A Non-Convex Approach to Blind Calibration for Linear Random Sensing Models

Dr. Valerio Cambareri
Prof. Laurent Jacques

Image and Signal Processing Group,
ICTEAM/ELEN, Université catholique de Louvain

iTWIST ’16, Aalborg, Denmark
Wednesday, August 24th 2016
Outline

- **Introduction**: Blind Calibration and Random Linear Models
- A Non-Convex Approach to Blind Calibration
- Solution by Projected Gradient Descent
- **Global Convergence Guarantees** (even if non-convex!)
- **Experiments I**: Empirical Phase Transition
- **Experiments II**: Computational Imaging
- **Conclusion** and outlook
Random Linear Sensing Model

\[ y_{i,l} = d_i \langle a_{i,l}, x \rangle, \quad d_i > 0, \quad x \in \mathbb{R}^n \]
\[ i = 1, \ldots, m, \quad l = 1, \ldots, p \]

\[ y^{(l)} = \text{diag}(d)A^{(l)}x, \quad d \in \mathbb{R}_+^m, \quad x \in \mathbb{R}^n \]
\[ l = 1, \ldots, p, \quad mp \geq n + m \]
Blind Calibration and Random Linear Models

Random Linear Sensing Model

\[ y_{i,l} = d_i \langle a_{i,l}, x \rangle, \quad d_i > 0, \quad x \in \mathbb{R}^n \]

\[ i = 1, \ldots, m, \quad l = 1, \ldots, p \]

\[ y^{(l)} = \text{diag}(d)A^{(l)}x, \quad d \in \mathbb{R}_+^m, \quad x \in \mathbb{R}^n \]

\[ l = 1, \ldots, p, \quad mp \geq n + m \]
Blind Calibration and Random Linear Models

Random Linear Sensing Model

\[ y_{i,l} = d_i \langle a_{i,l}, x \rangle, \quad d_i > 0, x \in \mathbb{R}^n \]
\[ i = 1, \ldots, m, \quad l = 1, \ldots, p \]

\[ y^{(l)} = \text{diag} (d) A^{(l)} x, \quad d \in \mathbb{R}_+^m, \quad x \in \mathbb{R}^n \]
\[ l = 1, \ldots, p, \quad mp \geq n + m \]

- Which algorithm to jointly recover \( x \) and \( d \)? Does it provably converge to the exact solution?
- How many snapshots and measurements suffice for an accurate recovery?
  Sample complexity bound (mp) required with i.i.d. sub-Gaussian random vectors \( a_{i,l} \)?
A Non-Convex Optimisation Problem

- We introduce the **Blind Calibration Problem**:

\[
(\hat{x}, \hat{d}) = \arg\min_{\xi \in \mathbb{R}^n, \gamma \in \mathbb{R}^m} \frac{1}{2mp} \sum_{l=1}^{p} \| \text{diag}(d) A^{(l)} x - \text{diag}(\gamma) A^{(l)} \xi \|_2^2
\]

(Scaled) probability simplex  
Sum of Euclidean data fidelity terms

- The problem is clearly *non-convex* (indefinite Hessian matrix in general): the sensing model is *bilinear*, while the problem is *biconvex* (similar to *blind deconvolution*).

- The objective has minima in: \( \{(\xi, \gamma) \in \mathbb{R}^n \times \mathbb{R}^m : \xi = 1/\alpha x, \gamma = \alpha d, \alpha \in \mathbb{R} \setminus \{0\}\} \)

- The constraint fixes one global minimiser: \((x^*, d^*) := \left( \frac{\|d\|_1}{m} x, \frac{m}{\|d\|_1} d \right)\)

- Since the gains are *positive* and *bounded*, we let:

\[
\gamma, d^* \in C_\rho \subset \mathbb{R}^m, \quad C_\rho := 1_m + 1_m^\perp \cap \rho \mathbb{B}_\infty^m
\]

\[
d^* = 1_m + \omega, \quad \omega \in 1_m^\perp \cap \rho \mathbb{B}_\infty^m
\]

\[
\gamma = 1_m + \epsilon, \quad \epsilon \in 1_m^\perp \cap \rho \mathbb{B}_\infty^m
\]

where:

\[
\rho > \|d^* - 1_m\|_\infty, \quad \rho < 1 \quad \text{(Perturbation analysis around 1!)}
\]
Solution by Projected Gradient Descent

The solution is obtained by *projected gradient descent*:

1. Initialise $\xi_0 := \frac{1}{mp} \sum_{l=1}^{p} (A_l)^\top y_l$, $\gamma_0 := 1_m$, $k := 0$.
2. **while** stop criteria not met **do**
   
   3. $\mu_\xi := \arg\min_{\nu \in \mathbb{R}} f(\xi_k - \nu \nabla_\xi f(\xi_k, \gamma_k), \gamma_k)$
   
   4. $\xi_{k+1} := \xi_k - \mu_\xi \nabla_\xi f(\xi_k, \gamma_k)$ \{Signal Update\}
   
   5. $\gamma_{k+1} := \gamma_k - \mu_\gamma \nabla_\gamma f(\xi_k, \gamma_k)$ \{Gain Update\}
   
   6. $\gamma_{k+1} := P_{C_\rho} \gamma_{k+1}$ \{Projection on $C_\rho$\}
   
   7. $k := k + 1$
3. **end while**

The chosen *initialisation* is crucial: in expectation (asymptotic $p$) it yields $x$ (unbiased estimator). For finite $p$ it can be shown to lie close to the *global minimum*.

Projection on $C_\rho$ is only a technical requirement for proofs (not required in experiments).
Consider a low-dimensional intuitive example for a random instance of the problem, at $n = 2, m = 2, \|\xi\|_2 = 1$, parametrising $\gamma(r)$.

To measure distances, we adopt the pre-metric:

$$\Delta(\xi, \gamma) := \|\xi - \mathbf{x}^*\|_2^2 + \frac{\|\mathbf{x}^*\|_2^2}{m} \|\gamma - \mathbf{d}^*\|_2^2.$$

Thus, we define a *neighbourhood* of the global minimiser as:

$$\mathcal{D}_{\kappa, \rho} := \{(\xi, \gamma) \in \mathbb{R}^n \times C_\rho : \Delta(\xi, \gamma) \leq \kappa^2 \|\mathbf{x}^*\|_2^2\}, \quad \rho \in [0, 1).$$
Convergence Guarantees

- Ideally: show via Hessian the local convexity of the problem in a given neighbourhood (for finite $p$, by concentration of measure).

- Simplification: first-order properties in the neighbourhood of the minimiser with i.i.d. sub-Gaussian random vectors. We need:
  1. **Initialisation**: fixes radius of neighbourhood, $(\xi_0, \gamma_0) \in D_{\kappa, \rho}, \rho \in [0, 1)$
  2. **Regularity Condition**: developing the distance at iterate $k+1$,

\[
\Delta(\xi_{k+1}, \gamma_{k+1}) = \Delta(\xi_k, \gamma_k) - 2 \left( \mu_\xi \langle \nabla \xi f(\xi_k, \gamma_k), \xi_k - x^* \rangle + \mu_\gamma \frac{\|x^*\|^2}{m} \langle \nabla \gamma f(\xi_k, \gamma_k), \gamma_k - g^* \rangle \right) + \mu_\xi^2 \| \nabla \xi f(\xi_k, \gamma_k) \|^2_2 + \mu_\gamma^2 \frac{\|x^*\|^2}{m} \| \nabla \gamma f(\xi_k, \gamma_k) \|^2_2 < \Delta(\xi_k, \gamma_k)
\]

3. **Projection** on convex set $C_\rho$ ensures $\Delta(\xi_{k+1}, \gamma_{k+1}) \leq \Delta(\xi_{k+1}, \gamma_{k+1})$

- Let’s have a look at the gradient:

\[
\nabla^\perp f(\xi, \gamma) = \frac{1}{mp} \sum_{l=1}^p \left[ A_l^T \text{diag} (\gamma) (\text{diag} (\gamma) A_l \xi - \text{diag} (d) A_l x) \right] \xrightarrow{p \rightarrow \infty} \frac{1}{m} \left[ \| \gamma \|^2_2 \xi - (\gamma^T d) x \right]
\]
To prove the minimum sample complexity that guarantees convergence, we need to verify two properties of the initialisation and the neighbourhood respectively.

**Proposition** (Initialisation Proximity). For any $\epsilon \in (0, 1)$ we have, with probability exceeding

$$1 - C e^{-c\epsilon^2 m\ell} - (m\ell)^{-t}$$

for some $C, c > 0$, that $\|\xi_0 - \mathbf{x}^*\|_2 \leq \epsilon \|\mathbf{x}^*\|_2$ provided $n \gtrsim t \log(m\ell)$ and

$$m\ell \gtrsim \epsilon^{-2}(n + m) \log \left( \frac{n}{\epsilon} \right).$$

Since $\gamma_0 = 1_m$ we also have $\|\gamma_0 - d^*\|_{\infty} \leq \rho < 1$. Thus $(\xi_0, \gamma_0) \in D_{\kappa, \rho}$ with the same probability and $\kappa := \sqrt{\epsilon^2 + \rho^2} \leq \sqrt{2}$. 

Initialise ...
**Proposition** (Regularity condition in $D_{\kappa, \rho}$). For any $\delta \in (0, 1), \rho \in [0, 1), t > 0$, provided $\rho < \frac{1-2\delta}{9}, \ n \gtrsim t \log(mp), \ p \gtrsim \delta^{-2} \log m$ and $\sqrt{mp} \gtrsim \delta^{-2}(n + m) \log\left(\frac{n}{\delta}\right)$, with probability exceeding

$$1 - C\left[me^{-c\delta^2\rho} + e^{-c\delta^2\sqrt{mp}} + (mp)^{-t}\right]$$

for some $C, c > 0$, we have that for all $(\xi, \gamma) \in D_{\kappa, \rho}$,

$$\left\langle \nabla^\perp f(\xi, \gamma), \left[\xi - x^* \atop \gamma - d^*\right]\right\rangle \geq \frac{1}{2} \eta \Delta(\xi, \gamma) \quad \text{(Bounded angle)}$$

$$\|\nabla^\perp f(\xi, \gamma)\|_2^2 \leq L^2 \Delta(\xi, \gamma) \quad \text{(Lipschitz gradient)}$$

for $\eta := 2(1 - 9\rho - 2\delta) > 0, \ L := 4\sqrt{2}[1 + \rho + (1 + \kappa)\|x^*\|_2]$.
Under the previous conditions, we can bound the \textit{error decay} of projected gradient descent when run in a neighbourhood of the global minimiser.

The projection step serves to ensure \textit{theoretically} that the neighbourhood does not change (\textit{i.e.}, for the regularity condition).

**Theorem** (Provable Convergence to the Exact Solution). \textit{Under the conditions of the previous Propositions we have that, with probability exceeding}

\[
1 - C \left[ m e^{-c \delta^2 p} + e^{-c \delta^2 \sqrt{mp}} + e^{-c e^2 mp} + (mp)^{-t} \right]
\]

\textit{for some} \( C, c > 0 \), \textit{our descent algorithm with} \( \mu_x := \mu, \mu_\gamma := \mu \frac{m}{\|x^*\|_2^2} \) \textit{has error decay}

\[
\Delta(\xi_k, \gamma_k) \leq (1 - \eta \mu + \frac{L^2}{\tau} \mu^2)^k (\epsilon^2 + \rho^2) \|x^*\|_2^2, (\xi_k, \gamma_k) \in D_{\kappa, \rho}
\]

\textit{at any iteration} \( k > 0 \) \textit{provided} \( \mu \in \left(0, \tau \eta/L^2\right)\), \( \tau := \min\{1, \|x^*\|_2^2/m\} \). \textit{Hence,}

\[
\Delta(\xi_k, \gamma_k) \xrightarrow{k \to \infty} 0.
\]
To test the problem’s phase transition we measure the probability of successful recovery

\[ P_\zeta := \mathbb{P} \left[ \max \left\{ \frac{||\hat{d} - d^*||_2}{||d^*||_2}, \frac{||\hat{x} - x^*||_2}{||x^*||_2} \right\} < \zeta \right], (x^*, d^*) \in \mathbb{R}^n \times \mathcal{C}_\rho, n = 2^8 \]

for 256 randomly generated problem instances (per point).

\( \rho \) increases \( (10^{-3} \rightarrow 10^{-1/2}) \)

---

Empirical Phase Transition

- To test the problem’s phase transition we measure the probability of successful recovery

\[ P_\zeta := \mathbb{P} \left[ \max \left\{ \frac{||\hat{d} - d^*||_2}{||d^*||_2}, \frac{||\hat{x} - x^*||_2}{||x^*||_2} \right\} < \zeta \right], (x^*, d^*) \in \mathbb{R}^n \times \mathcal{C}_\rho, n = 2^8 \]

for 256 randomly generated problem instances (per point).

\( \rho \) increases \( (10^{-3} \rightarrow 10^{-1/2}) \)
An application to computational (compressive) imaging under calibration errors yields the following results for $p = 4$ snapshots when $m = n = 4096$.

The achieved RMSE reads:

$$\max \left\{ \frac{\|\hat{d} - d^*\|_2}{\|d\|_2}, \frac{\|\hat{x} - x^*\|_2}{\|x\|_2} \right\} \approx -147.38 \text{ dB}$$

The algorithm (NC-BCP) scales gracefully to very large values of $n$, contrarily to other approaches with guarantees.

This experiment also converges with fast random matrices, such as a subsampled random convolution $A_i$ (not covered by current theory).
We have shown that a *simple* application of gradient descent provably solves this bilinear inverse problem with sample complexity:

$$\sqrt{mp} \gtrsim (n + m) \log n, \quad p \gtrsim \log m, \quad n \gtrsim \log mp$$

- **Proved extensions** of this approach:
  - Blind calibration with *known* subspace signal/gain models (lower sample complexity).
  - Stability analysis w.r.t. additive noise.
  - Better sample complexity is possible (*linear* in number of unknowns).

- **Future developments**:
  - Extension to signal-domain sparsity via hard thresholding: reduces sample complexity (*i.e.*, blind calibration for compressed sensing); empirically shown, not yet proved.
  - Extension to related problems: blind calibration with complex gains and sensing matrices; blind deconvolution.
  - Finding applications in which blind calibration of a sensor is critical and random measurements can be physically implemented in the sensing device.


Thank you for your attention.

For any question or suggestion, contact us at:
  laurent.jacques@uclouvain.be
  valerio.cambareri@uclouvain.be