

On Homotopy Continuation for Speech Restoration

Darian M. Onchis¹(✉) and Pedro Real²

¹ Faculty of Mathematics, University of Vienna, Vienna, Austria
darian.onchis@univie.ac.at

² Department of Applied Mathematics I, University of Seville, Seville, Spain
real@us.es

Abstract. In this paper, a homotopy-based method is employed for the recovery of speech recordings from missing or corrupted samples taken in a noisy environment. The model for the acquisition device is a compressed sensing scenario using Gabor frames. To recover an approximation of the speech file, we used the basis pursuit denoising method with the homotopy continuation algorithm. We tested the proposed method with various speech recordings.

Keywords: Homotopy continuation · Speech restoration · Basis pursuit · ℓ^1 regularization · Gabor frames · Numerical algorithm

1 Introduction

The reconstruction of an audio signal with missing sampled or clipped, is a classical problem in signal processing and it was largely discussed in the specialized scientific literature, see [1–3].

In this paper, we report the experiments performed with a method inspired from computational topology, namely the homotopy continuation method in order to enhance the typical recovery of audio speech recordings based on ℓ^1 –minimization, [4–6].

To precisely formulate the problem, we consider the following non parametric model with observations:

$$y = \Theta s + e \in \mathbb{R}^P$$

where $s \in \mathbb{R}^N$ is the speech signal to recover, $e \in \mathbb{R}^P$ is a noise vector, and $\Theta \in \mathbb{R}^{P \times N}$ models the acquisition device. This device is nowadays equipped with the additional assumption of sparsity, which refers to the circumstance that many natural signals can be expanded (using a suited dictionary Θ) with only few non zero coefficients. We assume a compressed sensing scenario where the operator Θ could be the realization of a random Gaussian, Bernoulli, or partial Fourier matrix satisfying the restricted isometry property (RIP) [7]. But given the special characteristics of nature signals as the speech recordings, which usually consist of sets of distinct components as transients and harmonics with orientation in time and frequency, we have used for the proposed method a Gabor frame generated by the Alltop sequences as proposed in [8, 9].

2 Gabor Frames and the ℓ^1 -minimization

Frames $(g_i)_{i \in I}$ generalize the idea of a basis in a Hilbert space \mathbf{H} and consist of the indexed families such that the so-called frame operator S

$$Sf = \sum_{i \in I} \langle f, g_i \rangle g_i \tag{1}$$

is invertible. The main tool for time-frequency analysis is the Short-Time Fourier Transform, defined for functions $f, g \in L^2(\mathbb{R}^d)$ at $\lambda = (\alpha, \beta) \in \mathbb{R}^{2d}$ by

$$V_g f(\lambda) = V_g f(\alpha, \beta) = \langle f, M_\beta T_\alpha g \rangle = \langle f, \pi(\lambda)g \rangle \tag{2}$$

where $T_\alpha f(t) = f(t - \alpha)$ is the translation (time shift) and $M_\beta f(t) = e^{2\pi i \beta \cdot t} f(t)$ is the modulation (frequency shift). The operators $\pi(\lambda) := M_\beta T_\alpha$ are called time-frequency shifts and the set $\Lambda = \{\lambda; \lambda = (\alpha, \beta) \in \mathbb{R}^d \times \widehat{\mathbb{R}}^d\}$ is a lattice, [11]. The Gabor system $\mathcal{G}(g, \Lambda) = \{\pi(\lambda)g; \lambda \in \Lambda\}$ over the lattice Λ consisting of the translated and modulated versions of one atom g , is a frame for the space $L^2(\mathbb{R}^d)$, if and only if there exist $0 < A \leq B < \infty$ (frame bounds) with

$$A\|f\|^2 \leq \sum_{\lambda \in \Lambda} |\langle f, \pi(\lambda)g \rangle|^2 \leq B\|f\|^2 \quad \text{for every } f \in L^2(\mathbb{R}^d), \tag{3}$$

We will use in the construction of Gabor frames the Alltop sequences as proposed in [8].

To recover an approximation of the signal s , a standard method is the basis pursuit denoising or ℓ^1 -minimization [10]. This method is based on using the ℓ^1 norm as a sparsity enforcing penalty. That turns into an optimization problem and allows us to recover the signal minimizing the expression:

$$s_\rho \in \operatorname{argmin}_{s \in \mathbb{R}^N} \frac{1}{2} \|y - \Theta s\|^2 + \rho \|s\|_1 \tag{4}$$

where the ℓ^1 norm is defined as $\|s\|_1 = \sum_i |s_i|$.

The parameter ρ should be set in accordance to the noise level $\|e\|$.

In the case where there is no noise, $e = 0$, we let $\lambda \rightarrow 0^+$ and solve the basis pursuit constrained optimization $s_{0^+} \in \operatorname{argmin}_{\Theta s = y} \|s\|_1$.

In order to avoid technical difficulties, we could further assume that Θ is such that s_ρ is uniquely defined.

In the following, for some index set $I \subset \{1, \dots, N\}$, we denote by

$$\Theta_I = (\theta_i)_{i \in I} \in \mathbb{R}^{P \times |I|}$$

the sub-matrix obtained by extracting the columns $\theta_i \in \mathbb{R}^P$ of Θ indexed by I . The support of a vector is $\operatorname{supp}(x) = \{i \in \{1, \dots, N\} : x_i \neq 0\}$.

Using results from the convex analysis, we obtain that s_ρ is a solution of (4) if and only if

$$\begin{cases} (C1) & \Theta_I^*(y - \Theta_I s_{\rho, I}) = \rho \operatorname{sign}(s_{\rho, I}), \\ (C2) & \|\Theta_J^*(y - \Theta_I s_{\rho, I})\|_\infty \leq \rho \end{cases}$$

where $I = \operatorname{supp}(s_\rho)$ and $J = I^c$ is the complementary.

3 The Homotopy-Continuation Algorithm and Experiments

Topology helps to understand the different degrees of connectivity a geometric object has. To deal with topological isomorphisms or homeomorphisms between continuous geometric objects is a hard task and discretization strategies, such as triangulations, are employed for reducing the computational complexity of the topological interrogation. While homology considers the notion of hole in linear algebra terms, the homotopy is dealing with the same issues in a purely combinatorial terms. Therefore, homotopy computation is much more harder in general than homology computation, but in combination with numerical methods it can be proven to be a useful tool for signal recovery but also in image recognition.

The proposed homotopy-based method for speech recovery is gradually deforming a trivial initialization of the speech vector into the original speech vector through the process of path-tracking. The numerical homotopy procedure is based on the fact that the objective function undergoes a homotopy from the ℓ^2 to the ℓ^1 optimization as the algorithm progresses. The homotopy algorithm proceeds by computing iteratively the value s_ρ .

We sum below the complete algorithm:

Homotopy-speech restoration algorithm

Input: y -noisy speech file,

Θ -Gabor frame compressed sensing operator,

*Initialization: $Corr = \Theta' * y$, $\rho = \max(Corr)$, $s_\rho = 0$,*

$I_{sparsity} = \text{supp}(s_\rho)$

Output: s_ρ -restored speech file, ρ ,

$I_{sparsity}$

Begin iteration

*Compute the correlations $Corr = \Theta' * (y - \Theta * s_\rho)$;*

*Update direction $dir = \Theta' * \Theta \text{ sign}(Corr)$;*

Compute J the complementary support of $I_{sparsity}$

Compute minimum α for condition (C1) and (C2)

*Update solution $s_\rho = s_\rho + \alpha * dir$; $\rho = \rho - \alpha$; $I_{sparsity} = \text{supp}(s_\rho)$.*

End Iteration

(ℓ^1 -minimization with homotopy deformation)

For the numerical experiments, we have used 5 speech data s of 2 to 5 s, recorded by a microphone and sampled at 16 kHz. All signals were normalized, and after that the following noise level $\sigma = 0.05 * \text{norm}(\Theta * s) / \text{sqrt}(P)$ was applied. We used $P = \text{round}(N/4)$ where $N = \text{size}(s)$. The distorted measurements were defined by the expression $y = \Theta * s + \sigma * \text{randn}(P, 1)$ as in [5]. These measurements were the input for our algorithm.

In Fig. 1, we displayed 6 iterations of the algorithm to visualize the homotopic progression towards the correct restoration. For clarity reasons, only the first 2000 samples of the speech signal are shown.

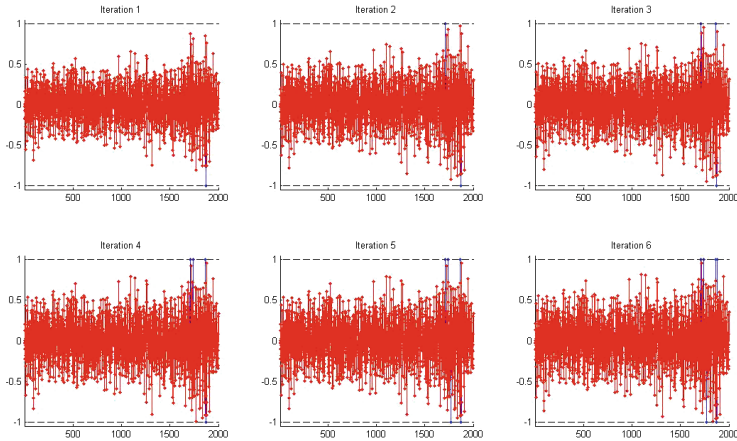


Fig. 1. First 6 iterations of the homotopy algorithm (original in red, recovered in blue) (Color figure online)

Even though the application of the algorithm provides a complete recovery of the original speech recording, a drawback is the large number of iterations. In our experiments, we managed to recover the 5 speech data, with a number of iterations proportional with almost half the size of the signal, depending on the distortion applied. In comparison with other ℓ^1 -minimization methods like the iterative shrinkage-thresholding, proximal gradient or augmented Lagrange multiplier, the homotopy achieves the best accuracy, even though, as mentioned before, in terms of speed, the homotopy takes longer time to converge when the distortion is high. But since speech recognition is usually a sensitive issue, the accuracy degree of the reconstruction made us confident in the utility of the proposed algorithm.

4 Conclusions

In this report, we presented the results of speech restoration using the basis pursuit algorithm in a sparse Gabor frames scenario, enhanced with a topology-inspired procedure entitled the homotopy-continuation method. The method allows a complete recovery of a speech recording with missing samples or clipped but with a high computational cost given by a large number of iteration necessary. Further parallelization of the algorithm are considered by the authors.

Acknowledgments. The first author gratefully acknowledge the support of the Austrian Science Fund (FWF): project number P27516.

References

1. Abel, J.S., Smith III., J.O.: Restoring a clipped signal. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing. IEEE, pp. 1745–1748 (1991)
2. Godsill, S.J., Rayner, P.J.: A Bayesian approach to the restoration of degraded audio signals. *IEEE Trans. Speech Audio Process.* **3**(4), 267–278 (1995)
3. Adler, A., Emiya, V., Jafari, M., Elad, M., Gribonval, R., Plumbley, M.D.: Audio inpainting. *IEEE Trans. Audio Speech Lang. Process.* **20**(3), 922–932 (2012)
4. Emmanuel Candes. <http://statweb.stanford.edu/~candes/l1magic/>
5. Numerical Tours of Signal Processing. <http://www.numerical-tours.com/matlab/optim8homotopy/>
6. Malioutov, D.M., Cetin, M., Willsky, A.S.: Homotopy continuation for sparse signal representation. In: IEEE International Conference on Acoustics, Speech and Signal Processing, Philadelphia, PA, vol. 5, pp. 733–736, March 2005
7. Candes, E.J., Tao, T.: Decoding by linear programming. *IEEE Trans. Inf. Theor.* **51**(12), 4203–4215 (2005)
8. Herman, M.A., Strohmer, T.: High-resolution radar via compressed sensing. *IEEE Trans. Signal Process.* **57**(6), 2275–2284 (2009)
9. Strohmer, T., Heath, R.: Grassmanian frames with applications to coding and communication. *Appl. Comput. Harmon. Anal.* **14**(3), 257–275 (2003)
10. Gill, P.R., Wang, A., Molnar, A.: The in-crowd algorithm for fast basis pursuit denoising. *IEEE Trans. Signal Process.* **59**(10), 4595–4605 (2011)
11. Ricaud, B., Stempfeler, G., Torresani, B., Wiesmeyer, C., Lachambre, H., Onchis, D.: An optimally concentrated Gabor transform for localized time-frequency components. *Adv. Comput. Math.* **40**(3), 683–702 (2014)