Sparsity in Linear Predictive Coding of Speech

Daniele Giacobello

Motivation
In one sentence

- Revisiting early concepts in speech and audio analysis in light of the new development in sparse representation (2010)...
  ...and the effectiveness of convex programming (2015).

2 Introduction

Sparse Linear Prediction

Finding a Sparse Residual

Finding a High-Order Sparse Predictor

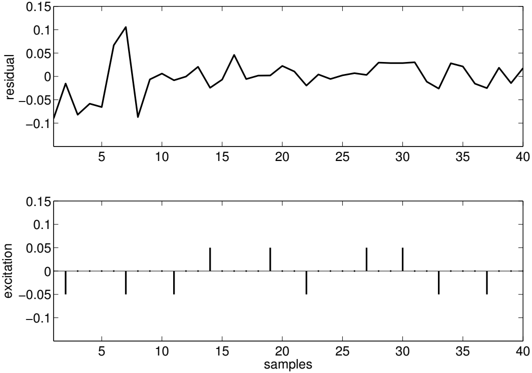
Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

Conclusions

Acknowledgements

- Initial idea: reduce the mismatch between a "white noise"-like prediction residual and a *sparse* approximation used for encoding.



Motivation
Why sparsity in Linear Prediction?

- One of the earliest problems in speech coding!

$$\widehat{\mathbf{r}} = \arg\min_{\mathbf{r}} \|\mathbf{W}(\mathbf{x} - \mathbf{Hr})\|_2^2 \text{ s.t. } \|\mathbf{r}\|_0 < K$$

- Solution is impractical due to the combinatorial nature of the problem.
- Suboptimal algorithm was proposed to find one pulse at the time: Multi-Pulse Encoding (MPE) (= Matching Pursuit).

Outline

Introduction

Sparse Linear Prediction

Finding a Sparse Residual

Finding a High-Order Sparse Predictor

Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

Conclusions

Outline

Introduction

Sparse Linear Prediction

Fundamentals

Finding a Sparse Residual

Finding a High-Order Sparse Predictor

Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

Conclusions

Acknowledgements

Sparse Linear Prediction

Fundamentals 1/2

- A speech sample $x(n)$ is approximated as a linear combination of past samples:

$$x(n) = \sum_{k=1}^{K} a_k x(n-k) + e(n),$$

where $\{a_k\}$ are the prediction coefficients, $e(n)$ is prediction error. In matrix form becomes:

$$\mathbf{x} = \mathbf{X}\mathbf{a} + \mathbf{e}.$$

- We can consider a generalized optimization framework to find $\mathbf{a}$:

$$\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_p^p + \gamma\|\mathbf{a}\|_k^k.$$

- How to choose p, k and γ depends on the kind of applications we want to implement.

Introduction
Sparse Linear Prediction
7 Fundamentals
Finding a Sparse Residual
Finding a High-Order Sparse Predictor
Reducing the 1-norm 0-norm mismatch
Compressed Sensing in Sparse LP
Conclusions
Acknowledgements

Sparse Linear Prediction
Fundamentals 2/2

- If we want to introduce sparsity in the LP optimization framework, we can set $p = 0$ and $k = 0$:

$$\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_0 + \gamma \|\mathbf{a}\|_0,$$

- γ relates to *how sparse* **a** is (prior knowledge of **a**).
- 1-norm used as a convex relaxation of the 0-norm:

$$\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_1 + \gamma \|\mathbf{a}\|_1.$$

- common formulation: $\widehat{\mathbf{a}} = \min_{\mathbf{a}} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_2^2 = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{x} = \mathbf{R}^{-1}\mathbf{r}$

Introduction
Sparse Linear Prediction
8 Fundamentals
Finding a Sparse Residual
Finding a High-Order Sparse Predictor
Reducing the 1-norm 0-norm mismatch
Compressed Sensing in Sparse LP
Conclusions
Acknowledgements

Outline

Introduction

Sparse Linear Prediction

Fundamentals

Finding a Sparse Residual

Finding a High-Order Sparse Predictor

Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

Conclusions

Acknowledgements

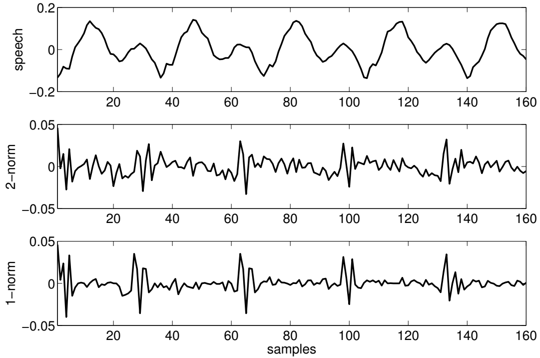
- Consider now the case of a short-term predictor that engenders a sparse residual ($\gamma = 0$):

$$\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_1.$$

- ML approach when the error sequence is considered to be a set of i.i.d. Laplacian random variables.
- Sparser residual beneficial for both analysis and coding purposes.

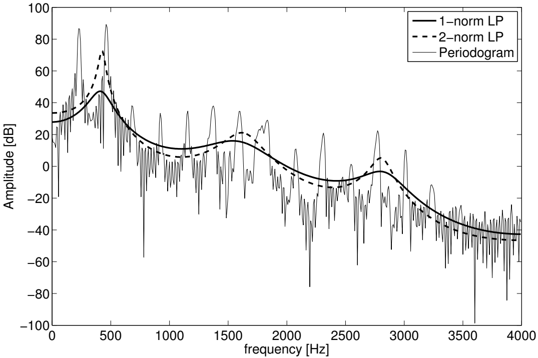
Sparsity in Linear Predictive Coding of Speech
Daniele Giacobello
Introduction
Sparse Linear Prediction
Finding a Sparse Residual
Finding a High-Order Sparse Predictor
Reducing the 1-norm 0-norm mismatch
Compressed Sensing in Sparse LP
Conclusions
Acknowledgements

Sparse Linear Prediction
Finding a Sparse Residual - Example



The spiky train characteristic of voiced speech is retrieved more accurately when we look for a sparse residual.

Sparsity in Linear Predictive Coding of Speech
Daniele Giacobello
Introduction
Sparse Linear Prediction
Finding a Sparse Residual
Finding a High-Order Sparse Predictor
Reducing the 1-norm 0-norm mismatch
Compressed Sensing in Sparse LP
Conclusions
Acknowledgements



The lower emphasis on peaks in the envelope, when 1-norm minimization is employed, is a direct consequence of the ability to retrieve the spiky pitch excitation.

Sparsity in Linear Predictive Coding of Speech
Daniele Giacobello
Introduction
Sparse Linear Prediction
12 Finding a Sparse Residual
Finding a High-Order Sparse Predictor
Reducing the 1-norm 0-norm mismatch
Compressed Sensing in Sparse LP
Conclusions
Acknowledgements
32

Outline

Introduction

Sparse Linear Prediction
 Fundamentals

Finding a Sparse Residual

Finding a High-Order Sparse Predictor

Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

Conclusions

Acknowledgements

- Consider the cascade of a short-term linear predictor $F(z)$ and a long-term linear predictor $P(z)$ to remove respectively near-sample and far-sample redundancies:

$$A(z) = \left(1 - \sum_{k=1}^{N_f} f_k z^{-k}\right)\left(1 - \sum_{k=1}^{N_p} g_k z^{-(T_p+k-1)}\right).$$

- The resulting prediction coefficient vector $\mathbf{a} = \{a_k\}$ of the high order polynomial $A(z)$ will therefore be highly sparse.
- We can impose sparsity on a high-order predictor:

$$\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_p^p + \gamma\|\mathbf{a}\|_1.$$

- When $p = 2$ minimum variance approach, $p = 1$ encourages sparsity also on the residual.

Sparsity in Linear Predictive Coding of Speech

Daniele Giacobello

Introduction

Sparse Linear Prediction

Finding a Sparse Residual

Finding a High-Order Sparse Predictor

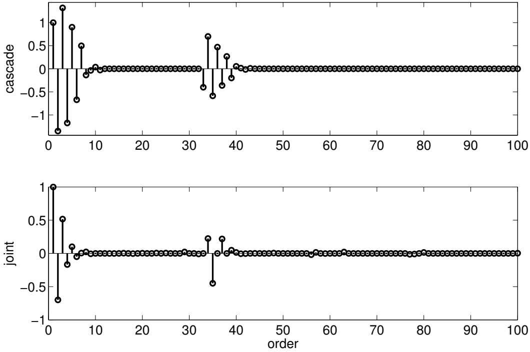
Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

Conclusions

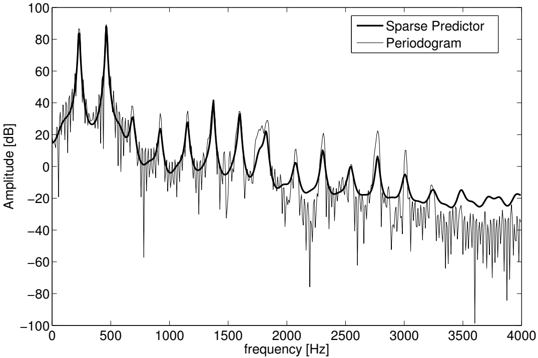
Acknowledgements

Sparse Linear Prediction
Finding a High-Order Sparse Predictor - Example



The prediction coefficients vector is similar to the multiplication of the short-term prediction filter and long-term prediction filter usually obtained in cascade.

Sparse Linear Prediction
Finding a High-Order Sparse Predictor - Example



Spectral modeling properties of a high order sparse predictor with only nine nonzero coefficients.

Sparsity in Linear Predictive Coding of Speech
Daniele Giacobello
Introduction
Sparse Linear Prediction
Finding a Sparse Residual
Finding a High-Order Sparse Predictor
Reducing the 1-norm 0-norm mismatch
Compressed Sensing in Sparse LP
Conclusions
Acknowledgements

- The purpose of the high order sparse predictor is to model the *whole* spectrum, i.e., the spectral envelope and the spectral harmonics.
- γ controls the sparsity of the prediction coefficient vector. If γ is chosen appropriately, we can obtained again $F(z)$ and $P(z)$ through approximate factorization.
- Intrinsic model order selection!

Sparsity in Linear Predictive Coding of Speech
Daniele Giacobello
Introduction
Sparse Linear Prediction
Finding a Sparse Residual
Finding a High-Order Sparse Predictor
Reducing the 1-norm 0-norm mismatch
Compressed Sensing in Sparse LP
Conclusions
Acknowledgements

Outline

Introduction

Sparse Linear Prediction

Fundamentals

Finding a Sparse Residual

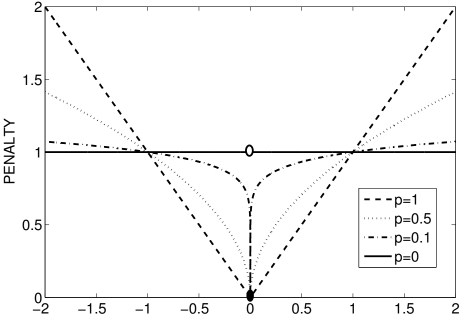
Finding a High-Order Sparse Predictor

Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

Conclusions

Acknowledgements



- Reweighted 1-norm minimization balances the dependence on the magnitude of the 1-norm.
- Changing the cost function and moving the problem towards the 0-norm minimization with convex tools (convergence to the log-sum penalty function).

Sparsity in Linear Predictive Coding of Speech
Daniele Giacobello
Introduction
Sparse Linear Prediction
Finding a Sparse Residual
Finding a High-Order Sparse Predictor
Reducing the 1-norm 0-norm mismatch
Compressed Sensing in Sparse LP
Conclusions
Acknowledgements

Outline

Introduction

Sparse Linear Prediction

Fundamentals

Finding a Sparse Residual

Finding a High-Order Sparse Predictor

Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

Conclusions

Acknowledgements

- If sparse, our solution lies in a subspace of reduced dimensionality.
- The Euclidean distance between all points in the signal model is preserved when applying a proper *random* projection (Restricted Isometry Property, RIP).
- Two ingredients needed for CS: a domain where the signal is sparse and the sparsity level T.
  - Sparsity in the residual domain means that a limited number of $M \propto T$ random projections are sufficient for recovery.
- The *shrinkage* of the minimization problem in a lower dimensional space will have a clear impact on the complexity.

Sparsity in Linear Predictive Coding of Speech
Daniele Giacobello
Introduction
Sparse Linear Prediction
Finding a Sparse Residual
Finding a High-Order Sparse Predictor
Reducing the 1-norm 0-norm mismatch
Compressed Sensing in Sparse LP
Conclusions
Acknowledgements

- Original MPE problem (known predictor):

$$\widehat{\mathbf{r}} = \arg\min_{\mathbf{r}} \|\mathbf{x} - \mathbf{H}\mathbf{r}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{r}\|_0 = T.$$

- CS Formulation:

$$\widehat{\mathbf{r}} = \arg\min_{\mathbf{r}} \|\mathbf{r}\|_1 + \gamma\|\Phi\mathbf{x} - \Phi\mathbf{H}\mathbf{r}\|_2^2.$$

- 1-norm global optimization as convex relaxation of the 0-norm: near-optimal selection of sparse excitation.
- Sparsity-knowledge-based *shrinkage*: reduction of constraints $\rightarrow$ computationally faster.

Sparsity in Linear Predictive Coding of Speech

Daniele Giacobello

Introduction

Sparse Linear Prediction

Finding a Sparse Residual

Finding a High-Order Sparse Predictor

Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

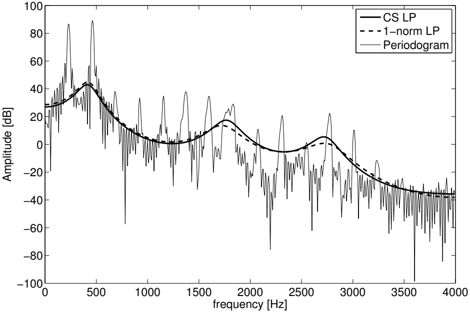
Conclusions

Acknowledgements

- To adapt CS principles to the estimation of the predictor as well, consider the relation between the synthesis matrix $\mathbf{H}$ and the analysis matrix $\mathbf{A}$ ($\mathbf{A} = \mathbf{H}^{+}$):

$$\min_{\mathbf{a},\mathbf{r}} \|\mathbf{r}\|_1 \quad \text{s.t.} \quad \Phi\mathbf{r} = \Phi(\mathbf{x} - \mathbf{X}\mathbf{a}).$$

- Equivalent to our original formulation projected onto a lower-dimensional space.
- We can involve a reweighting procedure also here.



1-norm solution with and without CS shrinkage (170 equations vs. 40 equations).

Sparsity in Linear Predictive Coding of Speech

Daniele Giacobello

Introduction

Sparse Linear Prediction

Finding a Sparse Residual

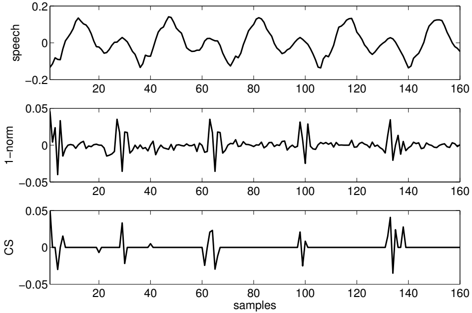
Finding a High-Order Sparse Predictor

Reducing the 1-norm 0-norm mismatch

24 Compressed Sensing in Sparse LP

Conclusions

Acknowledgements



CS recovery of the sparse residual. The imposed sparsity level is $T = 20$, corresponding to the size $M = 40$ for the sensing matrix.

Sparsity in Linear Predictive Coding of Speech

Daniele Giacobello

Introduction

Sparse Linear Prediction

Finding a Sparse Residual

Finding a High-Order Sparse Predictor

Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

Conclusions

Acknowledgements

Outline

Introduction

Sparse Linear Prediction
    Fundamentals

Finding a Sparse Residual

Finding a High-Order Sparse Predictor

Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

Conclusions

Acknowledgements

- Main advantage is to overcome some of 2-norm LP known issues of the envelope estimation
  - Lower spectral distortion.
  - Invariant to small shift of the analysis window.
  - Pitch Independence.
- Also accurate in estimating integer pitch lags when used in its high-order form

Introduction

Sparse Linear Prediction

Finding a Sparse Residual

Finding a High-Order Sparse Predictor

Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

Conclusions

Acknowledgements

- Synergistic multi-stage coding
  - sparse predictor $\rightarrow$ sparse encoding
- Possibility variable rate coding through high-order predictor modeling
  - model order selection and intrinsic V/UV classification
- Less parameters necessary.

Introduction
Sparse Linear Prediction
Finding a Sparse Residual
Finding a High-Order Sparse Predictor
Reducing the 1-norm 0-norm mismatch
Compressed Sensing in Sparse LP
Conclusions
Acknowledgements

Sparse Linear Prediction
Drawbacks

- Stability not guaranteed. Defined new methods to tackle this problem:
  - Reducing the numerical range of the shift operator.
  - Constrained 1-norm based on the alternative Cauchy bound.
- Computational Complexity:
  - Compressed Sensing reduces the number of constraints.
  - Efficient first-order convex optimization algorithms
    - ADMM algorithm can be interpreted as iterative "sparsification" of the 2-norm "classical" LP solution

$$\mathbf{y}^{(k+1)} = \mathcal{S}_{1/\rho}\left(\begin{bmatrix}\gamma\mathbf{a}^{(k+1)}\\ \mathbf{r}^{(k+1)}\end{bmatrix} + \mathbf{u}^{(k)}\right)$$

    - where $\mathcal{S}_{1/\rho}$ is the soft-thresholding operator (sparsity!)
  - Much of the total computational cost in a speech coder is saved by the "one-step" procedure.
- Non-Uniqueness (still optimal!).
- Lack of a Frequency-Domain interpretation.

Introduction
Sparse Linear Prediction
Finding a Sparse Residual
Finding a High-Order Sparse Predictor
Reducing the 1-norm 0-norm mismatch
Compressed Sensing in Sparse LP
29 Conclusions
Acknowledgements

Sparse Linear Prediction

References

- "Sparse Linear Prediction and Its Applications to Speech Processing," *IEEE T-ASL*, 2012.
- "Retrieving Sparse Patterns Using a Compressed Sensing Framework: Applications to Speech Coding Based on Sparse Linear Prediction," *IEEE SPL*, 2010.
- "High-Order Sparse Linear Predictors for Audio Processing," *Proc. EUSIPCO*, 2010.
- "Stable 1-norm Error Minimization Based Linear Predictors for Speech Modeling," *IEEE T-ASL*, 2014.
- *"Real-time Implementations of Sparse Linear Prediction for Speech Processing," Proc. ICASSP, 2013. + SOFTWARE*
- *"Fast Algorithms for High-Order Sparse Linear Prediction with Applications to Speech Processing," Speech Communication 2015. + SOFTWARE*

Outline

Introduction

Sparse Linear Prediction

Fundamentals

Finding a Sparse Residual

Finding a High-Order Sparse Predictor

Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

Conclusions

Acknowledgements

Acknowledgements

- Prof. Søren Holdt Jensen (adviser)
- Prof. Marc Moonen (adviser)
- Prof. Mads Græsbøll Christensen (estimation theory and sparse representation)
- Prof. Manohar N. Murthi (speech modeling and coding)
- Dr. Tobias Lindstrøm Jensen (convex optimization)
- Dr. Toon van Waterschoot (extension to audio processing and acoustics, e.g., dereverberation)

Introduction

Sparse Linear Prediction

Finding a Sparse Residual

Finding a High-Order Sparse Predictor

Reducing the 1-norm 0-norm mismatch

Compressed Sensing in Sparse LP

Conclusions

32 Acknowledgements