Sparse Linear Prediction and Its Applications to Speech Processing

Daniele Giacobello, Mads Græsbøll Christensen, Joachim Dahl, Søren Holdt Jensen
Multimedia Information And Signal Processing
Department of Electronic Systems
Aalborg University
Denmark

Introduction

- Typically the prediction coefficients are found such that the norm-2 of the residual is minimized

  →Maximum likelihood approach when the excitation is considered to be white Gaussian and identically distributed

- Problems

  →Excitation is not always Gaussian (ex. for voiced speech excitation is best represented by a pulse train)

  →Residual not easy to quantize

- Our idea is to use a linear prediction scheme that leaves a sparse residual rather than a minimum variance one

  →More efficient quantization!

Fundamentals

- Mathematically we can state the class of problems as those covered by the optimization problem:

  → Finding the prediction coefficient vector given set of observed real samples

$$x(n)=\sum_{k=1}^{K} a_k x(n-k)+e(n) \quad \Longrightarrow \quad \min_{\underline{a}} \left\|\underline{x}-\underline{\underline{X}}\,\underline{a}\right\|_p^p+\gamma\left\|\underline{a}\right\|_k^k$$

  (ex. $p=2 \wedge \gamma=0$ → standard LP, autocovariance method)

- Sparseness usually measured using the cardinality, which results in intractable (NP-hard) problems.

  → Instead we use the more tractable Norm-1

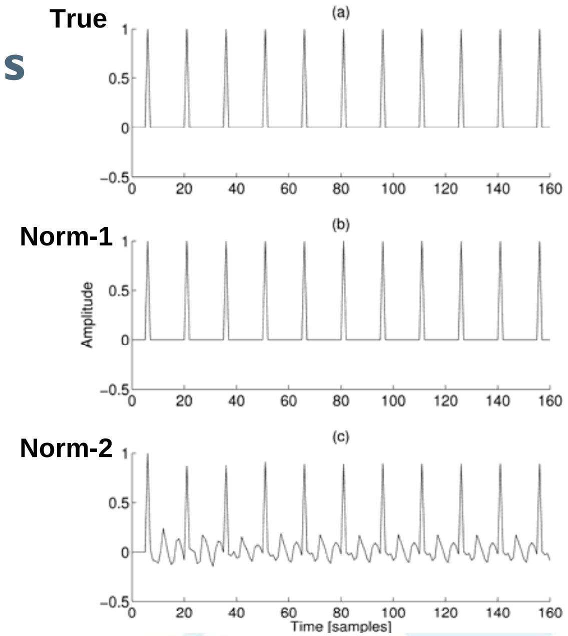
  (ex. $p=1 \wedge \gamma=0$ → ML for laplacian excitation)

- γ can have different interpretation:

  → Regularization term including prior knowledge about the coefficients

  → Minimization interpretation where gamma operates as a Lagrange multiplier

Some Properties of Norm-1 LP Analysis



- Less influenced by outliers
  → good for impulse train estimation

- Stability NOT guaranteed
  → Use of sub-optimal stable methods (BURG)
  → Other tricks (bandwidth expansion, poles reflection)

- Non-uniquess of the solution (there may be a stable solution in the set…)

- Easily solved using standard linear programming
  (or convex programming for alternative formulations)

Example of Applications: Joint Short & Long Term Linear Prediction

Considering the minimization problem: $\min_{\underline{a}} \left\| \underline{x} - \underline{\underline{X}}\,\underline{a} \right\|_p^p + \gamma \left\| \underline{a} \right\|_k^k$

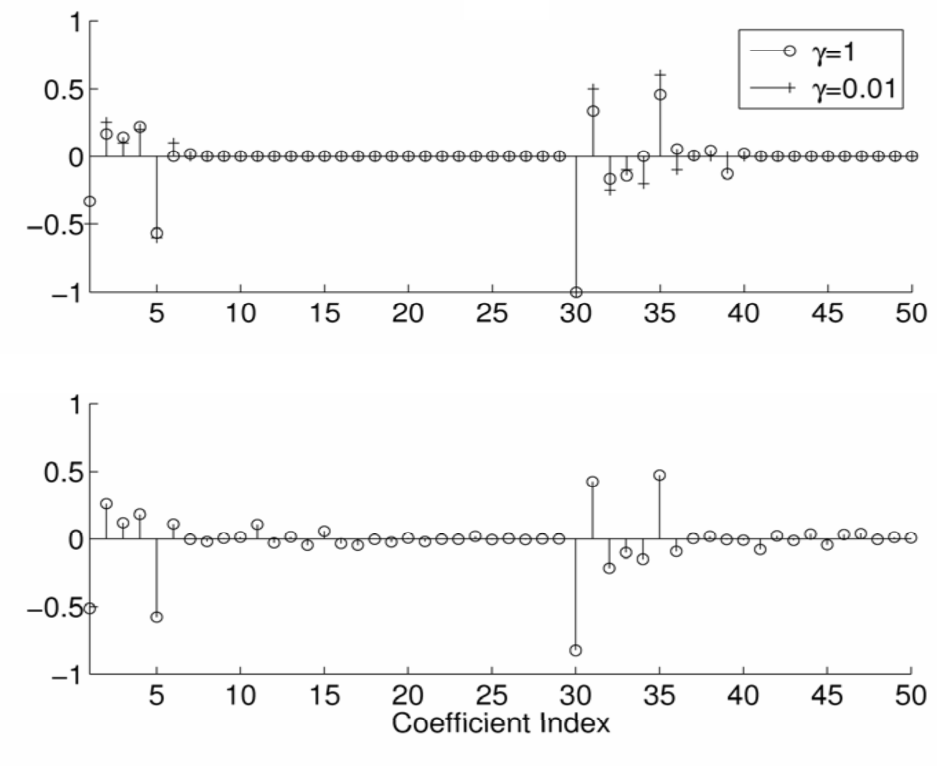
With: $p = 2, k = 1, \gamma > 0$

Using a high prediction order (ex. 30-40) and a high number of samples (300-400 @8KHz)

We will have a sparse coefficient vector (50-60% null) → factorizable in STLP and LTLP

$$\frac{1}{A_{lp}(z)} \cdot \frac{1}{1 - g_p z^{-T_p}} \simeq \frac{1}{A_{slp}(z)}$$

Example of Applications: Joint Short & Long Term Linear Prediction



$$\min_{\underline{a}} \left\|\underline{x} - \underline{\underline{X}}\,\underline{a}\right\|_2^2 + \gamma \left\|\underline{a}\right\|_1$$

$$\min_{\underline{a}} \left\|\underline{x} - \underline{\underline{X}}\,\underline{a}\right\|_2^2$$

Conclusions

- Sparse Linear Prediction based on Convex Optimization can be a breakthrough in Speech Coding
    - Residual adapted for the quantizer, rather than the other way around
    - Takes into account statistical properties ignored by the usual LP
- Main drawbacks:
    - Computational load still a bit heavy
      (1 Norm-1 minimization problem ~ 20-30 LS problems!)
    - Stability NOT guaranteed
- It's still a "work in progress" but it gave interesting results so far…

Main References

S. Boyd and L. Vandenberghe, ***Convex Optimization***. Cambridge University Press, 2004.

S. J. Wright, ***Primal-Dual Interior-Point Methods***. SIAM, 1997.

J. Makhoul, "**Linear Prediction: A Tutorial Review**," *Proc. IEEE*, vol. 63(4), pp. 561-580, Apr. 1975.

E. Denoël and J.-P. Solvay, "**Linear prediction of speech with a least absolute error criterion**," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33(6), pp. 1397-1403, Dec. 1985.

P. Kabal and R. P. Ramachandran, "**Joint optimization of linear predictors in speech coders**," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37(5), pp. 642–650, May 1989.

H. Zarrinkoub and P. Mermelstein, "**Joint optimization of short-term and long term predictors in CELP speech coders**," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 2, 2003, pp. 157-160.