Feature Affine Projection Algorithms

Hamed Yazdanpanah

University of São Paulo (USP)
Institute of Mathematics and Statistics (IME)
Department of Computer Science

{hamed.yazdanpanah}@smt.ufrj.br

IME-USP

International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2020)

Presentation Outline

Outline

Motivations

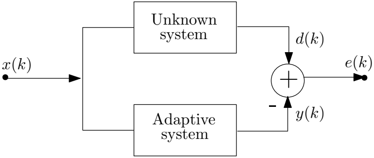


Figure: System identification problem.

We want to exploit hidden sparsity to improve the convergence rate and/or MSE steady-state

- When we have some *a priori* knowledge about the unknown system $\mathbf{w}_*$
  - We can exploit it for accelerating the convergence rate or MSE.
- When we want to obtain an estimate of the unknown system $\mathbf{w}_*$ such that a determined characteristic for the estimate is desirable
  - Lowpass, highpass, linear phase, etc.

Proposal

- Affine Projection (AP) algorithm ⇒ employ data-reuse strategy to improve the convergence rate when the input signal is correlated

- Feature Affine Projection (F-AP) algorithms ⇒ impose some structure on the system's coefficients ⇒ exploit hidden sparsity in system, such as sparsity in linear combination of coefficients
- In this paper, we present the F-AP algorithm in its general form. Also,
  - We discuss the stability of the F-AP algorithm
  - We propose the F-AP algorithm for unknown systems with lowpass narrowband spectrum
  - We propose the F-AP algorithm for unknown systems with highpass narrowband spectrum

Outline

F-AP Algorithm: Problem and Solution

- Problem:

$$\underset{\mathbf{w}(k+1)}{\text{minimize}} \; \underbrace{\|\mathbf{w}(k+1)-\mathbf{w}(k)\|^2}_{\text{standard AP term}} + \underbrace{\alpha\mathcal{P}(\mathbf{F}(k)\mathbf{w}(k+1))}_{\text{feature-inducing term}},$$
$$\text{subject to } \mathbf{d}(k) = \mathbf{X}^T(k)\mathbf{w}(k+1),$$

  - $\mathcal{P}(\cdot)$ is the sparsity promoting penalty function;
  - $\mathbf{F}(k)$ is the feature matrix that takes the unknown system to a sparse vector;
  - $\mathbf{X}(k) \in \mathbb{R}^{(N+1)\times(L+1)}$ and $\mathbf{d}(k) \in \mathbb{R}^{L+1}$ are the input matrix and the desired vector, respectively, and L is the data-reuse factor;

- Solution (using the method of Lagrange multipliers):

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k) + \frac{\mu\alpha}{2}\Big[\mathbf{X}(k)\mathbf{S}(k)\mathbf{X}^T(k) - \mathbf{I}_{N+1}\Big]\mathbf{p}(k),$$

  - $\mathbf{p}(k) \in \mathbb{R}^{N+1}$ is the gradient of function $\mathcal{P}(\mathbf{F}(k)\mathbf{w}(k))$;
  - $\mathbf{S}(k) \triangleq [\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}_{L+1}]^{-1}$;
  - μ and δ are the step-size and the regularization parameters, respectively;
  - $\mathbf{I}_{L+1}$ is the identity matrix of dimension $L+1$;

Stability Analysis: The Valid Ranges for μ and α

Theorem

Suppose that $\mathbf{w}_*$ is the unknown system to be identified. Let us denote the discrepancy between the adaptive filter and the optimal solution by $\widetilde{\mathbf{w}}(k) \triangleq \mathbf{w}_* - \mathbf{w}(k)$. In the noiseless environment, if we adopt $\mu \in (0,2)$, then we obtain $\|\widetilde{\mathbf{w}}(k+1)\|^2 < \|\widetilde{\mathbf{w}}(k)\|^2$; i.e., the convergence of the F-AP algorithm is guaranteed.

Theorem

For the F-AP algorithm, there exists $\alpha \in \mathbb{R}_+$ such that the sequence $\{\|\mathbf{w}_* - \mathbf{w}(k)\|\}_k$ is monotonically decreasing.

Remark

In practice, we choose $\mu \in (0,1)$. Also, we adopt α as small positive value; if the algorithm diverges, we reduce α.

Outline

F-AP Algorithm for Lowpass Systems

- Unknown system has lowpass narrowband spectrum $\Rightarrow$ its impulse response is smooth $\Rightarrow$ the difference between adjacent coefficients is small
- Choose $\mathbf{F}(k)$ to be the time-invariant $\mathbf{F}_l \in \mathbb{R}^{N\times(N+1)}$ as

$$\mathbf{F}_l \triangleq \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \Rightarrow \mathbf{F}_l\mathbf{w}(k) \text{ is a sparse vector}$$

- Adopt the l_1-norm as the sparsity-promoting penalty function $\mathcal{P}$
- Therefore, $\mathbf{p}(k) = \mathbf{p}_l(k) = [p_{l_0}(k) \ \cdots \ p_{l_N}(k)]^T$ is given by

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

F-AP Algorithm for Highpass Systems

- Unknown system has highpass narrowband spectrum $\Rightarrow$ adjacent coefficients have similar absolute values, but with opposite signs
- Choose $\mathbf{F}(k)$ to be the time-invariant $\mathbf{F}_h \in \mathbb{R}^{N\times(N+1)}$ as

$$\mathbf{F}_h \triangleq \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & 0 & \cdots & 1 & 1 \end{bmatrix} \Rightarrow \mathbf{F}_h\mathbf{w}(k) \text{ is a sparse vector}$$

- Adopt the l_1-norm as the sparsity-promoting penalty function $\mathcal{P}$
- Therefore, $\mathbf{p}(k) = \mathbf{p}_h(k) = [p_{h_0}(k) \ \cdots \ p_{h_N}(k)]^T$ is given by

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

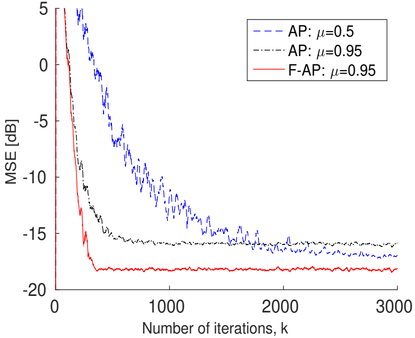
Outline

Scenario: System Identification

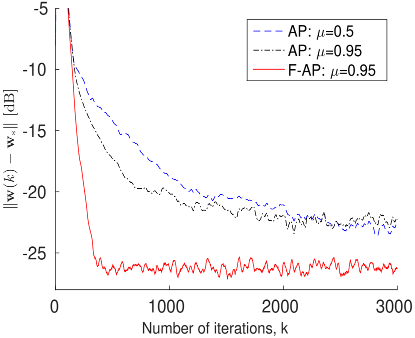
- Algorithms: AP and F-AP
- Input signal: a correlated signal with the eigenvalue spread equal to 20
- System order: $N = 39$, i.e., 40 coefficients
- $\mathbf{w}(0) = [0, \cdots, 0]^T$
- Data-reuse factor: $L = 2$
- $\alpha = 0.01$
- $\delta = 10^{-9}$
- SNR: 20 dB

F-AP Algorithm: Narrowband Lowpass System

- Narrowband lowpass system: $\mathbf{w}_*^l = [0.5\ 0.5\ \cdots\ 0.5]^T$



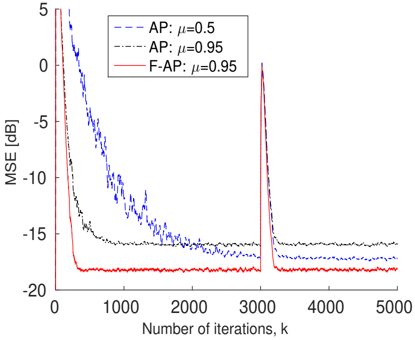
(a) MSE learning curves



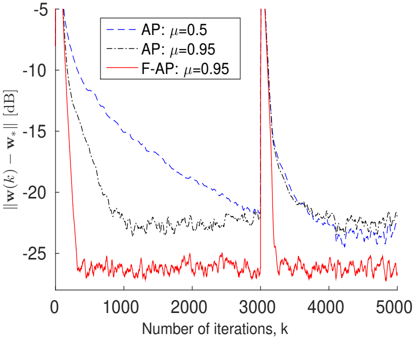
(b) Misalignment curves

F-AP Algorithm: Time-Varying Lowpass System

- $\mathbf{w}_*^l = [0.5\ 0.5\ \cdots\ 0.5]^T$, for $k \leq 3000$
- $\mathbf{w}_*^l = [0.7\ 0.7\ \cdots\ 0.7]^T$, for $k > 3000$



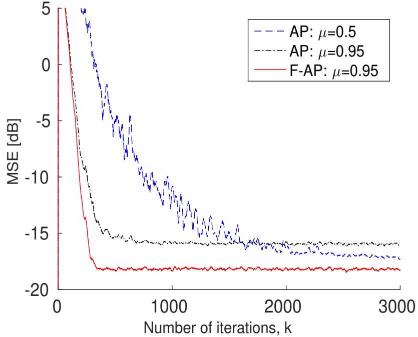
(c) MSE learning curves



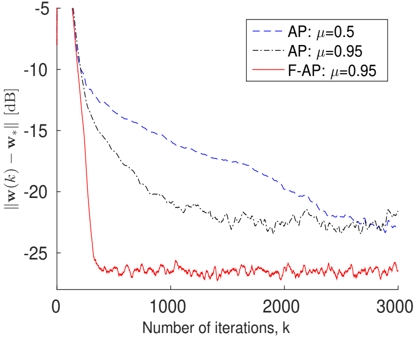
(d) Misalignment curves

F-AP Algorithm: Narrowband Highpass System

- Narrowband highpass system: $\mathbf{w}_*^h = [0.5 \; -0.5 \; 0.5 \; \cdots \; -0.5]^T$



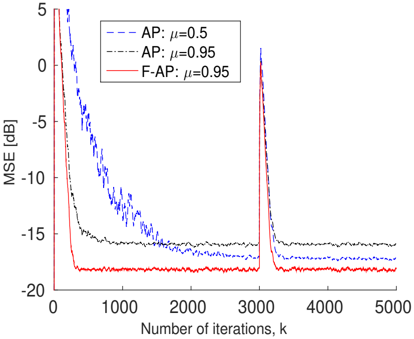
(e) MSE learning curves



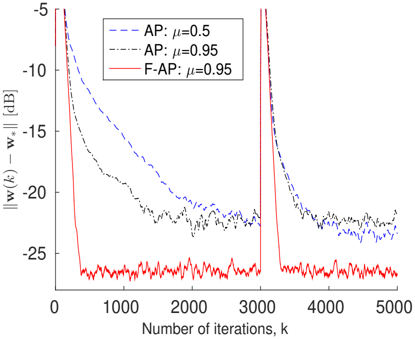
(f) Misalignment curves

F-AP Algorithm: Time-Varying Highpass System

- $\mathbf{w}_*^h = [0.5 \;\; -0.5 \; 0.5 \; \cdots \; -0.5]^T$, for $k \leq 3000$
- $\mathbf{w}_*^h = [0.7 \;\; -0.7 \; 0.7 \; \cdots \; -0.7]^T$, for $k > 3000$



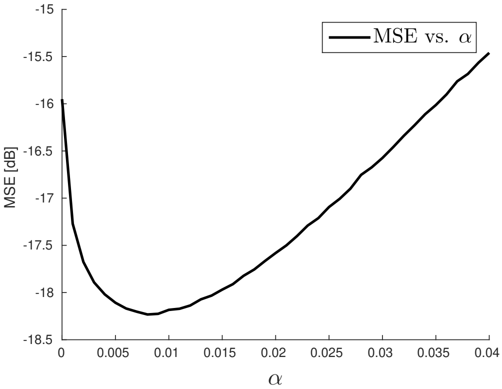
(g) MSE learning curves



(h) Misalignment curves

The MSE of The F-AP Algorithm Versus α

- $\mathbf{w}_*^l = [0.5\ 0.5\ \cdots\ 0.5]^T$



(i) The MSE of the F-AP algorithm versus α considering the lowpass system $\mathbf{w}_*^l$.

Outline

Conclusions

- In this presentation:
  - We exploit the feature of unknown systems in order to improve the convergence rate and the steady-state MSE by the F-AP algorithm
  - We have presented some examples of the F-AP algorithm for exploiting the lowpass and highpass characteristics of unknown systems
  - We have analyzed some properties of the proposed algorithm such as the step-size and the weight given to the sparsity-promoting penalty function
  - The numerical results corroborate the superiority of the F-AP algorithm as compared to the AP algorithm

Thank You!

The author would like to thank the São Paulo Research Foundation (FAPESP) grants #2015/22308-2 and #2019/06280-1.