DATA-SELECTIVE LMS-NEWTON AND LMS-QUASI-NEWTON ALGORITHMS

Christos Tsinos* and Paulo S. R. Diniz**

*University of Luxembourg
**Federal University of Rio de Janeiro

ICASSP 2019

14/05/2019
Outline

• Introduction
• System Model
• LMS-Newton and LMS-Quasi-Newton Algorithms
• Data-Selective Approaches
• Simulations
Introduction (1/3)

- In the era of big data, the processing will demand huge computational load if an effective strategy is not followed
- Data-selective processing
  - Process only the innovative new data
  - May avoid outliers
  - Performance close to the one of the non-data-selective counterparts
  - Reduced computational burden since only a very small portion of the data is processed
Introduction (2/3)

• This paper develops data-selective versions of
  – LMS-Newton (LMSN)
  – LMS-Quasi-Newton (LMSQN)

• LMSN/LMSQN are powerful alternatives to the classical LMS
  – Higher Complexity
  – Better Performance in several cases (e.g. when the spread of the eigenvalues of the input-signal correlation matrix is large)
  – Some versions of LMSQN appear to be very robust to quantization errors compared to algorithms of similar complexity/performance, i.e., RLS.
• The data are classified via two thresholds as
  – Non-innovative
  – Innovative
  – Outliers

• The thresholds are tuned based on a prescribed probability of update

• The latter probability is connected to the Mean Square Error (MSE) of the algorithms

• The performance is evaluated via simulations on synthetic and real world data
System Model (1/2)

• Linear System Identification Problem
• Input-Output

\[ d(k) = w_o^T x(k) + n(k) \]

• \( w_o \in \mathbb{R}^{L+1} \) is the unknown system
• \( x(k) = [x(k) \ x(k-1) \ldots \ x(k-L+1)]^T \) is the input signal
• \( n(k) \) is a Gaussian noise sample of variance \( \sigma_n^2 \)
• A filtering algorithm generates an output signal estimation via \( w^T(k)x(k) \)
System Model (2/2)

• Error Estimation Sequence for $k = 0, 1, \ldots, \infty$

\[ e(k) = d(k) - \mathbf{w}^T(k)\mathbf{x}(k) \]

• Mean Square Error

\[ \xi(k) = \sigma_n^2 + \mathbb{E}\{\Delta\mathbf{w}^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\Delta\mathbf{w}(k)\} \]
\[ = \sigma_n^2 + \xi_{exc}(k), \]

• $\Delta\mathbf{w}(k) = \mathbf{w}(k) - \mathbf{w}_o$

• The MSE formula is used to prescribe the desired probability of update
LMSN/LMSQN Algorithms (1/3)

• The aim is to minimize the cost function

\[ J(w(k)) = \frac{1}{2} |e(k)|^2 \]

• Update Step

\[ w(k) = w(k-1) + \frac{\mu}{x^T(k)\hat{R}^{-1}(k)x(k)} \hat{R}^{-1}(k)x(k)\tilde{e}(k) \]

• \( \mu \) is a step-size parameter

• \( \tilde{e}(k) = d(k) - w^T(k-1)x(k) \) is the a priori estimation error

• \( \hat{R}(k) \) is the estimation of \( R = \mathbb{E}\{x(k)x^T(k)\} \)
LMSN/LMSQN Algorithms (2/3)

- LMSN and LMSQN differ on how $\hat{R}^{-1}(k)$ is estimated.
- LMSN estimation is based on a Robbins-Monro procedure:
  \[
  \hat{R}^{-1}(k) = \frac{1}{1 - \alpha} \left\{ \hat{R}^{-1}(k - 1) - \frac{\hat{R}^{-1}(k - 1)x(k)x^T(k)\hat{R}^{-1}(k - 1)}{1 - \alpha} + \frac{x(k)^T\hat{R}^{-1}(k)x(k)}{\alpha} \right\}
  \]
- $\alpha$ is a step-size parameter.
For the LMSQN the estimation is given by

\[
\hat{R}^{-1}(k) = \frac{1}{1 - \alpha} \left\{ \hat{R}^{-1}(k - 1) + \left( \frac{\mu}{2x(k)^T \hat{R}^{-1}(k)x(k)} - 1 \right) \right. \\
\left. \times \frac{\hat{R}^{-1}(k - 1)x(k)x^T(k)\hat{R}^{-1}(k - 1)}{x(k)^T \hat{R}^{-1}(k)x(k)} \right\}.
\]


Data-Selective Approaches (1/5)

• New data are classified as innovative if \( |e(k)|^2 \) is greater than a scaled noise power level \( \tau(k) \sigma_n^2 \).

• If \( |e(k)|^2 \) is greater than \( \tau_{max} \sigma_n^2 \), an outlier is identified and no update is performed.

• Equivalent cost function

\[
J'(w(k)) = \begin{cases} 
\frac{1}{2} |e(k)|^2, & \text{if } \sqrt{\tau(k)} \leq \frac{|e(k)|}{\sigma_n} < \sqrt{\tau_{max}} \\
0, & \text{otherwise.}
\end{cases}
\]
Data-Selective Approaches (2/5)

• Update for the data-selective approach

\[ w(k) = \begin{cases} 
  w(k-1) + \mu \frac{\hat{R}^{-1}(k)x(k)e(k)}{x^T(k)\hat{R}^{-1}(k)x(k)}, & \sqrt{\tau(k)} \leq \frac{|e(k)|}{\sigma_n} < \sqrt{\tau_{max}} \\
  w(k-1), & \text{otherwise.} 
\end{cases} \]

• The data-selective strategy may be adopted for the update of \( \hat{R}^{-1} \), as well.

• Desired probability of update

\[ P_{up}(k) = P \left\{ \frac{|e(k)|}{\sigma_n} > \sqrt{\tau(k)} \right\} - P \left\{ \frac{|e(k)|}{\sigma_n} > \sqrt{\tau_{max}} \right\} \]
Data-Selective Approaches (3/5)

• Under the assumption of white Gaussian input signals, at the steady state we have

\[ P_{up} = 2Q \left( \frac{\sigma_n \sqrt{T}}{\sigma_e} \right) - 2Q \left( \frac{\sigma_n \sqrt{T_{max}}}{\sigma_e} \right) \]

• \( Q(\cdot) \) is the complementary Gaussian cumulative distribution function

• \( \sigma_e^2 \) is the error signal variance

• Index \( k \) is dropped under the assumption of stationarity
Data-Selective Approaches (4/5)

- **Proposition**: The excess mean square error at the steady-state can be approximated by

\[ \xi_{exc}(\infty) = \frac{\mu P_{up}}{2 - \mu P_{up}} \sigma_n^2 \]

- If no outliers are presented, the threshold is

\[ \sqrt{\tau} = \sqrt{1 + \beta Q^{-1}(0.5 P_{up})} \]

\[ \beta = \frac{\mu P_{up}}{2 - \mu P_{up}} \]

- For the case of outliers, some prior information of the signal sources and supporting circuitry, is needed for deriving the thresholds

Algorithm 1: Data-selective LMSN and LMSQN Algorithms

1: Inputs: $0 < \mu \leq 1, 0 < \alpha \leq 1$ (for LMSN), $\gamma$ small positive value, $P_{up}$ and $\tau_{max}$
2: Initialize $w(0) = 0_{L+1}$ and $\hat{R}^{-1}(0) = \gamma I_{L+1}$
3: Set $\beta = \frac{\mu P_{up}}{2 - \mu P_{up}}$
4: Calculate $\tau$ from (13), if outliers are present or from (16), otherwise
5: for $k = 1, 2, \ldots$ do
6: Acquire $x(k)$ and $d(k)$
7: $e(k) = d(k) - w^T(k)x(k)$
8: if $\sqrt{\tau \sigma_n} \leq |e(k)| \leq \sqrt{\tau_{max} \sigma_n}$ then
9: $t(k) \leftarrow \hat{R}^{-1}(k)x(k)$
10: $\psi(k) \leftarrow x^T(k)t(k)$
11: $w(k + 1) \leftarrow w(k) + \mu \frac{t(k)e(k)}{\psi(k)}$
12: $\hat{R}^{-1}(k + 1) \leftarrow \frac{1}{1 - \alpha} \left[ \hat{R}^{-1}(k) - \frac{t(k)t^T(k)}{1 - \alpha + \psi(k)} \right]$, for LMSN
13: else if $|e(k)| \leq \sqrt{\tau \sigma_n}$ then
14: $w(k + 1) \leftarrow w(k)$
15: else if $|e(k)| \geq \sqrt{\tau_{max} \sigma_n}$ then
16: $w(k + 1) \leftarrow w(k), e(k) = 0, d(k) = 0$
17: end if
18: end for
Simulations (1/5)

• System identification problem

\[
\begin{bmatrix}
0.1010 & 0.3030 & 0 \\
0.2020 & -0.4040 & -0.7071 \\
-0.4040 & 0.2020 & -0.7071
\end{bmatrix}.
\]

• Input signals

\[
x(k) = 0.88x(k - 1) + n_1(k),
\]
\[
x(k) = -0.55x(k - 1) - 1.221x(k - 2) - 0.49955x(k - 3)
- 0.4536x(k - 1) + n_2(k),
\]

• \(n_1(k)\) and \(n_2(k)\) are uncorrelated Gaussian noise variables
Simulations (2/5)

MSE Learning Curves - No Outliers - $P_{up} = 0.4$
Simulations (3/5)

(c) AR(1) + outliers
(d) AR(4) + outliers

MSE Learning Curves - Outliers - $P_{up} = 0.4$
Simulations (4/5)

Fig. 2. Comparison between prescribed $P_{up}$ and the achieved $\hat{P}_{up}^{LMSN}$ and $\hat{P}_{up}^{LMSQN}$ by the data-selective LMSN and LMSQN algorithms.
Simulations (5/5)

• Temperature prediction on a data-set provided by University of California at Irvine

• The prediction error variance $\sigma_e^2$ is derived according to the setup - $P_{up} = 0.4$

Conclusion

• Data Selective LMSN and LMSQN algorithms were developed
• Computational overhead reduction
• Outlier Exclusion from the learning process
• Performance evaluated via simulations on synthetic and real world data
• Extensions on distributed adaptive filtering under development
Thank you for your attention