Computational analysis of a fast algorithm for high-order sparse linear prediction

August 31, 2016

T. L. Jensen, D. Giacobello, T. van Waterschoot and M. G. Christensen

Dept. of Electronic Systems

Aalborg University

AALBORG UNIVERSITY
DENMARK

Outline

- Linear prediction
- Solving a high-order sparse linear prediction problem
- Solving positive-definite Toeplitz systems
- Simulations and timings

Principles of linear prediction

- A stationary set of samples (possible speech) $x[t]$, for $t = 1, \ldots, T$, are written as a linear combination of N past samples

$$x[t] = \sum_{n=1}^{N} \alpha_n x[t-n] + r[t], \tag{1}$$

- $\{\alpha_n\}$ are the prediction coefficients and $r[t]$ is the prediction error.
- Matrix formulation (certain boundary conditions)

$$x = X\alpha + r \tag{2}$$

- Find the prediction coefficients via

$$\underset{\alpha}{\text{minimize}} \quad \|x - X\alpha\|_p^p \tag{3}$$

Conventional linear prediction

- Select $p = 2$

$$\underset{\alpha}{\text{minimize}} \quad \|x - X\alpha\|_2^2 \tag{4}$$

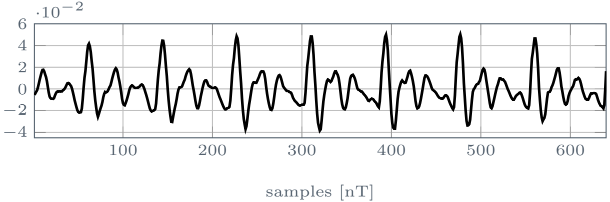
- Solution satisfying the normal equation

$$X^T X\alpha = X^T x \tag{5}$$

- Using the autocorrelation method, $R = X^T X$, is Toeplitz and with the special right-hand side, $X^T x$ the system can be solved using the Levinson–Durbin algorithm in $\mathcal{O}(N^2)$.

Exploiting redundancies: Joint short-term and long-term prediction

- The combination of short term and long prediction filters can be seen as a sparse high order filter:



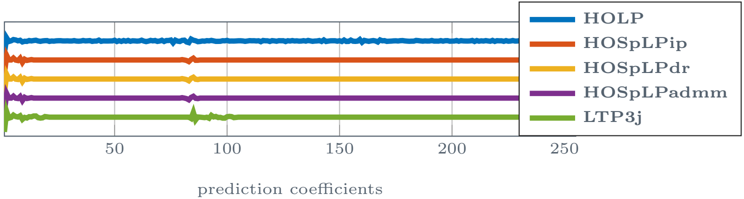


Figure : A 640 samples segment of the voiced speech (vowel /a/ uttered by a female speaker) and some predictors.

High-order sparse linear prediction

- Imposing sparsity via the 1-norm convex relaxation:

$$\underset{\alpha}{\text{minimize}} \quad \|x - X\alpha\|_2^2 + \gamma\|\alpha\|_1 \,. \tag{6}$$

- However, when imposing sparsity on both the residual vector and high-order predictor, gains can been obtained both in terms of modeling and coding performance[1]

$$\underset{\alpha}{\text{minimize}} \quad \|x - X\alpha\|_1 + \gamma\|\alpha\|_1 \,. \tag{7}$$

- In general, see[2].

---

[1] Daniele Giacobello et al. "Joint estimation of short-term and long-term predictors in speech coders". In: *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on.* IEEE. 2009, pp. 4109–4112.

[2] D. Giacobello et al. "Sparse linear prediction and its applications to speech processing". In: *Audio, Speech, and Language Processing, IEEE Transactions on* 20.5 (2012), pp. 1644–1657.

Solving the SLP problem I

- The objective:

$$f(\alpha) = \|x - X\alpha\|_1 + \gamma\|\alpha\|_1 \tag{8}$$

- is convex but not differentiable, neither is any of the terms $\rightarrow$ proximal gradient methods are not applicable[3][4][5].

---

[3] Yu. Nesterov. *Gradient methods for minimizing composite objective function.* Université catholique de Louvain, Center for Operations Research and Econometrics (CORE). No 2007076, CORE Discussion Papers. 2007.

[4] A. Beck and M. Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". In: *SIAM Journal of Imaging Sciences* 2 (1 2009), pp. 183–202.

[5] S. J. Wright, R.D. Nowak, and M. A. T. Figueiredo. "Sparse reconstruction by separable approximation". In: *Signal Processing, IEEE Transaction on* 57 (2009), pp. 2479–2493.

Solving the SLP problem II

- The problem can be solved as a general linear programming problem using interior point (IP) methods[6][7].
- Work-load concentrated on solving linear systems with coefficient matrix

$$C = X^T D_1 X + D_2, \quad D_1, D_2 \text{ diagonal, change at each iteration} \tag{9}$$

- $X^T D_1 X$ is not Toeplitz and do not have low displacement rank → resort to $\mathcal{O}(N^3)$ methods for solving linear systems.

---

[6] G. Alipoor and M. H. Savoji. "Wide-band speech coding based on bandwidth extension and sparse linear prediction". In: *Telecommunications and Signal Processing (TSP), 2012 35th International Conference on*. IEEE. 2012, pp. 454–459.

[7] T.L. Jensen et al. "Real-time implementations of sparse linear prediction for speech processing". In: *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE. 2013, pp. 8184–8188.

Solving the SLP problem III

- Is it possible to exploit the structure of X and $R = X^T X$?
- Is the high accuracy of the IP methods necessary?

An ADMM algorithm

One approach, is via the following straight-out-of-the-box alternating direction method of multipliers (ADMM) algorithm

$$\alpha^{(k+1)} = (X^T X + \gamma^2 I)^{-1} \begin{bmatrix} \gamma I & -X^T \end{bmatrix} \left(y^{(k)} + \begin{bmatrix} 0 \\ -x \end{bmatrix} - u^{(k)}\right) \tag{10}$$

$$e^{(k+1)} = x - X\alpha^{(k+1)} \tag{11}$$

$$y^{(k+1)} = S_{1/\rho}\left(\begin{bmatrix} \gamma\alpha^{(k+1)} \\ e^{(k+1)} \end{bmatrix} + u^{(k)}\right) \tag{12}$$

$$u^{(k+1)} = u^{(k)} + \begin{bmatrix} \gamma\alpha^{(k+1)} \\ e^{(k+1)} \end{bmatrix} - y^{(k+1)}\,. \tag{13}$$

- $S_{1/\rho}$ is the soft-threshold function.
- Operations with X is FIR filtering.

Solving positive-definite symmetric Toeplitz systems II

- Fast algorithms[8] $\rightarrow \mathcal{O}(N^2)$.
- Superfast algorithms[9] $\rightarrow \mathcal{O}(N \log^2 N) + \mathcal{O}(N \log N)$.
- "Intermediate"[10] $\rightarrow \mathcal{O}(N^2) + \mathcal{O}(N \log N)$.
- Break-even point in the number of operations at approximately $N = 256$ for N as a radix 2 number. We will use $N = 250$, so go for the intermediate.

---

[8] N. Levinson. "The Weiner RMS Error Criterion in Filter Design and Prediction". In: *Journal of Mathematics and Physics* 25 (1947), pp. 261–278.

[9] R.R. Bitmead and B.D.O Anderson. "Asymptotically fast solution of Toeplitz and related systems of linear equations". In: *Linear Algebra and its Applications* 34 (1980), pp. 103–116; G.S. Ammar and W.B Gragg. "Superfast solution of real positive definite Toeplitz systems". In: *SIAM Journal on Matrix Analysis and Applications* 9.1 (1988), pp. 61–76.

[10] J. R. Jain. "An efficient algorithm for a large Toeplitz set of linear equations". In: *Acoustics, Speech and Signal Processing, IEEE Transaction on* 27.6 (1979).

Solving positive-definite symmetric Toeplitz systems III

The inverse of a Toeplitz matrix can be described by the Gohberg-Semencul formula

$$\delta_N T^{-1} = T_1 T_1^T - T_0^T T_0 \qquad (14)$$

where

$$T_0 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \rho_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \rho_{N-1} & \cdots & \rho_0 & 0 \end{bmatrix}, \quad T_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \rho_{N-1} & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \rho_0 & \cdots & \rho_{N-1} & 1 \end{bmatrix}. \qquad (15)$$

- The variables δ_N and $\rho_0, \ldots, \rho_{N-1}$ can be computed using the Szegő recursions.
- Evaluation of matrix-vector products with T_0, T_0^T, T_1, T_1^T: FFTs/IFFTs.

Accuracy requirement

- The LP speech model only approximates the human speech production system and the model is not noise free[11],
- Moreover, the convex optimization framework uses the 1-norm because of its feasibility not because it was the best choice.
- Summary: solution should only be accurate enough to capture the essence of our endeavor.
- Expectation: saturation like effect as a function of accuracy[12].

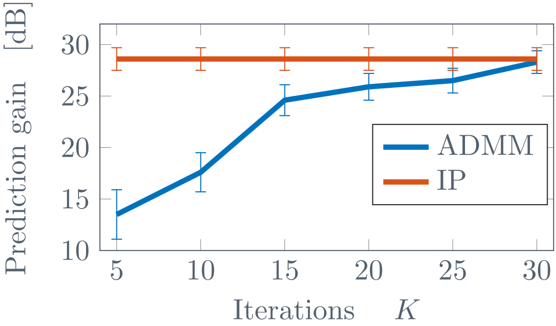
---

[11] J.R. Deller, J.G. Proakis, and J.H.L. Hansen. *Discrete-time processing of speech signals.* Ieee New York, NY, USA:, 2000.

[12] B. Defraene et al. "Real-Time Perception-Based Clipping of Audio Signals Using Convex Optimization". In: *Audio, Speech, and Language Processing, IEEE Transaction on* 20.10 (2012), pp. 2657–2671.

Prediction gain

- Processed only the vowel and semivowel phones from the TIMIT database, 3696 sentences from 462 speakers (≈40,000 voiced speech frames).
- Regularization γ obtained via modified L-curve analysis.



- Average prediction gains for a fixed number of iterations for the ADMM solution. A 95% confidence interval is shown.

Timings

- C++/ FFTW3 library / Intel Math Kernel Library (MKL)
- Varying k and linear regression.
- ADMM: Levinson:

$$t_k \approx 7 \cdot 10^{-6} + 159 \cdot 10^{-6} k \quad [\mathrm{s}], \quad C^2 = 0.999\,.$$

- ADMM: Szegö recursion + Gohberg-Semencul

$$t_k \approx 62 \cdot 10^{-6} + 55 \cdot 10^{-6} k \quad [\mathrm{s}], \quad C^2 = 0.999\,.$$

- Handtailored IP method:

$$t_k \approx 1206 \cdot 10^{-6} + 2033 \cdot 10^{-6} k \quad [\mathrm{s}], \quad C^2 = 0.999\,.$$

Summary

- High-order sparse linear prediction offers interesting properties for speech processing.
- Need to solve a linear program for each frame $\rightarrow$ real-time and embedded applications.
- An approach: efficient use of Toeplitz matrix structure using the alternating direction method of multipliers.
- Low accuracy is sufficient to obtain similar prediction gain as high accuracy methods.