Dually Regularized Recursive Prediction Error Identification for Acoustic Feedback and Echo Cancellation

Toon van Waterschoot, Geert Rombouts, and Marc Moonen

SISTA Seminar 30/08/2007

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:
  - From Least-Squares to Prediction Error Identification
  - Application 1: Double-talk-robust Acoustic Echo Cancellation
  - Application 2: Unbiased Acoustic Feedback Cancellation
- Problem Statement:
  - State-of-the-art: Stochastic Gradient RPE Algorithm
  - Core Problem: Scaling Ambiguity in the RPE Algorithm
  - Consequences of the Scaling Ambiguity
- Proposed Solution:
  - Dual Regularization: Regularizing the Auxiliary Variables Identification
  - Combined Primal and Dual Regularization
  - Dually Regularized Recursive Prediction Error Identification (DR-RPE) Algorithm
- Simulation Results:
  - Dual Regularization Matrix Construction using Prior Knowledge on Speech Characteristics
  - Simulation 1: Evaluation of Improvement in Convergence of RIR Coefficients Estimate
  - Simulation 2: Evaluation of Improvement in Accuracy of TVAR Coefficients Estimate
- Conclusions

From Least-Squares to Prediction Error Identification

- Goal: identify a room impulse response (RIR) $F(q,t) = f_0(t) + f_1(t)q^{-1} + \ldots + f_{n_F}(t)q^{-n_F}$, given a far-end signal $u(t)$ and microphone signal $y(t)$, possibly disturbed by the presence of a near-end signal $v(t)$.
- Least-squares (LS) identification:

$$\text{data model: } y(t) = F(q,t)u(t) + v(t)$$

$$\text{LS criterion: } \min_{\hat{\mathbf{f}}(t)} \frac{1}{2N} \sum_{k=1}^{t} \lambda^{t-k} [y(k) - \hat{F}(q,t)u(k)]^2$$

- Prediction error (PE) identification: include a time-varying AR (TVAR) near-end signal model

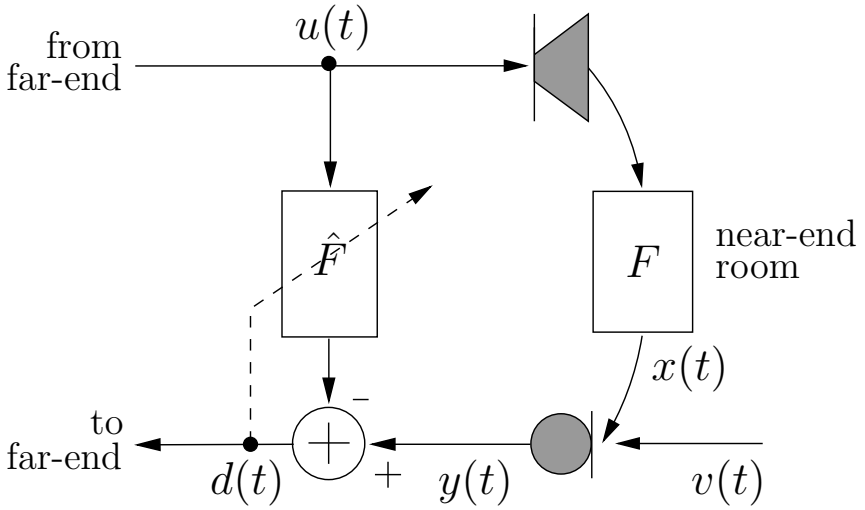
$$\text{data model: } y(t) = F(q,t)u(t) + \frac{1}{A(q,t)}e(t)$$

$$\text{PE criterion: } \min_{\hat{\mathbf{f}}(t),\hat{\mathbf{a}}(t)} \frac{1}{2N} \sum_{k=1}^{t} \frac{\lambda^{t-k}}{\hat{\sigma}_k^2} \{\hat{A}(q,t)[y(k) - \hat{F}(q,t)u(k)]\}^2$$

Application 1: Double-talk-robust Acoustic Echo Cancellation (1)

Double-talk in Acoustic Echo Cancellation (AEC):

- **bursting** double-talk may occur due to double-talk detector (DTD) imperfections
- **continuous** double-talk may occur for noisy near-end conditions or in applications other than teleconferencing, and does not allow the use of a DTD



- as a consequence of bursting and continuous double-talk, there is a need for increasing the **double-talk-robustness** of the AEC adaptive filtering algorithm, irrespective of the DTD algorithm which is used
- ideally, double-talk-robust adaptive filtering algorithms are able to operate **without a DTD** such that echo path changes during double-talk can still be tracked

Application 1: Double-talk-robust Acoustic Echo Cancellation (2)

How does the use of prediction error identification enable double-talk-robust AEC?

- if the near-end signal truly admits an autoregressive representation of the same order as the TVAR model, then the global solution to the PE criterion

$$\min_{\hat{\mathbf{f}}(t),\hat{\mathbf{a}}(t)} \frac{1}{2N} \sum_{k=1}^{t} \frac{\lambda^{t-k}}{\hat{\sigma}_k^2} \left\{ \hat{A}(q,t)[y(k) - \hat{F}(q,t)u(k)] \right\}^2$$

  is the so-called Best (i.e., minimum-variance) Linear Unbiased Estimate (BLUE) of the RIR [1]:

$$\hat{\mathbf{f}}_{\text{BLUE}} = (\mathbf{U}^T {\mathbf{R}_\mathbf{v}}^{-1} \mathbf{U})^{-1} \mathbf{U}^T {\mathbf{R}_\mathbf{v}}^{-1} \mathbf{y}$$

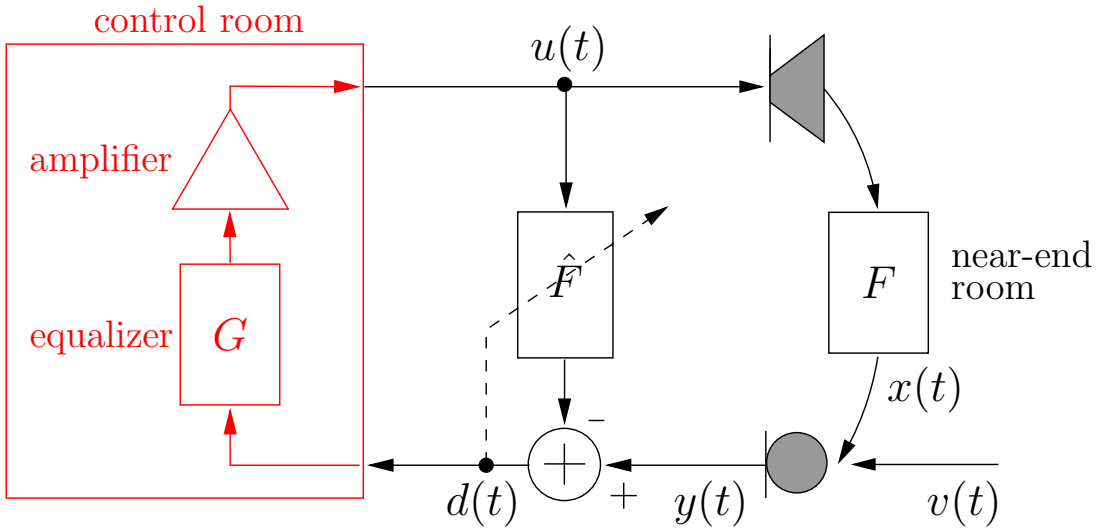
$$\mathbf{U} = \begin{bmatrix} u(t) & \dots & u(t-n_F) \\ \vdots & \ddots & \vdots \\ u(1) & \dots & u(1-n_F) \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y(t) \\ \vdots \\ y(1) \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v(t) \\ \vdots \\ v(1) \end{bmatrix}, \quad \mathbf{R}_\mathbf{v} = E\{\mathbf{v}\mathbf{v}^T\}$$

[1] T. van Waterschoot, G. Rombouts, P. Verhoeve, and M. Moonen, "Double-talk-robust prediction error identification algorithms for acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 846–858, Mar. 2007.

Application 2: Unbiased Acoustic Feedback Cancellation

Bias in Acoustic Feedback Cancellation (AFC):

- due to the closed signal loop, the far-end signal is correlated with the near-end signal, hence the LS estimate of the RIR is **biased**
- bias removal is possible by **signal decorrelation**, which is performed either in the closed signal loop or in the adaptive filtering algorithm
- signal decorrelation **in the adaptive filtering algorithm** is possible by prefiltering the adaptive filter's input and desired signal with the inverse near-end signal model, which is exactly what happens in PE identification [2]



[2] G. Rombouts, T. van Waterschoot, K. Struyve, and M. Moonen, "Acoustic feedback suppression for long acoustic paths using a nonstationary source model," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3426–3434, Sept. 2006.

State-of-the-art: Stochastic Gradient RPE Algorithm [1]

Prediction error:
$$\varepsilon[t, \hat{\boldsymbol{\theta}}(t-1)] = \begin{bmatrix} 1 & \hat{\mathbf{a}}^T(t-1) \end{bmatrix} \left(\begin{bmatrix} y(t) \\ \mathbf{y}(t) \end{bmatrix} - \begin{bmatrix} \mathbf{u}^T(t) \\ \mathbf{U}^T(t) \end{bmatrix} \hat{\mathbf{f}}(t-1) \right)$$

Prediction error variance:
$$\hat{\sigma}_t^2 = \lambda_A \hat{\sigma}_{t-1}^2 + (1-\lambda_A)\varepsilon^2[t, \hat{\boldsymbol{\theta}}(t-1)]$$

Regression vectors:
$$\boldsymbol{\psi}_F[t, \hat{\mathbf{a}}(t-1)] = \begin{bmatrix} \mathbf{u}(t) & \mathbf{U}(t) \end{bmatrix} \begin{bmatrix} 1 \\ \hat{\mathbf{a}}(t-1) \end{bmatrix}$$

$$\boldsymbol{\psi}_A[t, \hat{\mathbf{f}}(t-1)] = \mathbf{U}^T(t)\hat{\mathbf{f}}(t-1) - \mathbf{y}(t)$$

TVAR correlation matrix update:
$$\mathbf{R}_A(t) = \lambda_A \mathbf{R}_A(t-1) + \frac{1}{\hat{\sigma}_t^2} \boldsymbol{\psi}_A[t, \hat{\mathbf{f}}(t-1)] \boldsymbol{\psi}_A^T[t, \hat{\mathbf{f}}(t-1)]$$

TVAR and RIR weight updates:
$$\hat{\mathbf{a}}(t) = \hat{\mathbf{a}}(t-1) + \frac{1}{\hat{\sigma}_t^2} \mathbf{R}_A^{-1}(t) \boldsymbol{\psi}_A[t, \hat{\mathbf{f}}(t-1)] \varepsilon[t, \hat{\boldsymbol{\theta}}(t-1)]$$

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \mu_F \frac{\boldsymbol{\psi}_F[t, \hat{\mathbf{a}}(t-1)] \varepsilon[t, \hat{\boldsymbol{\theta}}(t-1)]}{\boldsymbol{\psi}_F^T[t, \hat{\mathbf{a}}(t-1)] \boldsymbol{\psi}_F[t, \hat{\mathbf{a}}(t-1)] + \hat{\sigma}_t^2}$$

Core Problem: Scaling Ambiguity in the RPE Algorithm

- Scaling ambiguity: if some of the auxiliary variables are replaced by their scaled counterparts

$$\left\{\hat{\mathbf{a}}(t), \hat{\mathbf{a}}(t-1), \hat{\sigma}_t^2, \hat{\sigma}_{t-1}^2, \boldsymbol{\psi}_F[t, \hat{\mathbf{a}}(t-1)], \varepsilon[t, \hat{\boldsymbol{\theta}}(t-1)], \mathbf{R}_A(t), \mathbf{R}_A(t-1)\right\}$$

$$\downarrow$$

$$\left\{K\hat{\mathbf{a}}(t), K\hat{\mathbf{a}}(t-1), K^2\hat{\sigma}_t^2, K^2\hat{\sigma}_{t-1}^2, K\boldsymbol{\psi}_F[t, \hat{\mathbf{a}}(t-1)], K\varepsilon[t, \hat{\boldsymbol{\theta}}(t-1)], K^{-2}\mathbf{R}_A(t), K^{-2}\mathbf{R}_A(t-1)\right\}$$

and if the scaling factor is large enough,

$$K >> \|\hat{\mathbf{a}}(t-1)\|_2^{-1}$$

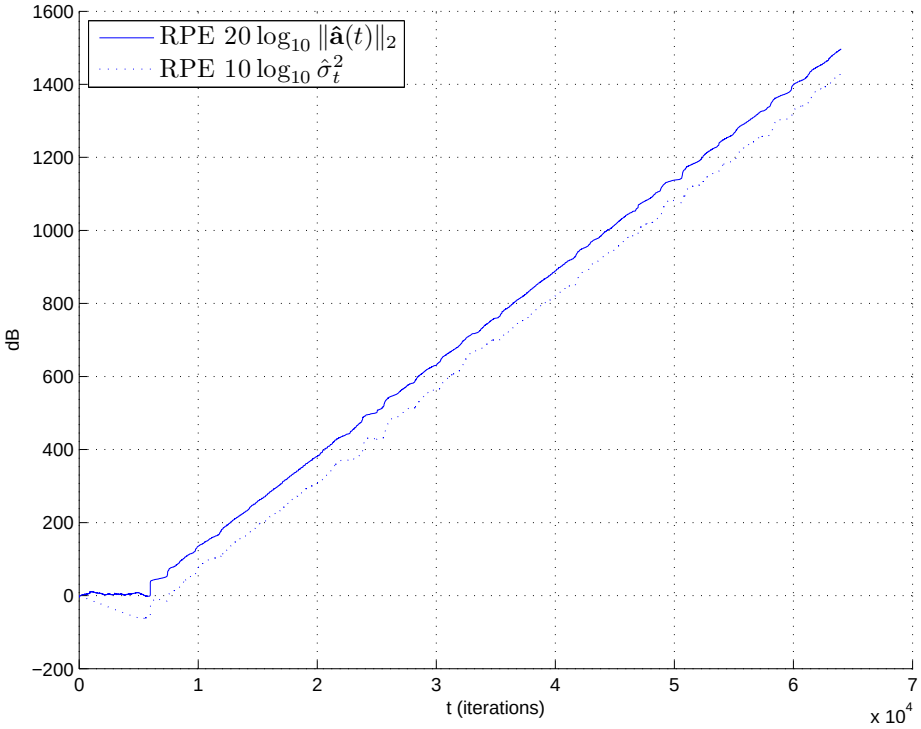
then the RPE algorithm produces **the same solution $\hat{\mathbf{f}}(t)$ as in the case without scaling**.

Consequences of the Scaling Ambiguity

- Consequences:
  - in many RPE simulations, the auxiliary variables undergo an **exponentially increasing scaling**, until numerical overflow occurs
  - this effect renders the RPE algorithm **numerically inaccurate and unstable**
- Example: log-scale plot of the squared norm of two auxiliary variables for an AFC simulation

TVAR coefficients norm: $\|\hat{\mathbf{a}}(t)\|_2^2$

prediction error variance: $\hat{\sigma}_t^2$



Dual Regularization: Regularizing Auxiliary Variables Identification

- Proposed solution: regularize the identification of one of the auxiliary variables, i.e., the TVAR coefficients identification. This approach is denoted as **Dual Regularization**.
- Tikhonov Regularization: includes a minimum-norm constraint in the PE criterion, featuring one regularization parameter β

$$\min_{\hat{\mathbf{f}}(t),\hat{\mathbf{a}}(t)} \frac{1}{2N} \sum_{k=1}^{t} \frac{\lambda^{t-k}}{\hat{\sigma}_k^2} \left\{ \hat{A}(q,t)[y(k) - \hat{F}(q,t)u(k)] \right\}^2 + \beta \|\hat{\mathbf{a}}(t)\|_2^2,$$

- Bayesian Regularization: adds a more flexible regularization term in the PE criterion

$$\min_{\hat{\mathbf{f}}(t),\hat{\mathbf{a}}(t)} \frac{1}{2N} \left\{ \sum_{k=1}^{t} \frac{\lambda^{t-k}}{\hat{\sigma}_k^2} \left\{ \hat{A}(q,t)[y(k) - \hat{F}(q,t)u(k)] \right\}^2 + [\hat{\mathbf{a}}(t) - \mathbf{a_0}]^T \mathbf{R_a}^{-1} [\hat{\mathbf{a}}(t) - \mathbf{a_0}] \right\}$$

which may incorporate prior knowledge on true TVAR process generating the near-end signal:

$$\mathbf{a_0} \triangleq E\{\mathbf{a}\}, \quad \mathbf{R_a} \triangleq \text{cov}\{\mathbf{a}\} = E\left\{ (\mathbf{a} - \mathbf{a_0})(\mathbf{a} - \mathbf{a_0})^T \right\}$$

Combined Primal and Dual Regularization

- Combined Primal and Dual Regularization:

$$\min_{\hat{\mathbf{f}}(t),\hat{\mathbf{a}}(t)} \frac{1}{2N}\Bigg\{\sum_{k=1}^{t}\frac{\lambda^{t-k}}{\hat{\sigma}_k^2}\{\hat{A}(q,t)[y(k)-\hat{F}(q,t)u(k)]\}^2 + \underbrace{[\hat{\mathbf{f}}(t)-\mathbf{f_0}]^T\mathbf{R_f}^{-1}[\hat{\mathbf{f}}(t)-\mathbf{f_0}]}_{\text{primal}} + \underbrace{[\hat{\mathbf{a}}(t)-\mathbf{a_0}]^T\mathbf{R_a}^{-1}[\hat{\mathbf{a}}(t)-\mathbf{a_0}]}_{\text{dual}}\Bigg\}$$

- Special cases [3]:
  - if $\mathbf{a_0} = \mathbf{0}$ the dual regularization corresponds to a generalized, MSE-optimal type of Tikhonov regularization (TR)
  - if $\mathbf{a_0} = \hat{\mathbf{a}}(t-1)$ the dual regularization corresponds to a generalized, MSE-optimal type of Levenberg-Marquardt regularization (LMR)

[3] T. van Waterschoot, G. Rombouts, and M. Moonen, "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement," *Signal Processing*, to appear.

Dually Regularized RPE (DR-RPE) Algorithm

Prediction error: $$\varepsilon[t,\hat{\boldsymbol{\theta}}(t-1)] = \begin{bmatrix}1 & \hat{\mathbf{a}}^T(t-1)\end{bmatrix}\left(\begin{bmatrix}y(t)\\ \mathbf{y}(t)\end{bmatrix} - \begin{bmatrix}\mathbf{u}^T(t)\\ \mathbf{U}^T(t)\end{bmatrix}\hat{\mathbf{f}}(t-1)\right)$$

Prediction error variance: $$\hat{\sigma}_t^2 = \lambda_A\hat{\sigma}_{t-1}^2 + (1-\lambda_A)\varepsilon^2[t,\hat{\boldsymbol{\theta}}(t-1)]$$

Regression vectors: $$\boldsymbol{\psi}_F[t,\hat{\mathbf{a}}(t-1)] = \begin{bmatrix}\mathbf{u}(t) & \mathbf{U}(t)\end{bmatrix}\begin{bmatrix}1\\ \hat{\mathbf{a}}(t-1)\end{bmatrix},\ \boldsymbol{\psi}_A[t,\hat{\mathbf{f}}(t-1)] = \mathbf{U}^T(t)\hat{\mathbf{f}}(t-1) - \mathbf{y}(t)$$

Regularized TVAR correlation matrix update:

[dual] $$\mathbf{R}_A(t) = \lambda_A\mathbf{R}_A(t-1) + \frac{1}{\hat{\sigma}_t^2}\boldsymbol{\psi}_A[t,\hat{\mathbf{f}}(t-1)]\boldsymbol{\psi}_A^T[t,\hat{\mathbf{f}}(t-1)] + (1-\lambda_A)\mathbf{R_a}^{-1}$$

Regularized TVAR and RIR weight updates:

[dual TR] $$\hat{\mathbf{a}}(t) = \hat{\mathbf{a}}(t-1) + \mathbf{R}_A^{-1}(t)\left\{\frac{1}{\hat{\sigma}_t^2}\boldsymbol{\psi}_A[t,\hat{\mathbf{f}}(t-1)]\varepsilon[t,\hat{\boldsymbol{\theta}}(t-1)] - (1-\lambda_A)\mathbf{R_a}^{-1}\hat{\mathbf{a}}(t-1)\right\}$$

[dual LMR] $$\hat{\mathbf{a}}(t) = \hat{\mathbf{a}}(t-1) + \frac{1}{\hat{\sigma}_t^2}\mathbf{R}_A^{-1}(t)\boldsymbol{\psi}_A[t,\hat{\mathbf{f}}(t-1)]\varepsilon[t,\hat{\boldsymbol{\theta}}(t-1)]$$

[primal LMR] $$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \mu_F\frac{\mathbf{R_f}\boldsymbol{\psi}_F[t,\hat{\mathbf{a}}(t-1)]\varepsilon[t,\hat{\boldsymbol{\theta}}(t-1)]}{\boldsymbol{\psi}_F^T[t,\hat{\mathbf{a}}(t-1)]\mathbf{R_f}\boldsymbol{\psi}_F[t,\hat{\mathbf{a}}(t-1)] + \hat{\sigma}_t^2}$$

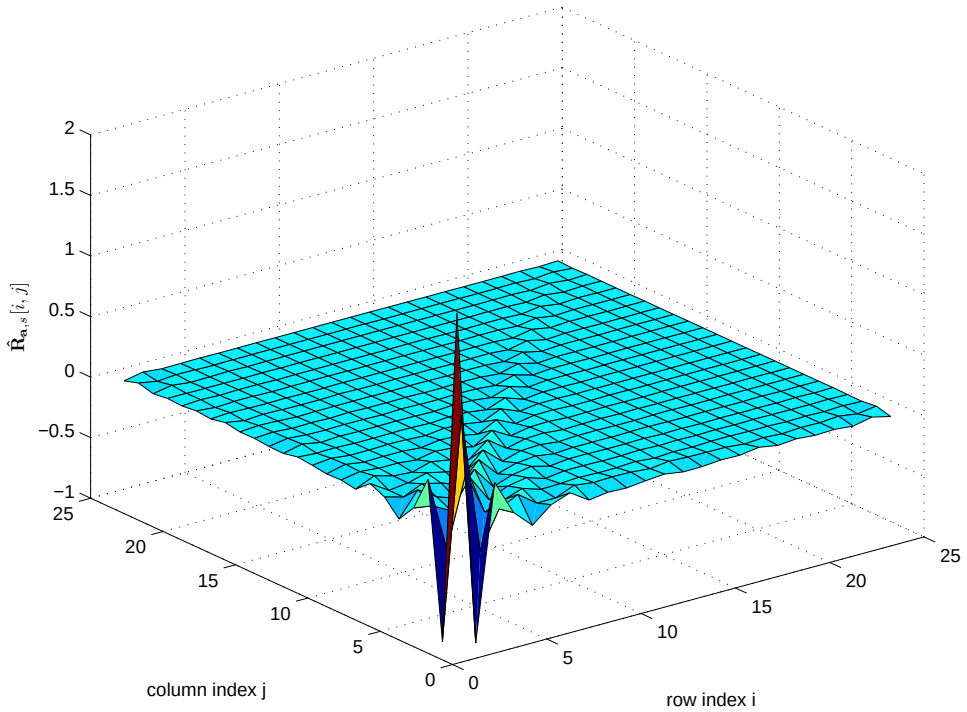
Simulations: Dual Regularization Matrix Construction

How to construct an appropriate dual regularization matrix $\mathbf{R_a}$?

- using no prior knowledge: $\mathbf{R_a} = \mathbf{I}$, which leads to traditional TR/LMR regularization

- using prior knowledge: $\mathbf{R_a} = \hat{\mathbf{R}}_{\mathbf{a},s}$
  - TVAR models $\hat{\mathbf{a}}_i$ of 310929 different 20 ms speech frames from the TIMIT database were identified
  - the corresponding TVAR coefficient vector outer products were ensemble-averaged:

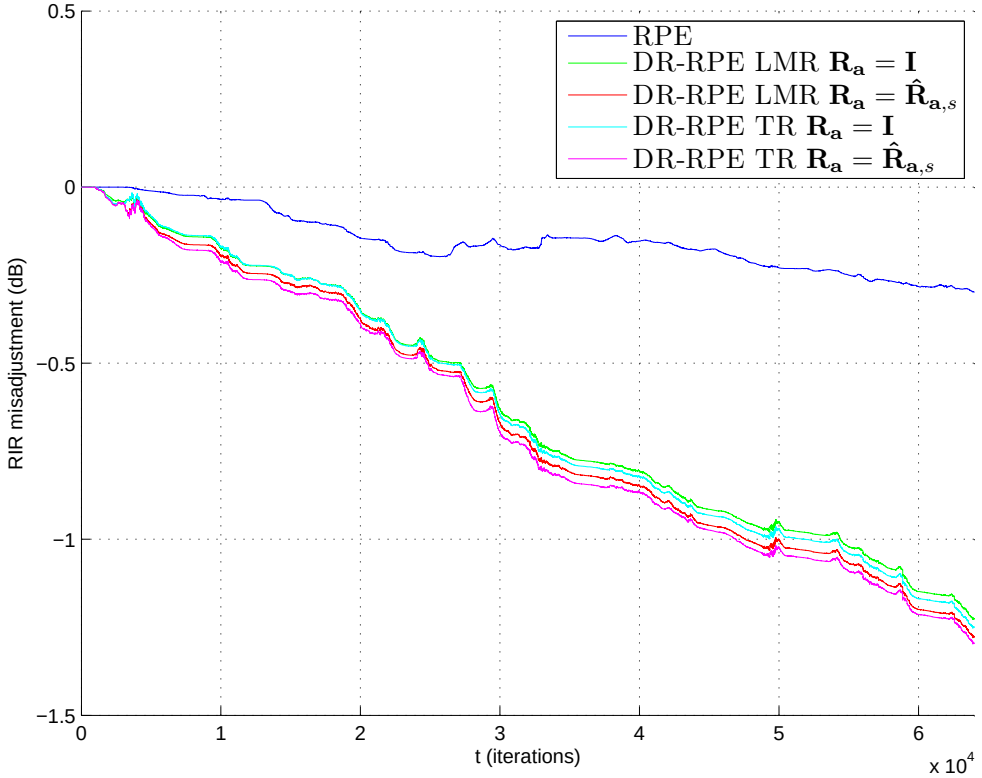
$$\hat{\mathbf{R}}_{\mathbf{a},s} = E\{\hat{\mathbf{a}}_i \hat{\mathbf{a}}_i^T\},$$
$$i = 1, \ldots, 310929$$



Simulation 1: AFC with real near-end speech (1)

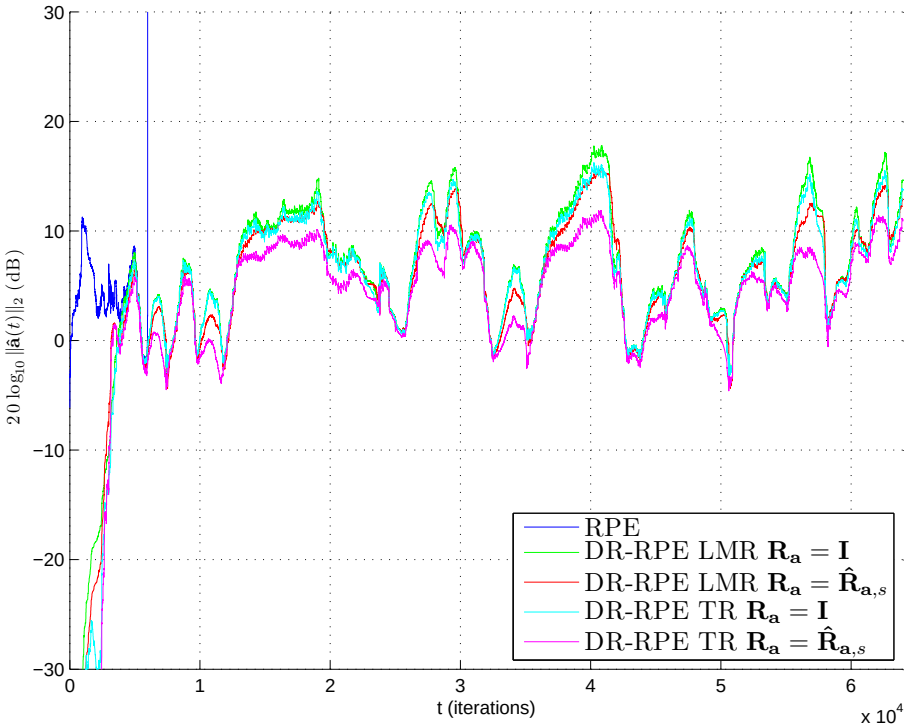
AFC simulation:

- sampling frequency: 16 kHz
- amplifier gain: -9 dB
  (gain margin $= 3$ dB)
- equalizer: pure time delay
  of 320 samples
- near-end signal: 4 s male speech
  (not from TIMIT database!)
- RIR order: $n_F = 2000$
- TVAR model order: $n_A = 24$
- TVAR forgetting factor:
  $\lambda_A = 0.9971$
- step size: $\mu_F = 10^{-3}$ (RPE),
  $\mu_F = 10^{-2}$ (DR-RPE)
- primal regularization
  matrix: $\mathbf{R}_\mathbf{f} = \mathbf{I}$

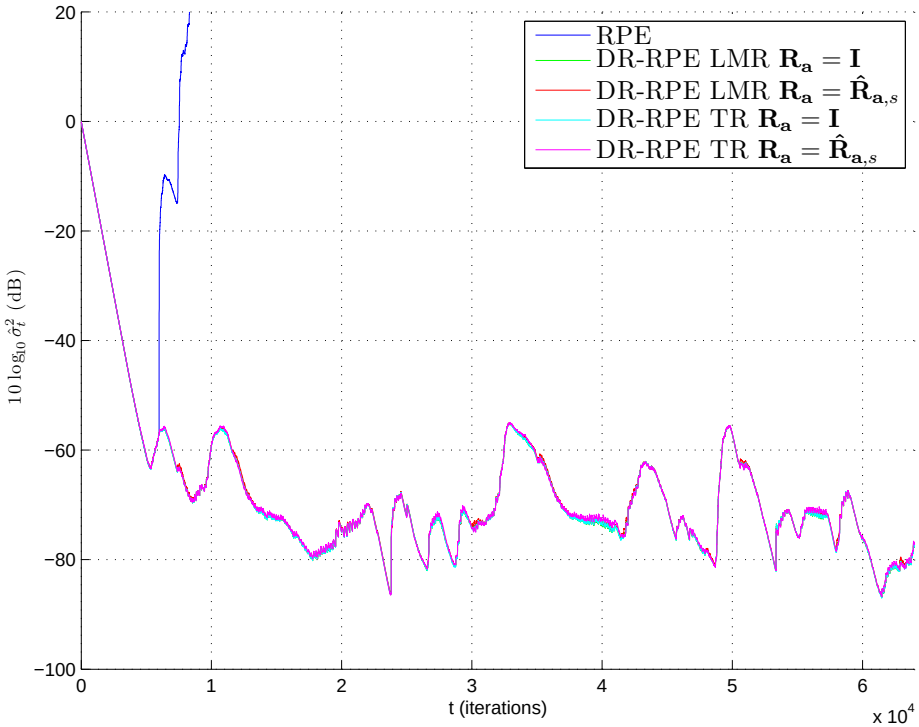


$$\text{RIR misadjustment (dB)} = 20\log_{10}\frac{\|\hat{\mathbf{f}}(t) - \mathbf{f}\|_2}{\|\mathbf{f}\|_2}.$$

Simulation 1: AFC with real near-end speech (2)



TVAR coefficients norm (dB) $= 20 \log_{10} \|\hat{\mathbf{a}}(t)\|_2$



Prediction error variance (dB) $= 10 \log_{10} \hat{\sigma}_t^2$

Simulation 2: AFC with synthetic near-end speech

- Synthetic speech generation: passing GWN through time-varying all-pole filter of order $n_A = 24$, whose coefficients $\mathbf{a}(t)$ change every 20 ms and are computed by linear prediction of real speech
- AFC simulation: same simulation parameters as in Simulation 1, except for near-end signal = 60 s synthetic speech (data record length $M = 960 \cdot 10^3$ samples)
- Results: we compare the accuracy of the TVAR coefficients identification using TR/LMR and with/without prior knowledge

Regularization method	$\mathbf{R_a}$	TVAR misadjustment
Levenberg-Marquardt	$\mathbf{I}$	-1.5322 dB
Levenberg-Marquardt	$\hat{\mathbf{R}}_{\mathbf{a},s}$	-1.9629 dB
Tikhonov	$\mathbf{I}$	-1.9814 dB
Tikhonov	$\hat{\mathbf{R}}_{\mathbf{a},s}$	-3.5738 dB

$$\text{TVAR misadjustment (dB)} = 20 \log_{10} \left(\frac{1}{M} \sum_{k=1}^{M} \frac{\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|_2}{\|\mathbf{a}(k)\|_2} \right)$$

Conclusions

- **Recursive prediction error (RPE) identification** algorithms are attractive alternatives to traditional least-squares-based adaptive filtering algorithms for RIR identification, e.g., to deliver double-talk-robust AEC or unbiased AFC solutions.
- The state-of-the-art stochastic gradient RPE algorithm suffers from a scaling ambiguity in the calculation of the auxiliary variables, which leads to **numerical inaccuracy and instability** being observed in many computer simulations.
- The proposed solution consists in regularizing the identification of the TVAR auxiliary variables, which is denoted as **Dual Regularization**.
- Combining primal and dual regularization, a class of **Dually Regularized RPE (DR-RPE) Algorithms** is obtained, with different choices of regularization methods (Tikhonov/Levenberg-Marquardt) and matrices (possibly incorporating prior knowledge).
- Simulation results confirm that, regardless of the choice of regularization method or matrix, the scaling ambiguity is solved by using the DR-RPE algorithm in an AFC application with real near-end speech, and the **RIR misadjustment is improved**.
- AFC simulations with synthetic near-end speech moreover illustrate that the **accuracy of the TVAR coefficients identification is increased** by using prior knowledge, and that the Tikhonov method is preferred over the Levenberg-Marquardt method.