DA^{-1}
\]

\[
A \leftarrow P_{N-1}^{-1}, \quad B \leftarrow u^*_N, \quad C \leftarrow 1, \quad D \leftarrow u_N
\]

\[
P_N = P_{N-1} - \frac{P_{N-1} u^*_N u_N P_{N-1}}{1 + u_N P_{N-1} u^*_N}, \quad P_{-1} = \Pi^{-1}
\]
**RLS Algorithm**

From the Weighted Regularized Least-Squares to RLS

**The time-update:**

\[ P_N = P_{N-1} - \frac{P_{N-1} u_N^* u_N P_{N-1}}{1 + u_N P_{N-1} u_N^*}, \quad P_{-1} = \Pi^{-1} \]

\[ w_N = P_N \left[ H_{N-1}^* y_{N-1} + u_N^* d(N) \right] \]

\[ = \left( P_{N-1} - \frac{P_{N-1} u_N^* u_N P_{N-1}}{1 + u_N P_{N-1} u_N^*} \right) \left[ H_{N-1}^* y_{N-1} + u_N^* d(N) \right] \]

\[ = \frac{P_{N-1} H_{N-1}^* y_{N-1}}{1 + u_N P_{N-1} u_N^*} - \frac{P_{N-1} u_N^*}{1 + u_N P_{N-1} u_N^*} u_N \frac{P_{N-1} H_{N-1}^* y_{N-1}}{1 + u_N P_{N-1} u_N^*} = w_{N-1} \]

\[ + P_{N-1} u_N^* \left( 1 - \frac{u_N P_{N-1} u_N^*}{1 + u_N P_{N-1} u_N^*} \right) d(N) \]

\[ w_N = w_{N-1} + \frac{P_{N-1} u_N^*}{1 + u_N P_{N-1} u_N^*} \left[ d(N) - u_N w_{N-1} \right], \quad w_{-1} = 0 \]
**RLS Algorithm**

From the Weighted Regularized Least-Squares to RLS

**The time-update:**

\[
P_N = P_{N-1} - \frac{P_{N-1}u_N^*u_NP_{N-1}}{1 + u_NP_{N-1}u_N^*}, \quad P_{-1} = \Pi^{-1}
\]

\[
w_N = P_N \left[ H_{N-1}^*y_{N-1} + u_N^*d(N) \right]
\]

\[
= \left( P_{N-1} - \frac{P_{N-1}u_N^*u_NP_{N-1}}{1 + u_NP_{N-1}u_N^*} \right) \left[ H_{N-1}^*y_{N-1} + u_N^*d(N) \right]
\]

\[
= \frac{P_{N-1}H_{N-1}^*y_{N-1}}{1 + u_NP_{N-1}u_N^*} - \frac{P_{N-1}u_N^*}{1 + u_NP_{N-1}u_N^*}u_N \frac{P_{N-1}H_{N-1}^*y_{N-1}}{1 + u_NP_{N-1}u_N^*} = w_{N-1}
\]

\[
+ P_{N-1}u_N^* \left( 1 - \frac{u_NP_{N-1}u_N^*}{1 + u_NP_{N-1}u_N^*} \right) d(N)
\]

\[
w_N = w_{N-1} + \frac{P_{N-1}u_N^*}{1 + u_NP_{N-1}u_N^*} [d(N) - u_Nw_{N-1}], \quad w_{-1} = 0
\]
**RLS Algorithm**

**Algorithm 30.1 (RLS algorithm)** Given $\Pi > 0$, the solution $w_N$ that minimizes the cost

$$w^* \Pi w + \|y_N - H_N w\|^2$$

can be computed recursively as follows. Start with $w_{-1} = 0$ and $P_{-1} = \Pi^{-1}$ and iterate for $i \geq 0$:

- $\gamma(i) = 1/(1 + u_i P_{i-1} u_i^*)$
- $g_i = P_{i-1} u_i^* \gamma(i)$
- $w_i = w_{i-1} + g_i [d(i) - u_i w_{i-1}]$
- $P_i = P_{i-1} - g_i g_i^* / \gamma(i)$

At each iteration, it holds that $w_i$ minimizes $w^* \Pi w + \|y_i - H_i w\|^2$, where $y_i = \text{col} \{d(0), d(1), \ldots, d(i)\}$ and the rows of $H_i$ are $\{u_0, u_1, \ldots, u_i\}$. Moreover, $P_i = (\Pi + H_i^* H_i)^{-1}$.

\[ \gamma(i) \quad \text{CONVERSION FACTOR} \quad ? \]
EXPONENTIALLY-WEIGHTED RLS ALGORITHM

It is more common in adaptive filtering to employ a *weighted* regularized least-squares cost function, as opposed to the unweighted cost in (30.6). More specifically, a diagonal weighting matrix is used whose purpose is to give more weight to recent data and less weight to data from the remote past.
EXPONENTIALLY-WEIGHTED RLS ALGORITHM

Let $\lambda$ be a positive scalar, usually very close to one (e.g., $\lambda = 0.998$ or some similar value), say, $0 \ll \lambda \leq 1$, and introduce the diagonal matrix

$$
\Lambda_N \overset{\Delta}{=} \text{diag}\{\lambda^N, \lambda^{N-1}, \ldots, \lambda, 1\}
$$

(30.25)

Then replace (30.6) by

$$
\min_w \left[ \lambda^{(N+1)}w^*\Pi w + (y_N - H_Nw)^*\Lambda_N(y_N - H_Nw) \right]
$$

(30.26)

or, more explicitly, by

$$
\min_w \left[ \lambda^{(N+1)}w^*\Pi w + \sum_{j=0}^{N} \lambda^{N-j} |d(j) - u_jw|^2 \right]
$$

(30.27)

The scalar $\lambda$ is called the forgetting factor since past data are exponentially weighted less heavily than more recent data.
Algorithm 30.2 (Exponentially-weighted RLS) Given $\Pi > 0$, and a forgetting factor $0 < \lambda \leq 1$, the solution $w_N$ of the exponentially-weighted regularized least-squares problem (30.27), and the corresponding minimum cost $\xi(N)$, can be computed recursively as follows. Start with $w_{-1} = 0$, $P_{-1} = \Pi^{-1}$, and $\xi(-1) = 0$, and iterate for $i \geq 0$:

\[
\begin{align*}
\gamma(i) &= 1/(1 + \lambda^{-1} u_i P_{i-1} u_i^*) \\
g_i &= \lambda^{-1} P_{i-1} u_i^* \gamma(i) \\
e(i) &= d(i) - u_i w_{i-1} \\
w_i &= w_{i-1} + g_i e(i) \\
P_i &= \lambda^{-1} P_{i-1} - g_i g_i^* / \gamma(i) \\
\xi(i) &= \lambda \xi(i-1) + \gamma(i) |e(i)|^2
\end{align*}
\]

At each iteration, $P_i$ has the interpretation $P_i = [\lambda^{(i+1)} \Pi + H_i^* \Lambda_i H_i]^{-1}$ and $w_i$ is the solution of

\[
\min_w \left[ \lambda^{(i+1)} w^* \Pi w + \sum_{j=0}^{i} \lambda^{i-j} |d(j) - u_j w|^2 \right]
\]

In addition, as was the case with (30.16)–(30.17), the following relations hold:

\[
g_i = P_i u_i^*, \quad \gamma(i) = 1 - u_i P_i u_i^* = 1 - u_i g_i, \quad r(i) = \gamma(i) e(i)
\]

where $r(i) = d(i) - u_i w_i$. 