l_0-norm Feature LMS algorithms

Hamed Yazdanpanah, José A. Apolinário Jr., Paulo S. R. Diniz, and Markus V. S. Lima
{hamed.yazdanpanah,markus.lima,diniz}@smt.ufrj.br and apolin@ime.eb.br
IME
UFRJ

Abstract

Feature LMS algorithms, by applying a feature matrix to the coefficients vector, can detect and exploit sparsity in linear combinations of filter coefficients (hidden sparsity). In many cases the unknown plant to be identified may contain not only hidden but also plain sparsity $\cdots$ we here use the l_0-norm, as a sparsity-promoting techniques, to the F-LMS algorithm. Experimental results show that the proposed algorithm outperforms (faster convergence rate) the F-LMS algorithm when dealing with hidden sparsity, particularly for highly sparse systems.

Introduction

- LMS: the most popular algorithm since the year I was born, but it may be improved for sparse systems
- Exploiting signals and systems sparsity can improve steady-state MSE, convergence rate, etc.
- Recently introduced (ICASSP 2018), the F-LMS exploits hidden sparsity
- The F-LMS algorithm is not able to exploit plain sparsity, sometimes observed along with hidden sparsity

l_0-norm F-LMS algorithms

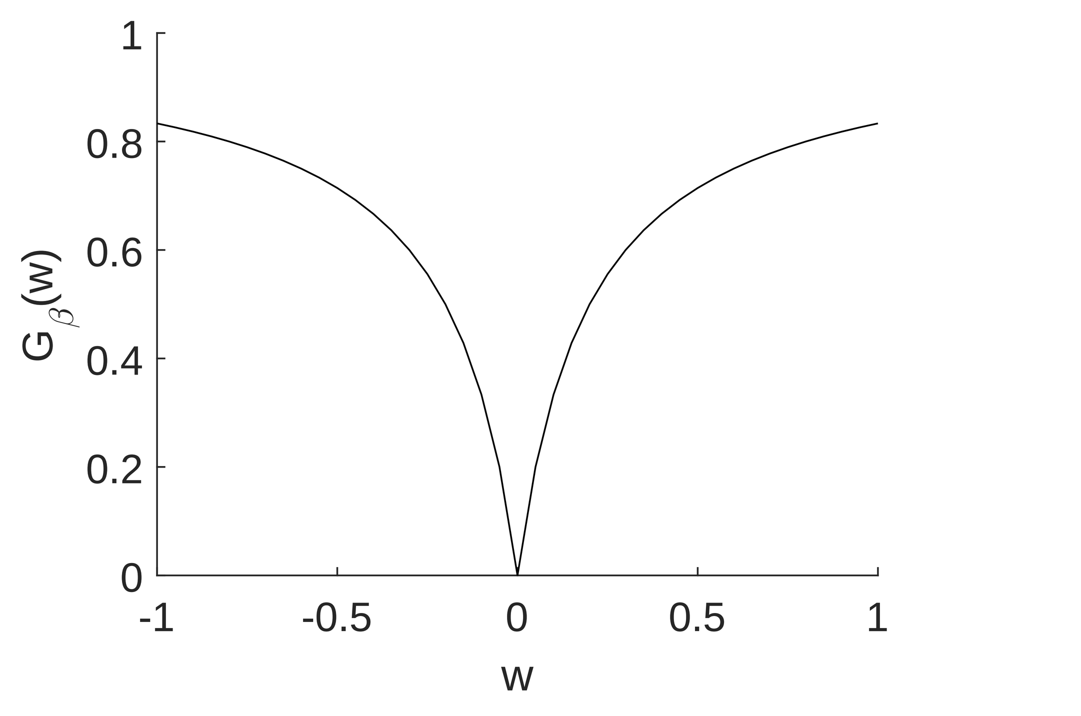
- **The objective function** ($\zeta_{l_0\text{-F-LMS}}$):

$$\underbrace{\frac{1}{2}|e(k)|^2}_{\text{LMS term}} + \underbrace{\alpha\|\mathbf{F}(k)\mathbf{w}(k)\|_1}_{\text{feature-inducing}} + \underbrace{\lambda\|\mathbf{w}(k)\|_0}_{\text{plain sparsity}}$$

- **Feature matrix:** $\mathbf{F}(k)$, utilized for exposing the hidden sparsity, is such that $\mathbf{F}(k)\mathbf{w}(k)$ becomes a sparse vector. Practical selections of $\mathbf{F}(k)$ must be based on some *a priori* information about the unknown system.
- **The plain sparsity promoting term** is non-differentiable and is replaced by $G_\beta(\mathbf{w}) \triangleq \sum_{i=0}^{N}\left(1 - \frac{1}{1+\beta|w_i|}\right)$, β trading off smoothness and quality of approximation.

l_0-norm F-LMS algorithms

- **The gradient** $\nabla G_\beta(\mathbf{w}) = \mathbf{g}_\beta(\mathbf{w})$ becomes $[g_\beta(w_0) \ \cdots \ g_\beta(w_N)]^T$, with $g_\beta(w_i) = \frac{\beta \text{sgn}(w_i)}{(1+\beta|w_i|)^2}$ (curve with $\beta = 5$).



- **The recursive expression** of the l_0-F-LMS algorithm is then obtained:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e(k)\mathbf{x}(k) - \mu(\alpha\mathbf{p}(k) + \lambda\mathbf{g}_\beta(\mathbf{w}(k))),$$

where $0 < \alpha$, $\lambda \ll 1$, μ is the step-size and $\mathbf{p}(k) \in \mathbb{R}^{N+1}$ is the gradient of $\|\mathbf{F}(k)\mathbf{w}(k)\|_1$.

Sparse lowpass systems

- The l_0-F-LMS Algorithm for sparse lowpass systems
- Unknown system has lowpass narrowband spectrum $\Rightarrow$ its impulse response is smooth $\Rightarrow$ the difference between adjacent coefficients is small
- Let $\mathbf{F} \in \mathbb{R}^{N\times(N+1)} = \begin{bmatrix} 1 & -1 & & \mathbf{0} \\ & \ddots & \ddots & \\ \mathbf{0} & & 1 & -1 \end{bmatrix}$

  $\Rightarrow$ **$\mathbf{Fw}(k)$ is a sparse vector**
- Therefore, $\mathbf{p}(k) = [p_0(k) \ \cdots \ p_N(k)]^T$ is given by

$$p_i(k) = \begin{cases} \text{sgn}(w_0(k) - w_1(k)) & \text{if } i = 0, \\ -\text{sgn}(w_{i-1}(k) - w_i(k)) & \\ +\text{sgn}(w_i(k) - w_{i+1}(k)) & \\ & \text{if } i = 1, \cdots, N-1, \\ -\text{sgn}(w_{N-1}(k) - w_N(k)) & \text{if } i = N. \end{cases}$$

Sparse highpass systems

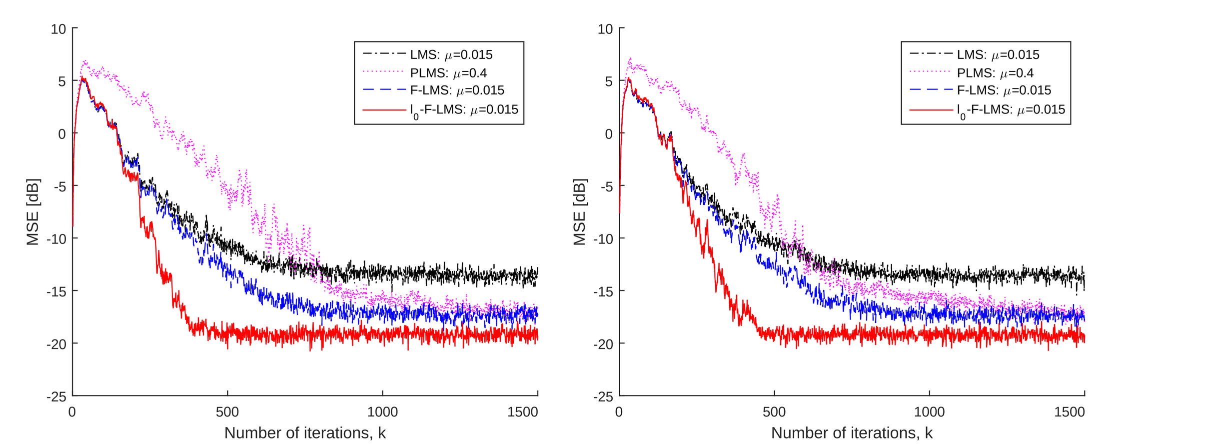
- The l_0-F-LMS Algorithm for sparse highpass systems
- Unknown system has highpass narrowband spectrum $\Rightarrow$ adjacent coefficients have similar absolute values, but with opposite signs $\Rightarrow$ the sum of adjacent coefficients is small
- Let $\mathbf{F} \in \mathbb{R}^{N\times(N+1)} = \begin{bmatrix} 1 & 1 & & \mathbf{0} \\ & \ddots & \ddots & \\ \mathbf{0} & & 1 & 1 \end{bmatrix}$

  $\Rightarrow$ **$\mathbf{Fw}(k)$ is a sparse vector**
- Therefore, $\mathbf{p}(k) = [p_0(k) \ \cdots \ p_N(k)]^{\mathrm{T}}$ is given by

$$p_i(k) = \begin{cases} \text{sgn}(w_0(k) + w_1(k)) & \text{if } i = 0, \\ \text{sgn}(w_{i-1}(k) + w_i(k)) & \\ +\text{sgn}(w_i(k) + w_{i+1}(k)) & \\ & \text{if } i = 1, \cdots, N-1, \\ \text{sgn}(w_{N-1}(k) + w_N(k)) & \text{if } i = N. \end{cases}$$

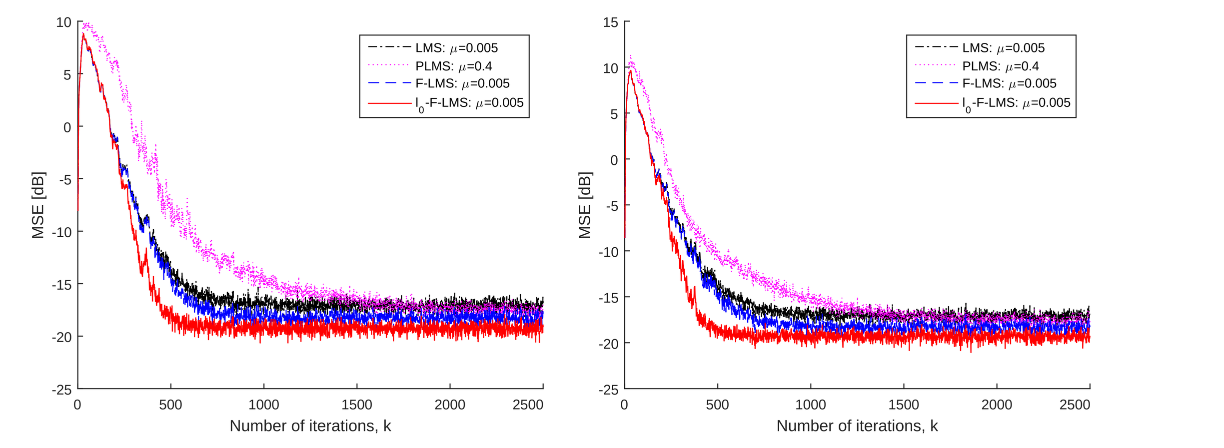
Simulations

- Algorithms tested: LMS, proportionate LMS (PLMS), F-LMS, and l_0-F-LMS
- Input signals (SNR=20dB): $x(k) \sim \mathcal{N}(0,1)$ and $x(k)$ correlated signal ($\lambda_{max}/\lambda_{min} = 20$)
- Filter order: $N = 99$ (100 coefficients) initialized as $\mathbf{w}(0) = [0, \cdots, 0]^{\mathrm{T}}$
- Constants: $\beta = 20$, $\alpha = 0.05$, and $\lambda = 0.005$
- Unknown block sparse lowpass system: $\mathbf{w}_{o,l} = [0.4, 0.4, \cdots, 0.4, \mathbf{0}_{1\times70}]^{\mathrm{T}}$
- Unknown block sparse highpass system: $\mathbf{w}_{o,h} = [0.4, -0.4, 0.4 \cdots, -0.4, \mathbf{0}_{1\times70}]^{\mathrm{T}}$
- $\mathbf{w}_{o,l}^2$ (refers to an upsampled by 2 version) and $\mathbf{w}_{o,l}^4$ (refers to an upsampled by 4 version)

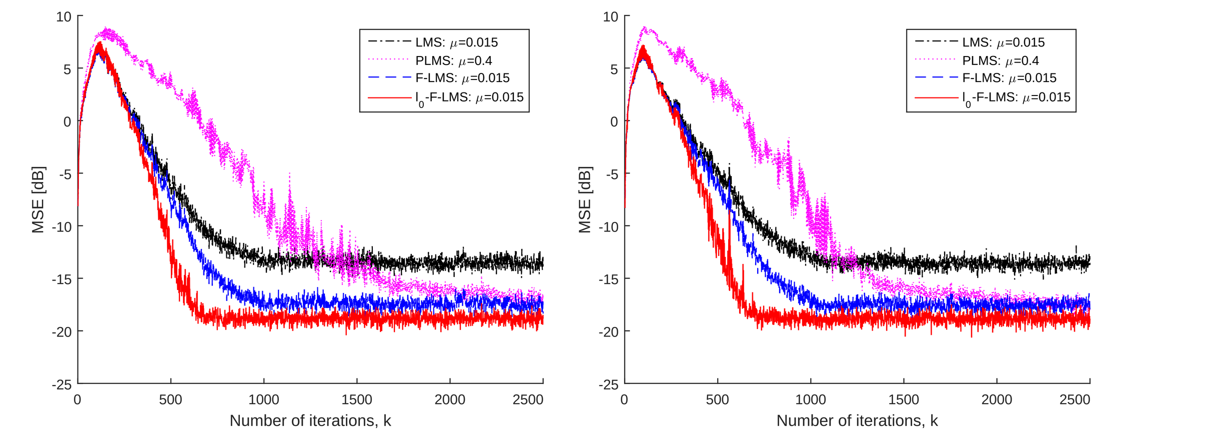


Learning curves of algorithms under test, considering: (1, left) $\mathbf{w}_{o,l}$ and $\mathbf{p}_l(k)$ (uncorrelated input when not mentioned explicitly); (2) $\mathbf{w}_{o,h}$ and $\mathbf{p}_h(k)$.

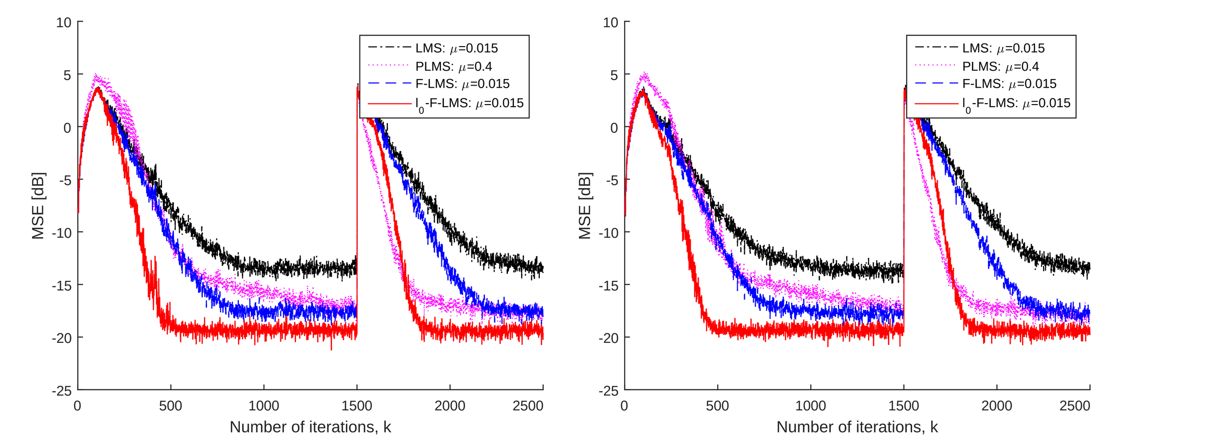
Simulations



Learning curves: (1, left) $\mathbf{w}_{o,l}$, $\mathbf{p}_l(k)$, and correlated input signal; (4) $\mathbf{w}_{o,h}$, $\mathbf{p}_h(k)$, and correlated input signal.



Learning curves: (1, left) $\mathbf{w}_{o,l}^2$ and $\mathbf{p}_l^2(k)$; (2) $\mathbf{w}_{o,h}^2$ and $\mathbf{p}_h^2(k)$.



Learning curves: (1, left) $\mathbf{w}_{o,l}^4$ and $\mathbf{p}_l^4(k)$; (2) $\mathbf{w}_{o,h}^4$ and $\mathbf{p}_h^4(k)$.

Conclusions

- We generalize the F-LMS algorithm and propose the l_0-F-LMS algorithm so that it can exploit hidden and plain sparsity in unknown systems
- Hidden sparsity is disclosed by applying the feature matrix while using the l_1-norm promotes a sparse vector
- Whenever we know the feature of system, this strategy can be utilized in the presence of hidden and plain sparsity
- Simulation results corroborate the superiority of the l_0-F-LMS algorithm over the F-LMS, the PLMS, and the LMS algorithms