An Alternating Projections Algorithm for Sparse Blind Deconvolution

Aniruddha Adiga

Department of Electrical Engineering
Indian Institute of Science, Bangalore
Email: aniruddha@ee.iisc.ernet.in

Advisor: Prof. Chandra Sekhar Seelamantula

Problem Formulation: Blind Deconvolution (BD)

- Consider the linear measurement model

$$\mathbf{y} = \mathbf{h} * \mathbf{e} + \mathbf{w} = \mathbf{H}\mathbf{e} + \mathbf{w} = \mathbf{E}\mathbf{h} + \mathbf{w},$$

$\mathbf{h} \in \mathbb{R}^L$: Point spread function (PSF).
$\mathbf{e} \in \mathbb{R}^M$: Excitation.
$\mathbf{y} \in \mathbb{R}^N$: observation.
$\mathbf{w} \in \mathbb{R}^N$: Noise ($N = L + M - 1$).
$\mathbf{E} \in \mathbb{R}^{N \times L}$: linear convolution matrices corresponding to $\mathbf{e}$.
$\mathbf{H} \in \mathbb{R}^{N \times M}$: linear convolution matrices corresponding to $\mathbf{h}$.

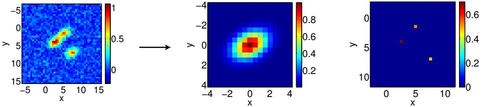
Some Examples

- Debluring: Blurring filter not known.

Some Examples

- Debluring: Blurring filter not known.

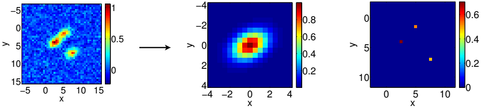
- Point source blurred due Gaussian PSF with unknown parameters.



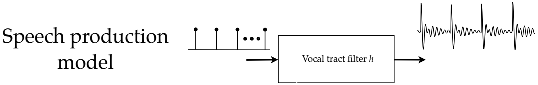
Some Examples

- Debluring: Blurring filter not known.

- Point source blurred due Gaussian PSF with unknown parameters.



- Linear speech production model: Factorize voiced speech into vocal-tract filter and excitation.



Objective and Typical Issues

Objective

- Develop a sparse blind deconvolution (SBD) algorithm to estimate **h** and **e** from **y**.

  Assumption: **e** is a sparse vector and **h** is a smooth and stable operator.

Objective and Typical Issues

Objective

- Develop a sparse blind deconvolution (SBD) algorithm to estimate $\mathbf{h}$ and $\mathbf{e}$ from $\mathbf{y}$.

  Assumption: $\mathbf{e}$ is a sparse vector and $\mathbf{h}$ is a smooth and stable operator.

Objective and Typical Issues

Objective

- Develop a sparse blind deconvolution (SBD) algorithm to estimate $\mathbf{h}$ and $\mathbf{e}$ from $\mathbf{y}$.

  Assumption: $\mathbf{e}$ is a sparse vector and $\mathbf{h}$ is a smooth and stable operator.

Typical issues

- Infinitely many solutions to the BD problem. Need priors on $\mathbf{e}$ and/or $\mathbf{h}$ to reduce search space.

Objective and Typical Issues

Objective

- Develop a sparse blind deconvolution (SBD) algorithm to estimate $\mathbf{h}$ and $\mathbf{e}$ from $\mathbf{y}$.

  Assumption: $\mathbf{e}$ is a sparse vector and $\mathbf{h}$ is a smooth and stable operator.

Typical issues

- Infinitely many solutions to the BD problem. Need priors on $\mathbf{e}$ and/or $\mathbf{h}$ to reduce search space.
- The formulated cost function is non-convex, thus leading to local minima issues.

Our Contribution

- We formulate the SBD problem using the MAP formulation.
- Propose an alternating minimization algorithm for the problem.
- We show that for well-conditioned systems, the pseudo-inverse solution is a good initialization.
- We analyze the convergence property of the algorithm.
- We show application to deconvolution of voiced speech signals.

Our Contribution

- We formulate the SBD problem using the MAP formulation.
- Propose an alternating minimization algorithm for the problem.
- We show that for well-conditioned systems, the pseudo-inverse solution is a good initialization.
- We analyze the convergence property of the algorithm.
- We show application to deconvolution of voiced speech signals.

Our Contribution

- We formulate the SBD problem using the MAP formulation.
- Propose an alternating minimization algorithm for the problem.
- We show that for well-conditioned systems, the pseudo-inverse solution is a good initialization.
- We analyze the convergence property of the algorithm.
- We show application to deconvolution of voiced speech signals.

Our Contribution

- We formulate the SBD problem using the MAP formulation.
- Propose an alternating minimization algorithm for the problem.
- We show that for well-conditioned systems, the pseudo-inverse solution is a good initialization.
- We analyze the convergence property of the algorithm.
- We show application to deconvolution of voiced speech signals.

Our Contribution

- We formulate the SBD problem using the MAP formulation.
- Propose an alternating minimization algorithm for the problem.
- We show that for well-conditioned systems, the pseudo-inverse solution is a good initialization.
- We analyze the convergence property of the algorithm.
- We show application to deconvolution of voiced speech signals.

MAP Formulation for BD problem

- **h**: deterministic but unknown.
- **e**: *i.i.d. generalized p-Gaussian distribution* (gpG).

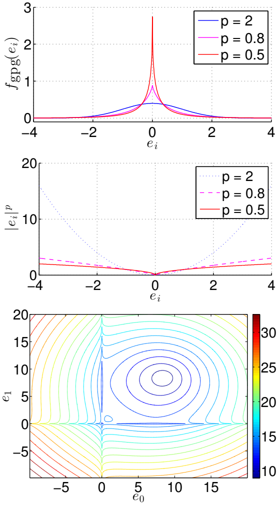
$$f(\mathbf{e}) = \left(\frac{p}{2\Gamma(1/p)\gamma\sigma_e}\right)^M \exp\left(-\sum_i \left(\frac{|e_i|}{\gamma\sigma_e}\right)^p\right),$$

where $\gamma = \left(\frac{\Gamma(1/p)}{\Gamma(3/p)}\right)^{1/2}$ and $0 \leq p \leq 1$ (Heavy-tailed distribution).

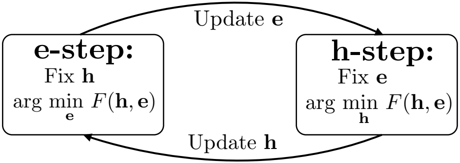
- The MAP estimates of the vectors **h** and **e** are

$$(\mathbf{h}_{\text{MAP}}, \mathbf{e}_{\text{MAP}}) = \arg\max_{\mathbf{h},\mathbf{e}} f(\mathbf{y}/\mathbf{e};\mathbf{h}) f(\mathbf{e}),$$
$$= \arg\min_{\mathbf{h},\mathbf{e}} \underbrace{\|\mathbf{y} - \mathbf{H}\mathbf{e}\|_2^2 + \lambda\|\mathbf{e}\|_p^p}_{F(\mathbf{h},\mathbf{e})}.$$

- The joint cost $F(\mathbf{h}, \mathbf{e})$ is non-convex and not straightforward to optimize.
- We resort to alternating minimization.



An Alternating $\ell_p - \ell_2$ Projections Algorithm (ALPA) I



e-step: $\mathbf{e}^{(k+1)} = \arg\min_{\mathbf{e}} F\left(\mathbf{h}^{(k)}, \mathbf{e}\right) = \arg\min_{\mathbf{e}} \|\mathbf{y} - \mathbf{H}^{(k)}\mathbf{e}\|_2^2 + \lambda\|\mathbf{e}\|_p^p,$

- For $0 \leq p < 1$, $F(\mathbf{h}^{(k)}, \mathbf{e})$ is non-convex.
- $\nabla_{\mathbf{e}} F\left(\mathbf{h}^{(k)}, \mathbf{e}\right)$ has a discontinuity at $\mathbf{e} = \mathbf{0}$.

An Alternating $\ell_p - \ell_2$ Projections Algorithm (ALPA) II

Iteratively reweighted least-squares (IRLS)

- Majorize $\|\mathbf{e}\|_p^p$ with weighted ℓ_2-norm function and minimize the cost iteratively.

$$\tilde{\mathbf{e}}^{(j+1,k)} = \arg\min_{\mathbf{e}} \underbrace{\|\mathbf{y} - \mathbf{H}^{(k)}\mathbf{e}\|_2^2 + \lambda \mathbf{e}^T \mathbf{W}^{j,k} \mathbf{e}}_{F_\epsilon(\mathbf{h}^{(k)},\mathbf{e})},$$

where $\mathbf{W}^{(j,k)} = \text{diag}\left(p \left((\tilde{e}_i^{(j,k)})^2 + \epsilon \right)^{p/2-1} \right).$

- Final **e-step solution** (after J iterations of IRLS): $\mathbf{e}^{(k+1)} = \tilde{\mathbf{e}}^{(J,k)}$,

$$\boxed{\mathbf{e}^{(k+1)} = \left((\mathbf{H}^{(k)})^T (\mathbf{H}^{(k)}) + \lambda \mathbf{W}^{(J,k)} \right)^{-1} (\mathbf{H}^{(k)})^T \mathbf{y},}$$

- Use $\mathbf{e}^{(k+1)}$ to update $\mathbf{h}^{(k+1)} = \arg\min_{\mathbf{h}} F_\epsilon\left(\mathbf{h}, \mathbf{e}^{(k+1)}\right)$.

h-step solution: $\boxed{\tilde{\mathbf{h}}^{(k+1)} = \arg\min_{\mathbf{h}} \|\mathbf{y} - \mathbf{E}^{(k+1)}\mathbf{h}\|_2^2 = \mathbf{E}^{(k+1)^\dagger}\mathbf{y},}$

$$\boxed{\mathbf{h}^{(k+1)} = \tilde{\mathbf{h}}^{(k+1)} / \|\tilde{\mathbf{h}}^{(k+1)}\|_2^2.}$$

Initialization: Pseudo-inverse Solution

- A good initialization might ensure we are in the "right" local minima.

Initialization: Pseudo-inverse Solution

- A good initialization might ensure we are in the “right” local minima.
- Initialization: Given some initial filter estimate $\tilde{\mathbf{h}}$, how good is the regularized least-squares solution ($\mathbf{W}^{(0,0)} = \mathbf{1}$)?

Initialization: Pseudo-inverse Solution

- A good initialization might ensure we are in the "right" local minima.
- Initialization: Given some initial filter estimate $\tilde{\mathbf{h}}$, how good is the regularized least-squares solution ($\mathbf{W}^{(0,0)} = \mathbf{1}$)?
- If true excitation and filter are $\mathbf{e}^*$ and $\mathbf{h}^*$ respectively, $\delta\mathbf{h} = \mathbf{h}^* - \tilde{\mathbf{h}}$ and pseudo-inverse solution is denoted $\hat{\mathbf{e}}_{\text{BLS}}$. The difference

$$\delta\mathbf{e}_{\text{BLS}} = \mathbf{e}^* - \hat{\mathbf{e}}_{\text{BLS}} = \mathbf{H}^{*\dagger}(\mathbf{w} - \delta\mathbf{H}\mathbf{e}^*),$$

where $\delta\mathbf{H} = \text{Conv. mtx}(\delta\mathbf{h})$.

Initialization: Pseudo-inverse Solution

- A good initialization might ensure we are in the "right" local minima.
- Initialization: Given some initial filter estimate $\tilde{\mathbf{h}}$, how good is the regularized least-squares solution ($\mathbf{W}^{(0,0)} = \mathbf{1}$)?
- If true excitation and filter are $\mathbf{e}^*$ and $\mathbf{h}^*$ respectively, $\delta\mathbf{h} = \mathbf{h}^* - \tilde{\mathbf{h}}$ and pseudo-inverse solution is denoted $\hat{\mathbf{e}}_{\text{BLS}}$. The difference

$$\delta\mathbf{e}_{\text{BLS}} = \mathbf{e}^* - \hat{\mathbf{e}}_{\text{BLS}} = \mathbf{H}^{*\dagger}(\mathbf{w} - \delta\mathbf{H}\mathbf{e}^*),$$

  where $\delta\mathbf{H} = \text{Conv. mtx}(\delta\mathbf{h})$.
- What is the probability of $\frac{1}{M}\|\delta\mathbf{e}_{\text{BLS}}\|_1$ exceeding a value ξ?

Initialization: Pseudo-inverse Solution

- A good initialization might ensure we are in the "right" local minima.
- Initialization: Given some initial filter estimate $\tilde{\mathbf{h}}$, how good is the regularized least-squares solution ($\mathbf{W}^{(0,0)} = \mathbf{1}$)?
- If true excitation and filter are $\mathbf{e}^*$ and $\mathbf{h}^*$ respectively, $\delta\mathbf{h} = \mathbf{h}^* - \tilde{\mathbf{h}}$ and pseudo-inverse solution is denoted $\hat{\mathbf{e}}_{\text{BLS}}$. The difference

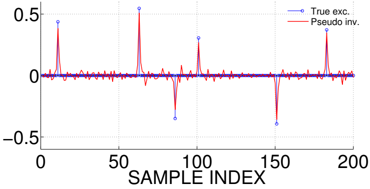
$$\delta\mathbf{e}_{\text{BLS}} = \mathbf{e}^* - \hat{\mathbf{e}}_{\text{BLS}} = \mathbf{H}^{*\dagger}(\mathbf{w} - \delta\mathbf{H}\mathbf{e}^*),$$

where $\delta\mathbf{H} = \text{Conv. mtx}(\delta\mathbf{h})$.

- What is the probability of $\frac{1}{M}\|\delta\mathbf{e}_{\text{BLS}}\|_1$ exceeding a value ξ?

$$P\left(\frac{1}{M}\|\delta\mathbf{e}_{\text{BLS}}\|_1 > \xi\right) \leq \frac{\kappa^2\sigma^2}{(\xi - \kappa\|\delta\mathbf{H}\|_2\|\mathbf{e}^*\|_2/\sqrt{M})^2}.$$

κ = condition number of $\mathbf{H}^*$.



Convergence Guarantees for ALPA

Behaviour of the majorized cost function $F_\epsilon(\mathbf{h}, \mathbf{e})$:

Convergence Guarantees for ALPA

Behaviour of the majorized cost function $F_\epsilon\left(\mathbf{h}, \mathbf{e}\right)$:

- After **e-step**,

$$F_\epsilon(\mathbf{h}^{(k)}, \mathbf{e}^{(k+1)}) \leq F_\epsilon\left(\mathbf{h}^{(k)}, \mathbf{e}^{(k)}\right).$$

Convergence Guarantees for ALPA

Behaviour of the majorized cost function $F_\epsilon(\mathbf{h}, \mathbf{e})$:

- After **e-step**,

$$F_\epsilon(\mathbf{h}^{(k)}, \mathbf{e}^{(k+1)}) \leq F_\epsilon\left(\mathbf{h}^{(k)}, \mathbf{e}^{(k)}\right).$$

- Similarly, after **h-step**,

$$F_\epsilon(\mathbf{h}^{(k+1)}, \mathbf{e}^{(k)}) \leq F_\epsilon\left(\mathbf{h}^{(k)}, \mathbf{e}^{(k)}\right).$$

Convergence Guarantees for ALPA

Behaviour of the majorized cost function $F_\epsilon(\mathbf{h}, \mathbf{e})$:

- After **e-step**,

$$F_\epsilon(\mathbf{h}^{(k)}, \mathbf{e}^{(k+1)}) \leq F_\epsilon\left(\mathbf{h}^{(k)}, \mathbf{e}^{(k)}\right).$$

- Similarly, after **h-step**,

$$F_\epsilon(\mathbf{h}^{(k+1)}, \mathbf{e}^{(k)}) \leq F_\epsilon\left(\mathbf{h}^{(k)}, \mathbf{e}^{(k)}\right).$$

- After one cycle of ALPA, (both **e-step** and **h-step**)

$$F_\epsilon(\mathbf{h}^{(k+1)}, \mathbf{e}^{(k+1)}) \leq F_\epsilon\left(\mathbf{h}^{(k)}, \mathbf{e}^{(k)}\right).$$

Convergence Guarantees for ALPA

Behaviour of the majorized cost function $F_\epsilon(\mathbf{h}, \mathbf{e})$:

- After **e-step**,

$$F_\epsilon(\mathbf{h}^{(k)}, \mathbf{e}^{(k+1)}) \leq F_\epsilon\left(\mathbf{h}^{(k)}, \mathbf{e}^{(k)}\right).$$

- Similarly, after **h-step**,

$$F_\epsilon(\mathbf{h}^{(k+1)}, \mathbf{e}^{(k)}) \leq F_\epsilon\left(\mathbf{h}^{(k)}, \mathbf{e}^{(k)}\right).$$

- After one cycle of ALPA, (both **e-step** and **h-step**)

$$F_\epsilon(\mathbf{h}^{(k+1)}, \mathbf{e}^{(k+1)}) \leq F_\epsilon\left(\mathbf{h}^{(k)}, \mathbf{e}^{(k)}\right).$$

Behaviour of the original Cost function $F(\mathbf{h}, \mathbf{e})$:

Convergence Guarantees for ALPA

Behaviour of the majorized cost function $F_\epsilon(\mathbf{h}, \mathbf{e})$:

- After **e-step**,

$$F_\epsilon(\mathbf{h}^{(k)}, \mathbf{e}^{(k+1)}) \leq F_\epsilon\left(\mathbf{h}^{(k)}, \mathbf{e}^{(k)}\right).$$

- Similarly, after **h-step**,

$$F_\epsilon(\mathbf{h}^{(k+1)}, \mathbf{e}^{(k)}) \leq F_\epsilon\left(\mathbf{h}^{(k)}, \mathbf{e}^{(k)}\right).$$

- After one cycle of ALPA, (both **e-step** and **h-step**)

$$F_\epsilon(\mathbf{h}^{(k+1)}, \mathbf{e}^{(k+1)}) \leq F_\epsilon\left(\mathbf{h}^{(k)}, \mathbf{e}^{(k)}\right).$$

Behaviour of the original Cost function $F(\mathbf{h}, \mathbf{e})$:

- With ϵ updated as $\epsilon^{(k)} = c\left(\max\left|e_j^{(k)}\right|\right)^{2-p}, \quad$ with $0 < c \ll 1$,

$$F\left(\mathbf{h}^{(k+1)}, \mathbf{e}^{(k+1)}\right) \leq F\left(\mathbf{h}^{(k)}, \mathbf{e}^{(k)}\right).$$

Deconvolution of Voiced Speech Signal

- 30 ms vowel segment /ae/ (female speaker),

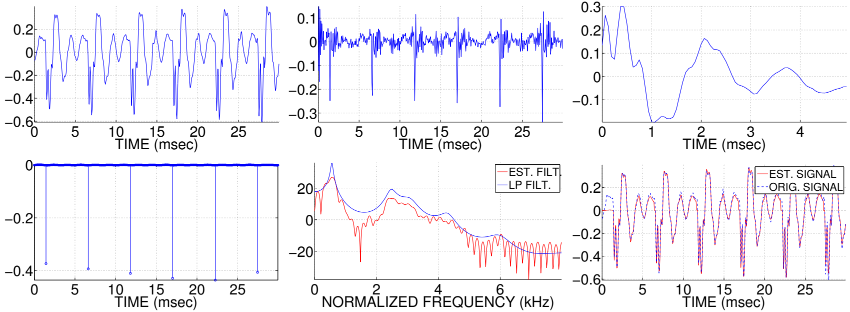


Figure: (a) Speech samples (sampling rate 16 kHz); (b) LP residue (MO 20); (c) ALPA estimate of the filter; (d) ALPA estimate of the excitation, (e) frequency response of the estimated filter; and (f) comparison between the original vowel segment shown in (a) and that synthesized based on the estimated filter and excitation.

Comparisons with state-of-the-art

Comparisons with Smooth ℓ_1/ℓ_2 blind deconvolution (SOOT)[1], sparse linear prediction (SLP)[2] and MM-based Sparse Deconvolution (SDMM)[3].

[1] A. Repetti et al. "Euclid in a Taxicab: Sparse Blind Deconvolution with Smoothed ℓ_1/ℓ_2 Regularization". In: *IEEE Signal Process. Lett.* 22.5 (2015), pp. 539–543. DOI: 10.1109/LSP.2014.2362861. eprint: 1407.5465.

[2] D. Giacobello et al. "Sparse Linear Prediction and Its Applications to Speech Processing". In: *IEEE Trans. Audio, Speech, Language Process.* 20.5 (2012), pp. 1644–1657. ISSN: 1558-7916. DOI: 10.1109/TASL.2012.2186807.

[3] Ivan Selesnick. *Sparse Deconvolution (An MM Algorithm)*. http://cnx.org/contents/f2738de6-b36d-458d-a2dd-2b50f375fe55@5/Sparse-Deconvolution-An-MM-Alg.

Comparisons with state-of-the-art

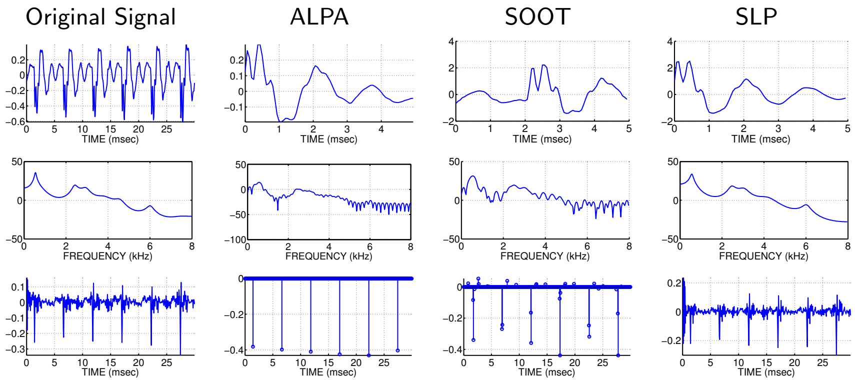


Figure: A comparison of sparse deconvolution methods for clean speech: Rows 1, 2, and 3 corresponding to Column 1 show the speech signal, frequency response of the LP filter, and the LP residue, respectively. For Columns 2−4, Rows 1, 2, and 3 show estimates of the filter, its frequency response, and the excitation, respectively.

Summary

Summary

- We considered the problem of sparse blind deconvolution.

Summary

- We considered the problem of sparse blind deconvolution.
- MAP formulation of the problem using super-Gaussian priors for a sparse solution.

Summary

- We considered the problem of sparse blind deconvolution.
- MAP formulation of the problem using super-Gaussian priors for a sparse solution.
- Resulting cost function is non-convex and was optimized using an alternating $\ell_p - \ell_2$ projections algorithm (ALPA).
- ALPA breaks the optimization problem into two convex sub problems for estimation of excitation and filter.

Questions?