Optimally Regularized Recursive Least Squares for Acoustic Echo Cancellation

Toon van Waterschoot, Geert Rombouts, and Marc Moonen

SPS-DARTS 2006

Katholieke Universiteit Leuven,
ESAT-SCD,
Kasteelpark Arenberg 10,
B-3001 Leuven, Belgium,
http://homes.esat.kuleuven.be/∼tvanwate
toon.vanwaterschoot@esat.kuleuven.be

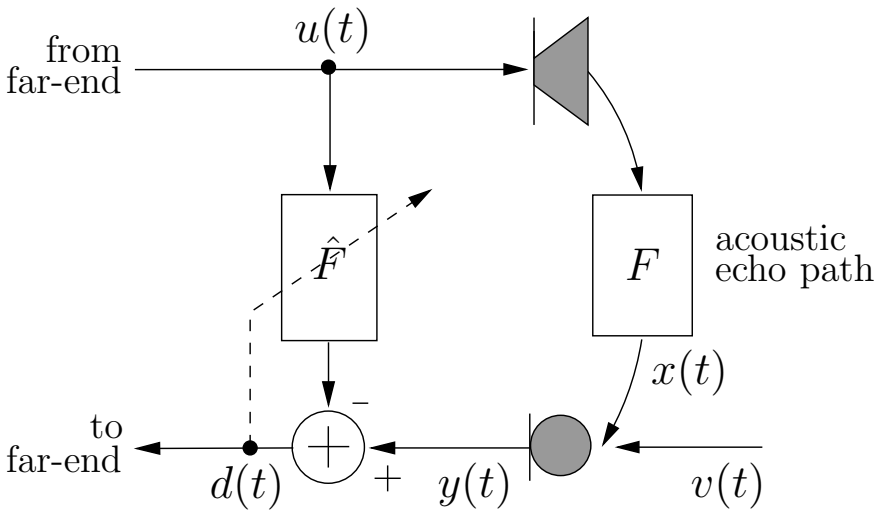


Outline

- Introduction:
  - Acoustic Echo Cancellation: applications
  - Acoustic Echo Cancellation: signal and system description
  - Exponentially weighted RLS
- The poor excitation problem and existing solutions:
  - Problem statement
  - Levenberg-Marquardt regularization (LMR-RLS)
  - Tikhonov regularization (TR-RLS)
- Optimal regularization:
  - Linear MMSE estimate
  - Optimally regularized RLS algorithms
  - Constructing an optimal regularization matrix
- Algorithm performance and complexity issues:
  - Simulation setup and parameters
  - Simulation results
  - Reducing computational complexity
- Conclusions and ongoing work

Introduction: Acoustic Echo Cancellation

- Echo Problem:
  $y(t) = \mathbf{u}^T(t)\mathbf{f} + v(t)$
- Echo Canceller:
  $d(t) = y(t) - \mathbf{u}^T(t)\hat{\mathbf{f}}$
- Room Impulse Response (RIR):
  $\mathbf{f} = \begin{bmatrix} f_0 & f_1 & \dots & f_{n_F} \end{bmatrix}^T$
- Data Model:
  $\mathbf{u}^T(t) = \begin{bmatrix} u(t) & \dots & u(t-n_F) \end{bmatrix}$



$$\underbrace{\begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(1) \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} u(t) & u(t-1) & \dots & u(t-n_F) \\ u(t-1) & u(t-2) & \dots & u(t-n_F-1) \\ \vdots & \vdots & \ddots & \vdots \\ u(1) & u(0) & \dots & u(1-n_F) \end{bmatrix}}_{\mathbf{U}} \mathbf{f} + \underbrace{\begin{bmatrix} v(t) \\ v(t-1) \\ \vdots \\ v(1) \end{bmatrix}}_{\mathbf{v}}$$

Introduction: AEC Applications

- Full-duplex hands-free communication:
  - teleconferencing, hands-free telephony, ...
  - far-end side is similar to near-end side, with communication channel in between
- Adaptive Feedback Cancellation:
  - Public Address systems, hearing aids, ...
  - far-end side is sound amplification system
  - signals are generated in closed-loop
    - $\Rightarrow$ continuous double-talk situation
    - $\Rightarrow$ near-end signal correlated with far-end signal
    - $\Rightarrow$ risk of system instability (howling)
- Automatic Gain Adjustment:
  - Public Address systems, ...
  - goal is to adjust the far-end signal volume to the near-end signal level
  - far-end side is gain control system
- ...

Introduction: Exponentially Weighted RLS

- The RIR $\mathbf{f}$ may be estimated using the well-known exponentially weighted RLS algorithm:

$$\begin{aligned} \hat{\mathbf{f}}(t) &= \hat{\mathbf{f}}(t-1) + \mathbf{R}^{-1}(t)\mathbf{u}(t)\varepsilon(t) \\ \mathbf{R}(t) &= \lambda\mathbf{R}(t-1) + \mathbf{u}(t)\mathbf{u}^T(t) \\ \varepsilon(t) &= y(t) - \mathbf{u}^T(t)\hat{\mathbf{f}}(t-1) \end{aligned} \tag{1}$$

- The $O(n_F^3)$ complexity can be reduced to $O(n_F^2)$ by applying the matrix inversion lemma to (1).
- Fast $O(n_F)$ algorithms may be derived by exploiting shift structure in the far-end signal vector $\mathbf{u}(t)$.
- Other $O(n_F)$ adaptive algorithms are the APA and NLMS algorithm, yet these exhibit a slower convergence behaviour for correlated far-end signals, such as speech and audio.

Problem statement: poor excitation in RLS

- If the far-end signal is not (or hardly) persistently exciting of order n_F, then the RLS covariance matrix $\mathbf{R}(t)$ will be singular (or ill-conditioned):

$$\mathbf{R}(t) = \lambda \mathbf{R}(t-1) + \mathbf{u}(t)\mathbf{u}^T(t)$$

- As a consequence, the RLS algorithm will become very sensitive to near-end signal activity (double-talk).

- In acoustic echo cancellation, this is a common situation, since
  - the far-end signal is a speech or audio signal, that may exhibit a nearly tonal spectrum
  - the RIR order n_F may be very high (e.g. for a small office $n_F \geq 1000$ at $f_s = 8$ kHz)

- Turning the above ill-posed problem into a well-posed one is known as regularization.

Existing solutions (1) and (2)

- Solution (1): Soft-constrained initialization [Haykin, 1986] consists in initializing the RLS covariance matrix as a scaled identity matrix:

$$\mathbf{R}(0) = \alpha\mathbf{I}$$

- Due to the exponential weighting, however, this initialization will be forgotten along with old data and hence the conditioning of $\mathbf{R}(t)$ deteriorates as time goes along.

- Solution (2): Levenberg-Marquardt regularization [Ljung & Söderström, 1986] can be performed by adding the forgotten part of the initial regularization matrix to the RLS covariance matrix in each iteration:

$$\mathbf{R}(t) = \lambda\mathbf{R}(t-1) + \mathbf{u}(t)\mathbf{u}^T(t) + (1-\lambda)\alpha\mathbf{I}$$

- The regularization parameter α is a small constant, which is in AEC typically chosen to be a rough estimate of the near-end signal power [Gay, 1994].

Existing solution (3)

- A similar poor excitation problem may occur in (non-recursive) least squares estimation, rendering the matrix $\mathbf{U}^T\mathbf{U}$ ill-conditioned:

$$\hat{\mathbf{f}} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{y}$$

- Solution (3): Tikhonov regularization [Tikhonov & Arsenin, 1977] consists in modifying the LS criterion by adding a quadratic constraint:

$$\min_{\hat{\mathbf{f}}}(\mathbf{y} - \mathbf{U}\hat{\mathbf{f}})^T(\mathbf{y} - \mathbf{U}\hat{\mathbf{f}}) + \alpha\hat{\mathbf{f}}^T\hat{\mathbf{f}}$$

  which leads to the regularized LS estimate:

$$\hat{\mathbf{f}} = (\mathbf{U}^T\mathbf{U} + \alpha\mathbf{I})^{-1}\mathbf{U}^T\mathbf{y}$$

- With exponential weighting, this results in a Tikhonov regularized (or "leaky") RLS algorithm [Horita, 2004]:

$$\begin{aligned}\hat{\mathbf{f}}(t) &= \hat{\mathbf{f}}(t-1) + \mathbf{R}^{-1}(t)\big(\mathbf{u}(t)\varepsilon(t) - (1-\lambda)\alpha\hat{\mathbf{f}}(t-1)\big)\\ \mathbf{R}(t) &= \lambda\mathbf{R}(t-1) + \mathbf{u}(t)\mathbf{u}^T(t) + (1-\lambda)\alpha\mathbf{I}\end{aligned}$$

Optimal regularization: LMMSE estimate (1)

- Motivation: can we optimize regularization by including more than just one regularization parameter?
- Bayesian Framework: we consider both the near-end signal and the true RIR as stochastic vector processes with:

$$\begin{cases} E\{\mathbf{v}\} = \mathbf{0} \\ \text{cov}\{\mathbf{v}\} = E\{\mathbf{v}\mathbf{v}^T\} = \mathbf{R}_\mathbf{v} \end{cases} \quad \text{and} \quad \begin{cases} E\{\mathbf{f}\} = \mathbf{0} \\ \text{cov}\{\mathbf{f}\} = E\{\mathbf{f}\mathbf{f}^T\} = \mathbf{R}_\mathbf{f} \end{cases}$$

- LMMSE: in this framework, the linear minimum mean square error estimate of the RIR appears to be a regularized and weighted LS estimate [Luenberger, 1969]:

$$\begin{aligned} \hat{\mathbf{f}} &= \min_{\hat{\mathbf{f}}} E\{(\hat{\mathbf{f}} - \mathbf{f})^T(\hat{\mathbf{f}} - \mathbf{f})\} \\ &= (\mathbf{U}^T {\mathbf{R}_\mathbf{v}}^{-1}\mathbf{U} + {\mathbf{R}_\mathbf{f}}^{-1})^{-1}\mathbf{U}^T {\mathbf{R}_\mathbf{v}}^{-1}\mathbf{y} \end{aligned}$$

Optimal regularization: LMMSE estimate (2)

- In the LMMSE estimate, $\hat{\mathbf{f}} = (\mathbf{U}^T {\mathbf{R_v}}^{-1}\mathbf{U} + {\mathbf{R_f}}^{-1})^{-1}\mathbf{U}^T {\mathbf{R_v}}^{-1}\mathbf{y}$, data weighting is performed with the inverse near-end signal covariance matrix ${\mathbf{R_v}}^{-1}$. Without loss of generality, we assume that the near-end signal is white (non-white near-end signals are tackled in [van Waterschoot et al., 2005]):

$$\mathbf{R_v} = \begin{bmatrix} \sigma_v^2(t) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_v^2(1) \end{bmatrix}$$

- Regularization is performed with the inverse true RIR covariance matrix ${\mathbf{R_f}}^{-1}$, which is not necessarily a scaled identity matrix. In this way, prior knowlegde on the true RIR can be incorporated in the regularization matrix. For computational reasons, we will restrict ${\mathbf{R_f}}^{-1}$ to be diagonal.
- The LMMSE estimate is equivalent to the classical Tikhonov regularized LS estimate if both the near-end signal and the true RIR are drawn from a stationary GWN vector process (i.e. $\mathbf{R_v} = \sigma_v^2\mathbf{I}$ and $\mathbf{R_f} = \sigma_f^2\mathbf{I}$), and if

$$\alpha = \frac{\sigma_v^2}{\sigma_f^2}$$

Optimally regularized RLS algorithms

- Tikhonov Regularized Recursive Least Squares (TR-RLS)

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \mathbf{R}^{-1}(t)\Big(\frac{1}{\sigma_v^2(t)}\mathbf{u}(t)\varepsilon(t) - (1-\lambda)\mathbf{R_f}^{-1}\hat{\mathbf{f}}(t-1)\Big)$$

$$\mathbf{R}(t) = \lambda\mathbf{R}(t-1) + \frac{1}{\sigma_v^2(t)}\mathbf{u}(t)\mathbf{u}^T(t) + (1-\lambda)\mathbf{R_f}^{-1}$$

$$\varepsilon(t) = y(t) - \mathbf{u}^T(t)\hat{\mathbf{f}}(t-1)$$

- Levenberg-Marquardt Regularized Recursive Least Squares (LMR-RLS)

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \frac{1}{\sigma_v^2(t)}\mathbf{R}^{-1}(t)\mathbf{u}(t)\varepsilon(t)$$

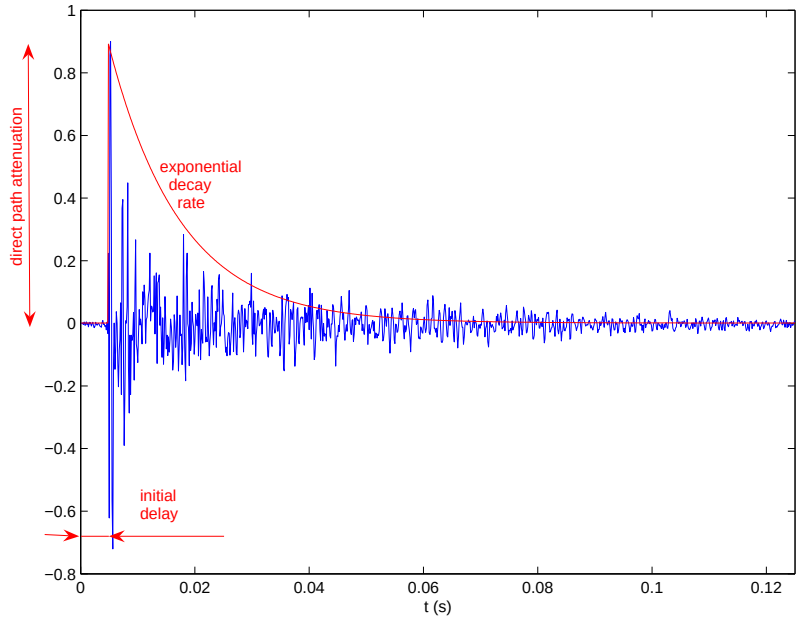
$$\mathbf{R}(t) = \lambda\mathbf{R}(t-1) + \frac{1}{\sigma_v^2(t)}\mathbf{u}(t)\mathbf{u}^T(t) + (1-\lambda)\mathbf{R_f}^{-1}$$

$$\varepsilon(t) = y(t) - \mathbf{u}^T(t)\hat{\mathbf{f}}(t-1)$$

Optimal regularization: RIR prior knowledge (1)

- Optimal regularization may be achieved by incorporating prior knowledge on the true RIR in the regularization matrix.
- A RIR has a very typical form, which may be characterized by three parameters, that can be calculated a priori from the acoustic setup:

- initial delay d: depends on the loudspeaker-microphone distance
- direct path attenuation A: depends on the loudspeaker-microphone distance
- exponential decay time constant τ: depends on the room geometry and absorption coefficients



Optimal regularization: RIR prior knowledge (2)

- A diagonal regularization matrix $\hat{\mathbf{R}}_{\mathbf{f},\mathbf{3}}^{-1}$ based on this 3-parameter RIR model may be constructed with:

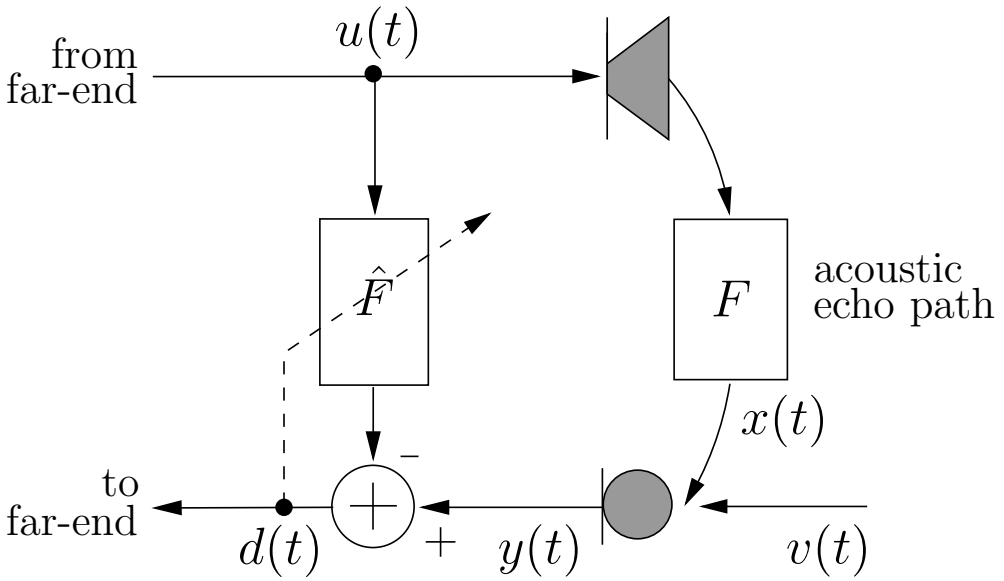
$$\hat{\mathbf{R}}_{\mathbf{f},\mathbf{3}} = A \cdot \begin{bmatrix} \beta & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \beta & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & e^{-2\frac{1}{\tau}} & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & e^{-2\frac{n_F-d}{\tau}} \end{bmatrix}$$

$$\underbrace{\qquad\qquad}_{d} \qquad \underbrace{\qquad\qquad\qquad\qquad}_{n_F-d}$$

- For the RIR displayed previously, the parameters may be set as $d = 38$, $A = 0.9$, $\tau = 100$ and $\beta = 10^{-6}$.

Algorithm performance: Simulation setup and parameters

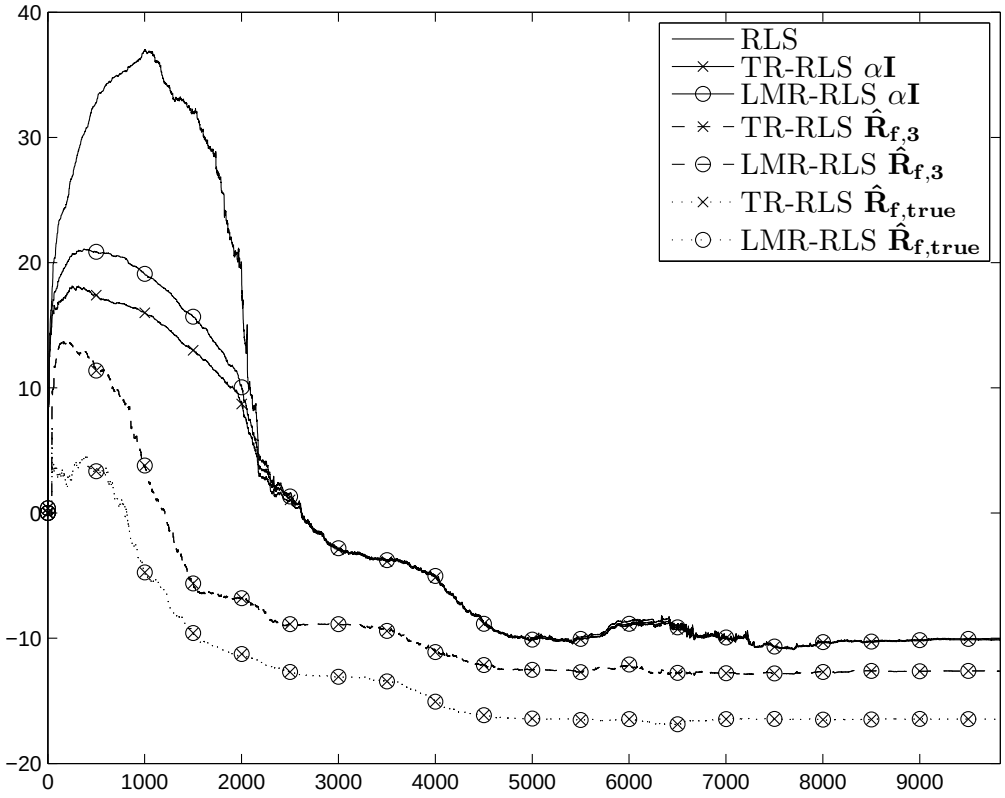
- Simulation parameters:
  - $f_s = 8kHz$
  - $n_F = 1000$, $\lambda = 1 - (3n_F)^{-1}$
  - $u(t)$ = male speech
  - $v(t)$ = stationary GWN with known variance σ_v^2
  - echo-to-near-end ratio $\frac{\sum_{k=1}^{N} |x(k)|^2}{\sum_{k=1}^{N} |v(k)|^2} = 18dB$



- The TR-RLS and LMR-RLS algorithms have been evaluated with three different regularization matrices:
  - $\sigma_v^2\mathbf{I}$: traditional one-parameter regularization with $\alpha = \sigma_v^2$
  - $\sigma_v^2\hat{\mathbf{R}}_{\mathbf{f,3}}^{-1}$: regularization using proposed 3-parameter RIR model
  - $\sigma_v^2\hat{\mathbf{R}}_{\mathbf{f,true}}^{-1}$: regularization using diagonal matrix based on true RIR

Algorithm performance: Simulation results

Performance measure: misadjustment (dB) $= 20 \log_{10} \frac{\|\hat{\mathbf{f}}(t) - \mathbf{f}\|}{\|\mathbf{f}\|}$



Reducing computational complexity

$$\mathbf{R}(t) = \lambda \mathbf{R}(t-1) + \frac{1}{\sigma_v^2(t)} \mathbf{u}(t)\mathbf{u}^T(t) + (1-\lambda){\mathbf{R}_\mathbf{f}}^{-1}$$

Due to adding a full-rank regularization matrix in each iteration of the RLS covariance matrix update, the matrix inversion lemma (MIL) cannot be applied to reduce RLS complexity from $O(n_F^3)$ to $0(n_F^2)$. Still, an $0(n_F^2)$ complexity can be achieved by applying one of the following techniques:

1. Rectangular window: if a rectangular data window is used instead of an exponential window, the initial regulariation matrix will not be forgotten. Applying the MIL twice then leads to an $0(n_F^2)$ complexity [Houacine, 1990]. This procedure is however numerically not appealing.
2. Decomposing the regularization matrix: every rank-$(n_F + 1)$ regularization matrix can be approximated by a sum of $n_F + 1$ rank-one matrices. Adding only one of these rank-one matrices in each iteration (with appropriate scaling), and applying the MIL twice leads to an $0(n_F^2)$ algorithm [Ljung & Soderstrom, 1986; Gay, 1996].
3. SVD updating: updating [Gu & Eisenstat, 1994], regularizing and inverting the eigenvalue decomposition of the RLS covariance matrix $\mathbf{R}(t)$ instead of $\mathbf{R}(t)$ itself can be performed in $0(n_F^2)$ operations [Sayed et al., 1997; Horita et al., 2004]

Conclusions

- From Bayesian LMMSE estimation we know that incorporating prior knowledge on both the near-end signal and the true RIR can lead to a low-variance RIR estimate.
- We have therefore proposed two optimally regularized RLS algorithms, focussing on a regularization matrix in which true RIR prior knowledge can be contained, in analogy to the traditional Tikhonov and Levenberg-Marquardt regularized algorithms.
- A method for extracting prior knowledge on the RIR from the acoustic setup has been proposed, using a 3-parameter RIR model.
- The optimally regularized RLS algorithms outperform the standard RLS algorithm as well as the traditionally regularized algorithms, in terms of convergence speed and steady-state misadjustment.
- We have proposed three techniques for reducing the computational complexity of regularized RLS algorithms from $O(n_F^3)$ to $O(n_F^2)$ flops per iteration.

Ongoing work

- Derivation of optimally regularized underdetermined recursive least squares (URLS) algorithms, such as the affine projection algorithm (APA) and the normalized least mean squares (NLMS) algorithm. [van Waterschoot et al., 2006, see $http://homes.esat.kuleuven.be/{\sim}tvanwate$]
- Derivation of fast (FTF- or lattice-based) and numerically stable (QRD-based) optimally regularized RLS algorithms.