Enhancing Sparsity in Linear Prediction of Speech by Iteratively Reweighted 1-norm Minimization

Daniele Giacobello[1] Mads Græsbøll Christensen[1] Manohar N. Murthi[2] Søren Holdt Jensen[1] Marc Moonen[3]

[1]Aalborg Universitet, Aalborg, Denmark
[2]University of Miami, Florida, USA
[3]Katholieke Universiteit Leuven, Leuven, Belgium

1 Introduction

- Linear prediction of speech based on 1-norm minimization has already proved to be an interesting alternative to 2-norm minimization.
- Choosing the 1-norm as a convex relaxation of the 0-norm, the corresponding linear prediction model offers a sparser residual better suited for coding applications.
- Purpose of this paper is to overcome the mismatch between 0-norm minimization and 1-norm minimization while keeping the problem solvable with convex optimization tools.
- Experimental results prove the effectiveness of the reweighted 1-norm minimization, offering better coding properties compared to 1-norm minimization.

2 Sparse Linear Prediction

- The problem considered in this paper is based on the following Auto-Regressive (AR) speech production model, where a sample of speech $x(n)$ is written as a linear combination of K past samples:

$$x(n) = \sum_{k=1}^{K} a_k x(n-k) + r(n), \quad 0 < n \leq N,$$

where $\{a_k\}$ are the prediction coefficients and $r(n)$ is the driving noise process (commonly referred to as the prediction residual).
- The prediction coefficient vector $\mathbf{a}$ (order K) is found by minimizing the $1-$norm of the residual $\mathbf{r}$:

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

- An interesting alternative problem formulation is obtained when sparsity is also imposed on the predictor:

$$\hat{\mathbf{a}}, \hat{\mathbf{r}} = \arg\min_{\mathbf{a}} \|\mathbf{r}\|_1 + \gamma\|\mathbf{a}\|_1, \quad \text{s.t.} \quad \mathbf{r} = \mathbf{x} - \mathbf{X}\mathbf{a};$$

in this case the sparse structure of the predictor (in this case high order) allows a joint estimation of a short-term and a long-term predictor.

3 Reweighted 1-norm

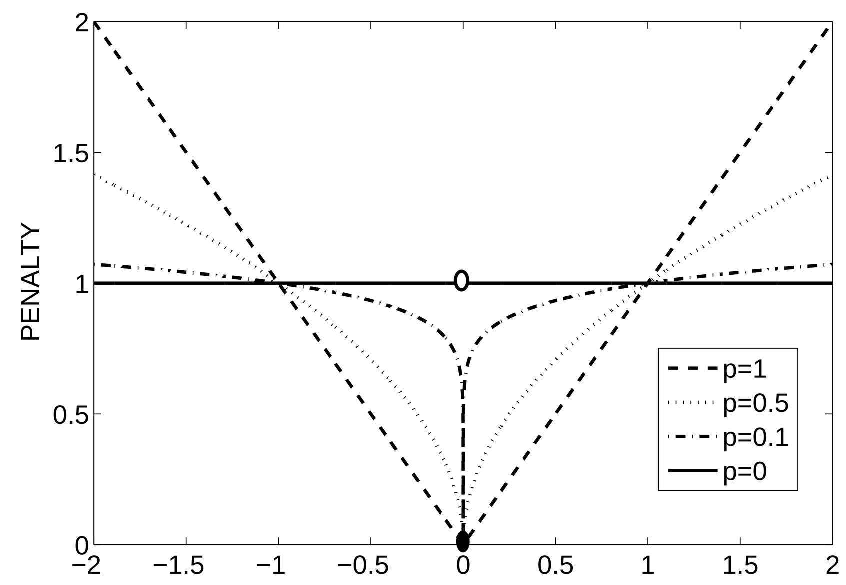
- **Iteratively Reweighted 1-norm Minimization of the Residual**

Inputs: speech segment $\mathbf{x}$
Outputs: predictor $\hat{\mathbf{a}}^i$, residual $\hat{\mathbf{r}}^i$
$i = 0$, initial weights $\mathbf{W}^{i=0} = \mathbf{I}$
while halting criterion false **do**
1. $\hat{\mathbf{a}}^i, \hat{\mathbf{r}}^i \leftarrow \arg\min_{\mathbf{a}} \|\mathbf{W}^i\mathbf{r}\|_1$
2. $\mathbf{W}^{i+1} \leftarrow \text{diag}\left(|\hat{\mathbf{r}}^i| + \epsilon\right)^{-1}$
3. $i \leftarrow i+1$

- **Iteratively Reweighted 1-norm Minimization of Residual and Predictor**

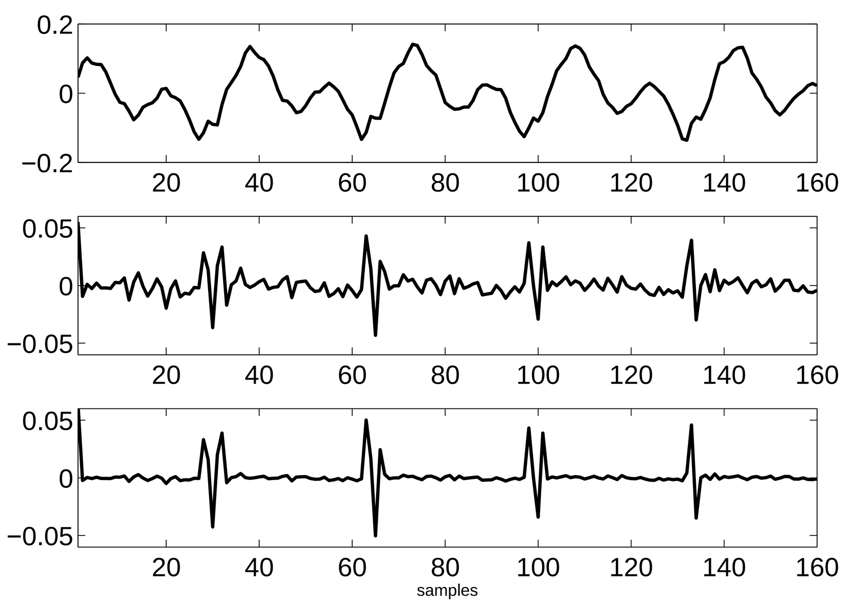
Inputs: speech segment $\mathbf{x}$
Outputs: predictor $\hat{\mathbf{a}}^i$, residual $\hat{\mathbf{r}}^i$
$i = 0$, initial weights $\mathbf{W}^{i=0} = \mathbf{I}$ and $\mathbf{D}^{i=0} = \mathbf{I}$
while halting criterion false **do**
1. $\hat{\mathbf{a}}^i, \hat{\mathbf{r}}^i \leftarrow \arg\min_{\mathbf{a}} \|\mathbf{W}^i\mathbf{r}\|_1 + \gamma\|\mathbf{D}^i\mathbf{a}\|_1$
2. $\mathbf{W}^{i+1} \leftarrow \text{diag}\left(|\hat{\mathbf{r}}^i| + \epsilon\right)^{-1}$
3. $\mathbf{D}^{i+1} \leftarrow \text{diag}\left(|\hat{\mathbf{a}}^i| + \epsilon\right)^{-1}$
4. $i \leftarrow i+1$

- The reweighting process will balance the dependence on the magnitude of the 1-norm, changing the cost function and moving the problem towards the 0-norm minimization.
- The weights are chosen as the inverse of the magnitude of the residual.
- $\epsilon > 0$ is used to provide stability when a component of $\hat{\mathbf{r}}$ goes to zero.
- $\|\hat{\mathbf{r}}^{i+1}\|_1 \leq \|\hat{\mathbf{r}}^i\|_1$
- The halting criterion can be chosen as either a maximum number of iterations or as a convergence criterion.

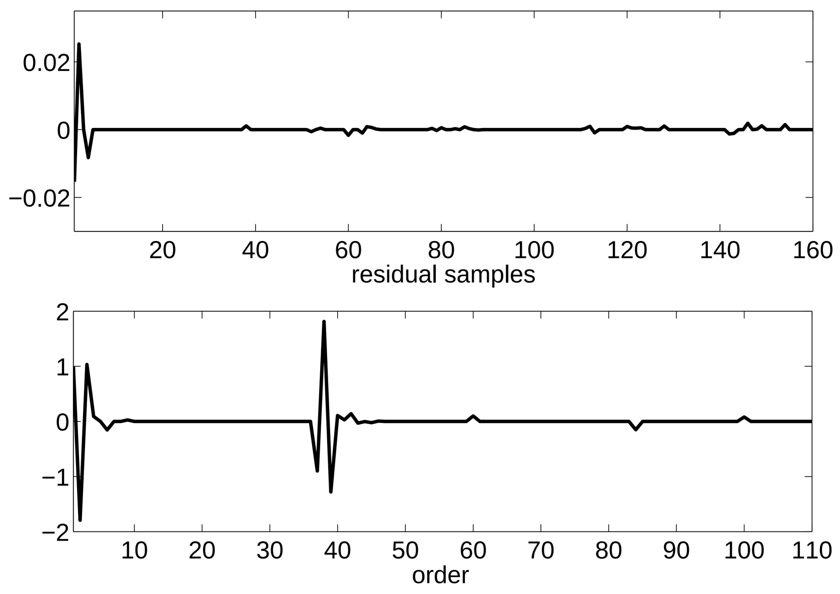


Comparison between cost functions for $p \leq 1$. The 0-norm can be seen as more "democratic" than any other norm by weighting all the nonzero coefficients equally.

4 Experiments



Comparison between true 1-norm 10th order LP residual (middle) and iteratively reweighted 1-norm LP residual (bottom). The original voiced speech is shown on top. Three iteration where performed, sufficient to reach convergence



Residual and 110th order linear predictor at the convergence after three iterations. As it is clear, all the information of the speech signal is transferred to the predictor which also show a very clear sparse structure, similar to the convolution between the coefficients of short-term and long-term predictors.

5 Validation

- The residual sequence is coded with sparse multipulse techniques ($M = 20$) after the reweighted scheme. The frame length is $N = 160$ (20 ms). The prediction order is $K = 10$.
- In **MPE1r** it is not necessary to calculate the positions of the nonzero pulses are located (as it is usually done in MPE coding), we simply exploit the information coming out of the predictive analysis. We are then clearly moving our problem towards a more synergistic way to code a signal.

METHOD	SSNR	MOS
MPE1r	20.9±1.9	3.24±0.03
MPE1	20.0±3.2	3.20±0.12
MPE2r	19.3±2.9	3.17±0.10
MPE2	18.5±2.1	3.17±0.22

Comparison between the MPE residual estimation methods in terms of Segmental SNR and Mean Opinion Score (PESQ evaluation). A 95% confidence intervals is given for each value. Each signal is coded with 9500 bits/s.

- When both predictor and residual are processed in the reweighted minimization, we can further reduce the number of nonzero samples in the MPE process ($M = 5$ and $M = 10$ pulses respectively in the voiced and unvoiced case).

METHOD	SSNR	MOS
J11r+os	27.9±0.9	3.59±0.02
J11r	25.3±1.3	3.43±0.03
J11os	24.7±1.0	3.40±0.09
J11	23.9±1.9	3.22±0.09

Comparison between the coding methods with joint estimation of residual and predictor in terms of Segmental SNR and Mean Opinion Score (PESQ evaluation). A 95% confidence intervals is given for each value. The approximate average bit rate is 5175 bit/s.

6 Discussion

- With just few iterations, we were able to move the error minimization criterion toward the 0-norm solution, showing general improvements over conventional 1-norm minimization in coding purposes.
- We have no prior knowledge of where the residual should be nonzero. This brings the bit allocated to describe the position of few samples to significantly increase the rate.
- An interesting case would be to *structure* the reweighting process by imposing where we would like to have the nonzero pulses located, as in RPE encoding scheme.

References

[1] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen and M. Moonen, "Sparse linear prediction and its applications to speech processing," submitted to *IEEE Trans. Speech, Audio and Language Processing*, 2010.

[2] E. J. Candés, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14(5-6), pp. 877–905, 2008.