A simple gradient descent algorithm for blind gain calibration of randomized sensing devices

Valerio Cambareri and Laurent Jacques
UCLouvain, Belgium

October 6th, 2016. INRIA, Rennes, France
Compressed Sensing & Random Linear Models

$M$ questions

$y$

Sensing method

$A$

Signal

$x$

$\sim \text{noise}$

$m$

$m \times n$

$n$

low-complexity signal (e.g., sparse, compressible, low-rank)
Compressed Sensing & Random Linear Models

\[ y_i \sim \langle a_i, x \rangle = a_i^T x \quad 1 \leq i \leq m \]

Generalized Linear Sensing!

\( M \) questions \hspace{1cm} Sensing method \hspace{1cm} Signal

\( y \)

\( y_i \)

\( m \)

\( m \times n \)

low-complexity signal (e.g., sparse, compressible, low-rank)
Blind Calibration and Random Linear Models

\[ y_i \sim \langle a_i, x \rangle = a_{i}^{T} x \]

additive noise
Blind Calibration and Random Linear Models

\[ y_i \sim \langle a_i, x \rangle = a_i^T x \]

\[ \text{additive noise} \]

\[ \text{what if unknown gains?} \]
Blind Calibration and Random Linear Models

$$y_i \approx \langle a_i, x \rangle = a_i^T x$$

additive noise ✓
what if unknown gains?

Blind Calibration Problem:

Recover $x$ (signal) and $d$ (gains) in

$$y = \text{diag}(d) Ax + \eta \text{ with } d_i \approx 1$$

unknown noise

Recent related works:

- Blind calibration: [Friedlander, Strohmer, 14] [Li, Ling, Strohmer, 16]
- Blind deconvolution: [Ali, Rech, Romberg, 14], [Bilen, 14] [Li, Ling, Strohmer, 16]
Blind Calibration and Random Linear Models

\[ y_i \sim \langle a_i, x \rangle = a_i^T x \]

\[ \text{additive noise} \quad \checkmark \]

\[ \text{what if unknown gains?} \]

Blind Calibration Problem: our approach

Recover \( x \) (signal) and \( d \) (gains) in

\[ y_l = \text{diag}(d) A_l x + \eta, \quad 1 \leq l \leq p \]

with random sensing model:

\[ A_l \sim_{\text{iid}} A \in \mathbb{R}^{m \times n}, \]

with \( A_{ij} \) sub-Gaussian, zero mean & unit variance.

(e.g., Gaussian, Bernoulli, Bounded)
Blind Calibration and Random Linear Models

\[ y_i \sim \langle a_i, x \rangle = a_i^T x \]

- additive noise
- what if unknown gains?

Blind Calibration Problem: our approach

Recover \( x \) (signal) and \( d \) (gains) in

\[ y_l = \text{diag}(d) A_l x + \eta, \quad 1 \leq l \leq p \]

Inspirations:
- Programmable Compressive Imagers
- Rice single pixel camera (Baraniuk, Kelly et al)
- Coded aperture CS imagers (CASSI, Brady et al)
Blind Calibration and Random Linear Models

\[ y_i \sim \langle a_i, x \rangle = a_i^T x \]

- additive noise
- what if unknown gains?

Blind Calibration Problem: 

Recover \( x \) (signal) and \( d \) (gains) in

\[ y_l = \text{diag}(d) A_l x + \eta, \quad 1 \leq l \leq p \]

**Central questions:**

- Efficient algorithm?
- Minimal sample complexity: \( mp \) ?
- Minimal snapshot number: \( p \) ?
- Robustness vs \( \eta \) ?

(for sub-Gaussian \( A_l \))
Intrinsic ambiguity (in noiseless case)

- Let $S := \{(x', d') : \text{diag}(d') A_l x' = \text{diag}(d) A_l x = y_l, 1 \leq l \leq p\}$
- Scaling ambiguity:
  $$(x^*, d^*) \in S \iff \forall \alpha \neq 0, (\frac{1}{\alpha} x^*, \alpha d^*) \in S !$$
Intrinsic ambiguity (in noiseless case)

- Let $S := \{(x', d') : \text{diag}(d')A_l x' = \text{diag}(d)A_l x = y_l, 1 \leq l \leq p\}$
- Scaling ambiguity:
  $$(x^*, d^*) \in S \iff \forall \alpha \neq 0, \left(\frac{1}{\alpha}x^*, \alpha d^*\right) \in S$$

Our context:
- Gain calibration: $0 \leq d_i \approx 1, 1 \leq i \leq m$
- Let’s assume (wlog):
  $$\sum_i d_i = m,$$
  or $d \in \Pi_m^+ = \{w \in \mathbb{R}_+^m : 1^T_m w = \sum_i w_i = m\}$

(Scaled) probability simplex
Intrinsic ambiguity (in noiseless case)

- Let \( S := \{(x', d') : \text{diag}(d')A_l x' = \text{diag}(d)A_l x = y_l, 1 \leq l \leq p\} \)
- Scaling ambiguity:
  \[(x^*, d^*) \in S \iff \forall \alpha \neq 0, (\frac{1}{\alpha} x^*, \alpha d^*) \in S! \]

Our context:

- Gain calibration: \( 0 \leq d_i \approx 1, 1 \leq i \leq m \)
- Let’s assume (wlog):
  \[ \sum_i d_i = m, \]
  or \( d \in \Pi_m^+ = \{w \in \mathbb{R}_+^m : 1^T_m w = \sum_i w_i = m\} \)
  + perturbation analysis: \( |d_i - 1| \leq \rho < 1 \)
  (for some \( 0 \leq \rho < 1 \))
  \[ \Rightarrow d \in 1 + \rho B_m^\infty \]

\[ \Rightarrow \text{We define } C_\rho := \Pi_m^+ \cap (1 + \rho B_m^\infty) \text{ our optimization space!} \]
A Non-Convex Optimisation Problem

- Blind Calibration Problem:
  \[
  (\hat{x}, \hat{d}) = \arg\min_{\xi \in \mathbb{R}^n, \gamma \in C_\rho} \frac{1}{2mp} \sum_{l=1}^p \| \text{diag}(d) A_l x - \text{diag}(\gamma) A_l \xi \|^2_{y_l}
  \]

- Non-convex (bi-convex) but maybe locally convex?
- Idea: initialize + (projected) gradient descent

  (as in Phase-Retrival via Wirtinger flow,
  e.g., [Candès, Li, 2015] [White et al., 2015]
  [Ling, Strohmer, Wei, 2016])
Geometric Analysis

- Low-dimensional intuitive example:

  \[ \gamma, \xi \in \mathbb{R}^2, \text{i.e., } n = m = 2, \]
  \[ \|\xi\| = 1, \quad \gamma = (1 + r, 1 - r) \in \Pi_2^+, \quad r \in \mathbb{R} \]
  \[ \rightarrow \text{Optimization space: } (\xi_1, \xi_2, r) \text{ on a cylinder.} \]

  We study the variations of
  \[ f(\xi, \gamma) := \frac{1}{2mp} \sum_{l=1}^p \| y_l - \text{diag}(\gamma) A_l \xi \|^2 \]
  around \((x_1, x_2, \rho) = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0.08)\)

\[ \log f(\xi, \gamma), \quad p = 1 \]
Geometric Analysis

- Low-dimensional intuitive example:

\[ \gamma, \xi \in \mathbb{R}^2, \text{ i.e., } n = m = 2, \]
\[ \|\xi\| = 1, \gamma = (1 + r, 1 - r) \in \Pi_2^+, r \in \mathbb{R} \]

\[ \rightarrow \text{Optimization space: } (\xi_1, \xi_2, r) \text{ on a cylinder.} \]

We study the variations of \( f(\xi, \gamma) := \frac{1}{2mp} \sum_{l=1}^{p} \|y_l - \text{diag}(\gamma)A_l\xi\|^2 \)
around \((x_1, x_2, \rho) = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0.08\right)\)

\[ \log f(\xi, \gamma), \ p = 1 \]
\[ \log f(\xi, \gamma), \ p = 2 \]
\[ \log f(\xi, \gamma), \ p = 3 \]
\[ \log \mathbb{E} f(\xi, \gamma) \]

Increasing \(p\)

\(p = \infty\)
Geometric Analysis

- Conclusion:
  - Hope for *local convexity*
    in a neighborhood (an “ellipsoid” of radius $\kappa$)

$$
\mathcal{D}_{\kappa, \rho} := \{ (\xi, \gamma) \in \mathbb{R}^n \times \mathcal{C}_\rho : \Delta(\xi, \gamma) \leq \kappa^2 \| x^* \|_2^2 \}
$$

with distance

$$
\Delta(\xi, \gamma) := \| \xi - x^* \|_2^2 + \frac{\| x^* \|_2^2}{m} \| \gamma - d^* \|_2^2
\approx_{\rho} 2 \mathbb{E} f(\xi, \gamma) \text{ if } \gamma \in \mathcal{C}_\rho.
$$
Solution by Projected Gradient Descent

Algorithm:

1. Initialize $\xi_0 := \frac{1}{mp} \sum_{l=1}^{p} (A_l)^\top y_l$, $\gamma_0 := 1_m$, $k := 0$.  

\[
 f(\xi, \gamma) := \frac{1}{2mp} \sum_{l=1}^{p} \|y_l - \text{diag}(\gamma) A_l \xi\|^2
\] 

(almost dumb ...
Solution by Projected Gradient Descent

Algorithm:

1. Initialize $\xi_0 := \frac{1}{mp} \sum_{l=1}^{p} (A_l)_{\top} y_l$, $\gamma_0 := 1_m$, $k := 0$. (almost dumb ...

... but not so bad initialization!)

Prop. Let $0 < \delta < 1$, $t > 0$, and define $\kappa^2 := \delta^2 + \rho^2$. If $\text{mp} \gtrsim \delta^{-2}(m + n) \log(n/\delta)$ and $n \gtrsim t \log(mp)$, then $(\xi_0, \gamma_0) \in D_{\kappa, \rho}$, with prob. failure $\lesssim e^{-c \delta^2 mp} + (mp)^{-t}$ ($c > 0$).

$\Rightarrow \|\xi_0 - x^*\|^2 + \frac{\|x^*\|^2}{m} \|\gamma_0 - d^*\|^2 \leq \kappa^2 \|x^*\|^2$
Solution by Projected Gradient Descent

\[ f(\xi, \gamma) := \frac{1}{2mp} \sum_{l=1}^{p} \|y_l - \text{diag}(\gamma) A_l \xi \|^2 \]

**Algorithm:**

1. Initialize \( \xi_0 := \frac{1}{mp} \sum_{l=1}^{p} (A_l)^\top y_l, \) \( \gamma_0 := 1_m, \) \( k := 0. \)

(for some step sizes \( \mu_\xi, \mu_\gamma > 0 \))

2. while stop criteria not met do

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. \( k := k + 1 \)

8. end while

\( \nabla_\gamma f(\xi, \gamma) := P_{1_m^\perp} \nabla_\gamma f(\xi, \gamma) \)

technical requirement for proofs
(not required in experiments)
Solution by Projected Gradient Descent

Algorithm:

1: Initialize $\xi_0 := \frac{1}{mp} \sum_{l=1}^{p} (A_l)^\top y_l$, $\gamma_0 := 1_m$, $k := 0$. 
2: while stop criteria not met do 
   3: $\mu_\xi := \text{argmin}_{\nu \in \mathbb{R}} f(\xi_k - \nu \nabla_\xi f(\xi_k, \gamma_k), \gamma_k)$ 
   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: $\gamma_{k+1} := PC_\rho \gamma_{k+1}$ \{Projection on $C_\rho$\} 
   7: $k := k + 1$ 
8: end while

$f(\xi, \gamma) := \frac{1}{2mp} \sum_{l=1}^{p} \|y_l - \text{diag}(\gamma)A_l \xi\|^2$

$\nabla_\gamma f(\xi, \gamma) := P_{1_m}^\perp \nabla f(\xi, \gamma)$

technical requirement for proofs (not required in experiments)
Solution by Projected Gradient Descent

Algorithm:

1: Initialize $\xi_0 := \frac{1}{mp} \sum_{l=1}^{p} (A_l)^\top y_l$, $\gamma_0 := 1_m$, $k := 0$.

(for some step sizes $\mu_\xi, \mu_\gamma > 0$)

2: while stop criteria not met do

4: $\xi_{k+1} := \xi_k - \mu_\xi \nabla_\xi f(\xi_k, \gamma_k)$ \{Signal Update\}

5: $\gamma_{k+1} := \gamma_k - \mu_\gamma \nabla_\gamma \perp f(\xi_k, \gamma_k)$ \{Gain Update\}

6: $\gamma_{k+1} := P_{C_\rho} \gamma_{k+1}$ \{Projection on $C_\rho$\}

7: $k := k + 1$

8: end while

Convergence to $(x^*, d^*)$?

$\Delta(\xi_{k+1}, \gamma_{k+1}) \leq \Delta(\xi_{k+1}, \gamma_{k+1}) \leq \Delta(\xi_k, \gamma_k)$ ? (at any $k \geq 0$)

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

... we need to prove regularity on angles and magnitudes!
Solution by Projected Gradient Descent

Algorithm:

1. Initialize \( \xi_0 := \frac{1}{mp} \sum_{l=1}^{p} (A_l)\top y_l \), \( \gamma_0 := 1_m \), \( k := 0 \).

(for some step sizes \( \mu_\xi, \mu_\gamma > 0 \))

2. while stop criteria not met do

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. \( k := k + 1 \)

8. end while

Convergence to \((x^*, d^*)\)?

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

\[
\text{(must be } > 0) \quad \text{Gradient Angle Part}
\]

\[
\text{Gradient Magnitude Part} \quad (\text{must bounded})
\]

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

... we need to prove regularity on \textit{angles} and \textit{magnitudes}!
Regularity condition in $D_{\kappa,\rho}$

**Prop.** Let $0 < \delta < 1$, $t > 0$, and define $\kappa^2 := \delta^2 + \rho^2$. If $n \gtrsim t \log(mp)$,

$$mp \gtrsim \delta^{-2}(m+n)\log(n/\delta) \quad \text{and} \quad p \gtrsim \delta^{-2}\log m,$$

and if

$$\rho < \frac{1}{9}(1 - 2\delta),$$

then, $\exists \eta, L > 0$ (only depending on $\delta$ and $\rho$) such that, $\forall (\xi, \gamma) \in D_{\kappa,\rho}$,

1. $\left\langle \nabla^\perp f(\xi, \gamma), \left[ \frac{\xi - x^*}{\gamma - d^*} \right] \right\rangle \geq \frac{1}{2} \eta \Delta(\xi, \gamma)$ (Bounded angle)
2. $\|\nabla^\perp f(\xi, \gamma)\|^2 \leq L^2 \Delta(\xi, \gamma)$ (Lipschitz gradient)

with prob. failure $\lesssim e^{-c\delta^2 mp} + e^{-c'\delta^2 p} + (mp)^{-t}$ (for some $c, c' > 0$).

*Proof ingredients*:

Measure concentration on sub-Gaussian r.v., Matrix Bernstein inequality, non-uniformity wrt $x^*$ and $d^*$.
Convergence to \((x^*, d^*)\)?

- **Regularity condition in** \(D_{\kappa, \rho}\)

**Prop.** Let \(0 < \delta < 1\), \(t > 0\), and define \(\kappa^2 := \delta^2 + \rho^2\). If \(n \gtrsim t \log(mp)\),

\[
mp \gtrsim \delta^{-2}(m + n) \log(n/\delta)
\quad \text{and} \quad
p \gtrsim \delta^{-2} \log m,
\]

and if

\[
\rho < \frac{1}{9}(1 - 2\delta),
\]

then, \(\exists \eta, L > 0\) (only depending on \(\delta\) and \(\rho\)) such that, \(\forall (\xi, \gamma) \in D_{\kappa, \rho}\),

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

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

with prob. failure \(\lesssim e^{-c\delta^2 mp} + e^{-c'\delta^2 p} + (mp)^{-t}\) (for some \(c, c' > 0\)).

\[\Rightarrow \|\nabla^\perp f(\xi, \gamma)\| \neq 0 \text{ but on the solution in } D_{\kappa, \rho}! \]

\[\Rightarrow \text{allows convergence for } \mu_\gamma = \mu_\xi \frac{m}{\|x^*\|}.\]
**Theorem.** Let $0 < \delta < 1$, $t > 0$, and define $\kappa^2 := \delta^2 + \rho^2$. If

\[
 n \gtrsim t \log(mp), \quad mp \gtrsim \delta^{-2}(m + n) \log(n/\delta) \quad \text{and} \quad p \gtrsim \delta^{-2} \log m,
\]

and if

\[
 \rho < \frac{1}{9}(1 - 2\delta),
\]

then, $\exists \eta, L > 0$ (only depending on $\delta$ and $\rho$) such that, with probability exceeding

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

for some $C, c > 0$, our descent algorithm initialized on $(\xi_0, \gamma_0)$ with $\mu_\xi = \mu$ and $\mu_\gamma = \mu \frac{m}{\|x^*\|_2}$

gives jointly, at each iteration $k$,

\[
 (\xi_k, \gamma_k) \in \mathcal{D}_{\kappa, \rho} \quad \text{and} \quad \Delta(\xi_k, \gamma_k) \leq (1 - \eta \mu + \frac{L^2}{\tau} \mu^2)^k \kappa^2 \|x^*\|_2^2,
\]

provided $\mu \in (0, \frac{\eta \|x^*\|_2^2}{mL^2 + \|x^*\|_2^2L^2})$. Hence, $\Delta(\xi_k, \gamma_k) \xrightarrow{k \to \infty} 0$.

Roughly speaking, for $\rho$ small enough,

we need $n$ and $p > 1$ large enough,

and $mp \gtrsim (m + n) \log(n/\delta)$ observations.
Empirical Phase Transition

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

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

for 256 randomly generated problem instances (per point).

\[ \rho \text{ increases } (10^{-3} \rightarrow 10^{-1/2}) \]
Empirical Phase Transition

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

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

for 256 randomly generated problem instances (per point).

$\rho$ increases ($10^{-3} \rightarrow 10^{-1/2}$)
(Randomized) Computational Imaging

Imaging you must recalibrate an imager that is far far away?

\[
\hat{x} \leftarrow k_1 \hat{x} + r(k_2 y_2) \leftarrow m_0 \nabla \theta_0 \leftarrow m_0 \|
\]

\[
\nabla \theta_0 = \left\{ \nabla \theta_0 \leftarrow m_0 \right\}
\]

\[
\nabla \theta_0 = \left\{ \nabla \theta_0 \leftarrow m_0 \right\}
\]

Fixed signal \( \hat{x} \)

\[
\mathrm{Pluto}
\]

\[
\mathrm{(NewHorizon \ 2015)}
\]

\[
\mathrm{e.g.,}
\]

\[
\mathrm{e.g.,}
\]

\[
\mathrm{e.g.,}
\]

\[
\mathrm{e.g.,}
\]
(Randomized) Computational Imaging

- Computational (compressive) imaging under calibration errors for $p = 4$ snapshots when $m = n = 4096$.
  (with Gaussian random matrices)

- LS SNR: 5.5 dB on signal
(Randomized) Computational Imaging

- Computational (compressive) imaging under calibration errors for $p = 4$ snapshots when $m = n = 4096$.
  (with Gaussian random matrices)

- LS SNR: 5.5 dB on signal

- PGD:
  min. gain/signal SNR = 147.38 dB

- PGD c. time: 2’ here
  and still ok for large $n$
  (in paper: 40’ for n=16384, m=1024, p=32)

- Also converges with fast and structured random matrices $A_i$
  (e.g., subsampled random convolution, spread-spectrum)
  (not covered by current theory).
Preliminary extension to sparse signals (ICASSP’17)

- Assumption: $\mathbf{x}$ is $k$-sparse in an ONB $\Psi$
  \[ (i.e., |\text{supp } \Psi^\top \mathbf{x}| =: \|\Psi^\top \mathbf{x}\|_0 \leq k) \]

- Hard thresholding “projection”:
  \[ \mathcal{H}_k(\mathbf{u}) := \text{argmin}_\mathbf{v} \|\mathbf{u} - \mathbf{v}\| \text{ s.t. } \|\mathbf{v}\|_0 = k \]
Preliminary extension to sparse signals (ICASSP’17)

- **Assumption:** \( \mathbf{x} \) is \( k \)-sparse in an ONB \( \Psi \)
  \[ (i.e., |\text{supp} \Psi^\top \mathbf{x}| =: \|\Psi^\top \mathbf{x}\|_0 \leq k) \]

- **Hard thresholding “projection”:**
  \[ \mathcal{H}_k(\mathbf{u}) := \arg\min_{\mathbf{v}} \|\mathbf{u} - \mathbf{v}\| \text{ s.t. } \|\mathbf{v}\|_0 = k \]

- **Objective:** (under similar assumptions: \( d \in \Pi_m^+, \|d - 1\|_\infty \leq \rho \))
  Recover \((\mathbf{x}, d)\) from \( \{\mathbf{y}_l = \text{diag}(d)A_l\mathbf{x} : 1 \leq p \leq p\} \)

- **Hope:** sample complexity should be like
  \[ mp \gtrsim m + k \Rightarrow m(p - 1) \gtrsim k \]
Preliminary extension to sparse signals (ICASSP’17)

- **Assumption:** $x$ is $k$-sparse in an ONB $\Psi$
  \[ \text{i.e., } |\text{supp } \Psi^\top x| =: \|\Psi^\top x\|_0 \leq k \]

- **Hard thresholding “projection”:**
  \[ \mathcal{H}_k(u) := \arg\min_v \|u - v\| \text{ s.t. } \|v\|_0 = k \]

- **Blind Calibration with Iterative Hard Thresholding**

  1. Initialize $\xi_0 := \frac{1}{mp} \sum_{l=1}^{p} (A_l)^\top y_l, \gamma_0 := 1_m, j := 0.$
  2. while stop criteria not met do
  3. \[ \xi_{j+1} := \Psi \mathcal{H}_k[\Psi^\top (\xi_j - \mu_\xi \nabla_\xi f(\xi_j, \gamma_j))] \text{ \{Signal Update\}} \]
  4. \[ \gamma_{j+1} := \gamma_j - \mu_\gamma \nabla_\gamma f(\xi_j, \gamma_j) \text{ \{Gain Update\}} \]
  5. \[ \gamma_{j+1} := P_{C_\rho} \gamma_{j+1} \text{ \{Projection on } C_\rho\} \]
  6. \[ j := j + 1 \]
  7. end while
Empirical Phase Transition (bis)

\[ n = 2^{10} = 1024, \, k = 2^5 = 32, \, m \in [2k, 32k = n] \]
Empirical Phase Transition (bis)

\[ n = 2^{10} = 1024, \quad k = 2^5 = 32, \quad m \in [2k, 32k = n] \]

\[ \log_2 mp = \log_2(C(m + k)) \]
“Tous les jours”, by René Magritte

\[ \Psi = \text{Daubechies-4}, \ k = 1800, \ n = 256^2, \ m = 103^2, \ \frac{m}{k} \approx 6 \text{ and } p = 5 \]

(sparsified) true signal

(a) True signal \(x\), \(n = 256 \times 256\) px

“Tous les jours”, René Magritte, 1966 (Charly Herscovici 2011, Wikiart.org)
“Tous les jours”, by René Magritte

\[
\Psi = \text{Daubechies - 4}, \ k = 1800, \ n = 256^2, \ m = 103^2, \ \frac{m}{k} \approx 6 \ \text{and} \ p = 5
\]

(sparsified) true signal

IHT

(a) True signal \( \mathbf{x} \), \( n = 256 \times 256 \) px

(b) Recovery \( \hat{\mathbf{x}} \) provided by IHT, 
\[ \text{RSNR}_{\mathbf{x}, \hat{\mathbf{x}}} = 17.83 \text{ dB} \]

“Tous les jours”, René Magritte, 1966 (Charly Herscovici 2011, Wikiart.org)
“Tous les jours”, by René Magritte

\[ \Psi = \text{Daubechies-4, } k = 1800, \quad n = 256^2, \quad m = 103^2, \quad \frac{m}{k} \approx 6 \quad \text{and} \quad p = 5 \]

(sparsified) true signal

IHT

BC-IHT

(a) True signal \( \mathbf{x} \), \( n = 256 \times 256 \) px

(b) Recovery \( \hat{\mathbf{x}} \) provided by IHT, \( \text{RSNR}_{\mathbf{x}, \hat{\mathbf{x}}} = 17.83 \) dB

(c) Recovery \( \hat{\mathbf{x}} \) provided by BC-IHT, \( \text{RSNR}_{\mathbf{x}, \hat{\mathbf{x}}} = 153.16 \) dB

(d) True gains \( \mathbf{g} \), \( m = 103 \times 103 \) px

(e) Recovery \( \hat{\mathbf{g}} \) provided by BC-IHT, \( \text{RSNR}_{\mathbf{g}, \hat{\mathbf{g}}} = 122.76 \) dB

“Tous les jours”, René Magritte, 1966 (Charly Herscovici 2011, Wikiart.org)
Conclusion

- We have shown that a simple application of gradient descent provably solves this bilinear inverse problem with sample complexity:

\[ mp \gtrsim (n + m) \log n, \ p \gtrsim \log m, \ n \gtrsim \log mp \]

(*: note: it was “(\sqrt{m})p \gtrsim (n + m) \log(n)” in our CoSeRa’16 paper)

- Proved extension of this approach:
  - Stability analysis w.r.t. additive noise, in fact:
    \[ \Delta(\xi_k, \gamma_k) \xrightarrow[k \to \infty]{} C \|\text{noise}\|^2 \]
  - Known subspaces on signal and gains (no shown here)
Conclusion

- We have shown that a simple application of gradient descent provably solves this bilinear inverse problem with sample complexity:

\[ mp \gtrsim (n + m) \log n, \quad p \gtrsim \log m, \quad n \gtrsim \log mp \]

(*: note: it was “\((\sqrt{m})p \gtrsim (n + m) \log(n)\)” in our CoSeRa’16 paper)

- **Proved extension** of this approach:
  - Stability analysis w.r.t. additive noise, in fact:
    \[ \Delta(\xi_k, \gamma_k) \xrightarrow{k \to \infty} C||\text{noise}||^2 \]
  - Known subspaces on signal and gains (no shown here)

- **Connections with other works:** e.g., [Li, Ling, Strohmer, 16]

- **Future developments:**
  - Extension to signal-domain *sparsity* via hard thresholding: reduces sample complexity (*i.e.*, blind calibration for compressed sensing); **empirically shown** (+ conf paper), not yet proved.
  - More advanced calibration? (e.g., through matrix probing).
Thank you for you attention!

Bibliography


