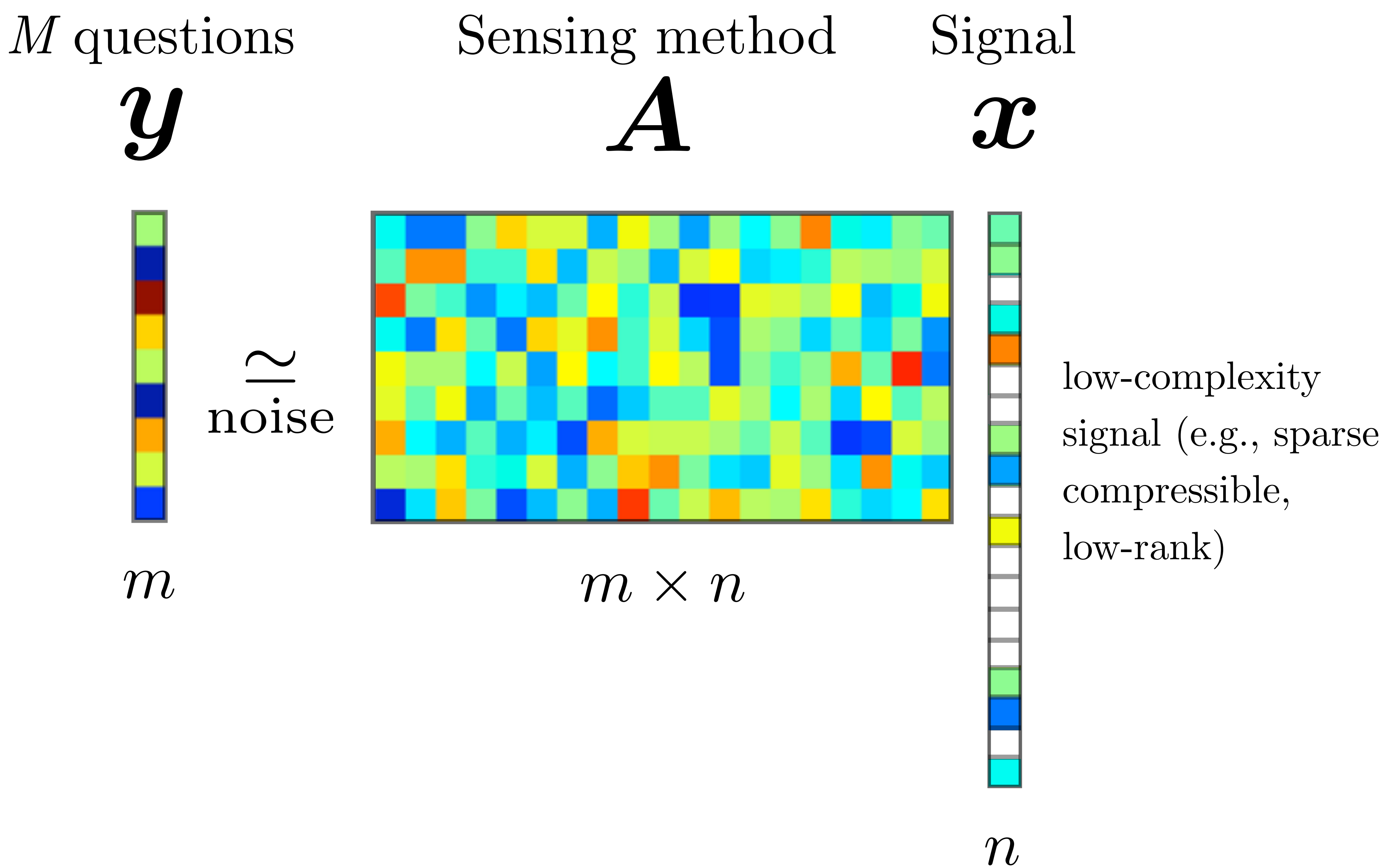
A simple gradient descent algorithm for blind gain calibration of randomized sensing devices

Valerio Cambareri and Laurent Jacques
UCLouvain, Belgium

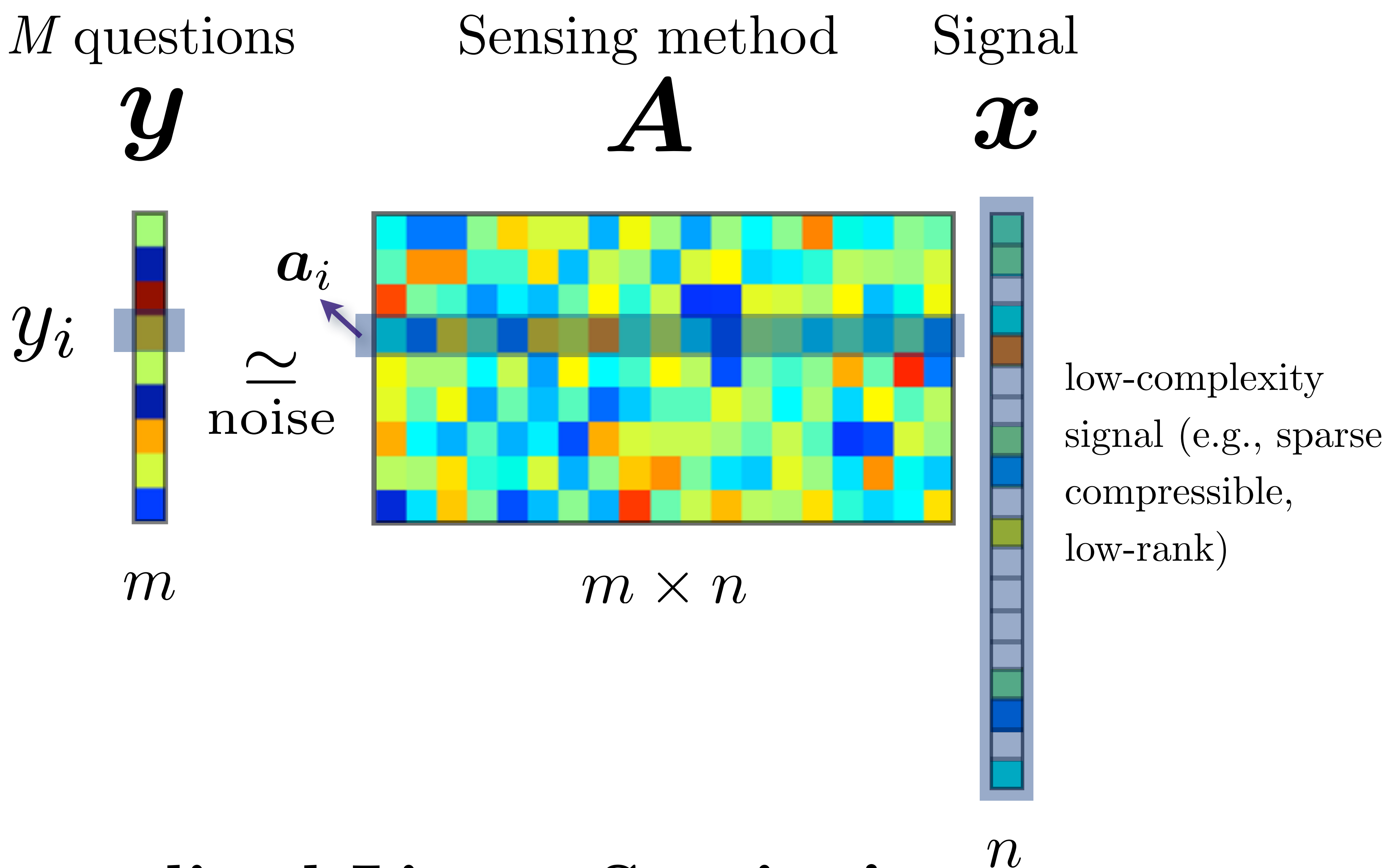
October 6th, 2016. INRIA, Rennes, France

UCL Université catholique de Louvain
fnrs LA LIBERTÉ DE CHERCHER

Compressed Sensing & Random Linear Models



Compressed Sensing & Random Linear Models



Generalized Linear Sensing!

$$y_i \simeq \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle = \boldsymbol{a}_i^T \boldsymbol{x} \quad 1 \leq i \leq m$$

Blind Calibration and Random Linear Models

$$y_i \simeq \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle = \boldsymbol{a}_i^T \boldsymbol{x}$$

additive noise ✓

Blind Calibration and Random Linear Models

$$y_i \simeq \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle = \boldsymbol{a}_i^T \boldsymbol{x}$$

→ additive noise ✓

→ what if unknown gains?

Blind Calibration and Random Linear Models

$$y_i \simeq \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle = \boldsymbol{a}_i^T \boldsymbol{x}$$

→ additive noise ✓

→ what if unknown gains?

Blind Calibration Problem:

Recover $\boldsymbol{x}$ (signal) and $\boldsymbol{d}$ (gains) in

$$\boldsymbol{y} = \mathrm{diag}(\boldsymbol{d})\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\eta} \quad \text{with } d_i \approx 1$$

unknown ($\mathrm{diag}(\boldsymbol{d})$, $\boldsymbol{x}$) noise ($\boldsymbol{\eta}$)

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

$$y_i \simeq \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle = \boldsymbol{a}_i^T \boldsymbol{x}$$

additive noise ✓

what if unknown gains?

Blind Calibration Problem: our approach

Recover $\boldsymbol{x}$ (signal) and $\boldsymbol{d}$ (gains) in

$$\boldsymbol{y}_l = \mathrm{diag}(\boldsymbol{d})\, \boldsymbol{A}_l \boldsymbol{x} + \boldsymbol{\eta},\ 1 \leqslant l \leqslant p$$

Multiple "snapshots"

with random sensing model:

$\boldsymbol{A}_l \sim_{\mathrm{iid}} \boldsymbol{A} \in \mathbb{R}^{m \times n}$,
with A_{ij} sub-Gaussian, zero mean & unit variance.

(e.g., Gaussian, Bernoulli, Bounded)

Blind Calibration and Random Linear Models

$$y_i \simeq \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle = \boldsymbol{a}_i^T \boldsymbol{x}$$

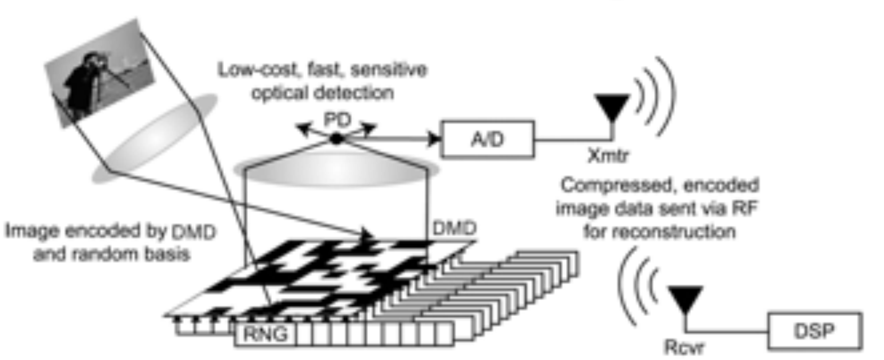
- additive noise ✓
- what if unknown gains?

Blind Calibration Problem: our approach

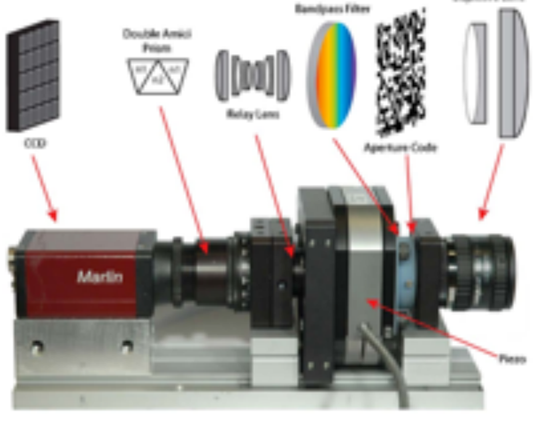
Recover $\boldsymbol{x}$ (signal) and $\boldsymbol{d}$ (gains) in

$$\boldsymbol{y}_l = \mathrm{diag}(\boldsymbol{d})\, \boldsymbol{A}_l \boldsymbol{x} + \boldsymbol{\eta},\ 1 \leqslant l \leqslant p$$

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

$$y_i \simeq \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle = \boldsymbol{a}_i^T \boldsymbol{x}$$

↘ additive noise ✓

↘ what if unknown gains?

Blind Calibration Problem: **our approach**

Recover $\boldsymbol{x}$ (signal) and $\boldsymbol{d}$ (gains) in

$$\boldsymbol{y}_l = \mathrm{diag}(\boldsymbol{d})\, \boldsymbol{A}_l \boldsymbol{x} + \boldsymbol{\eta},\ 1 \leqslant l \leqslant p$$

Central questions: (for sub-Gaussian $\boldsymbol{A}_l$)

- Efficient algorithm?
- Minimal sample complexity: mp ?
- Minimal snapshot number: p ?
- Robustness vs $\boldsymbol{\eta}$?

Intrinsic ambiguity (in noiseless case)

- Let $\mathcal{S} := \{(\boldsymbol{x}', \boldsymbol{d}') : \mathrm{diag}(\boldsymbol{d}')\boldsymbol{A}_l\boldsymbol{x}' = \mathrm{diag}(\boldsymbol{d})\boldsymbol{A}_l\boldsymbol{x} = \boldsymbol{y}_l, 1 \leqslant l \leqslant p\}$
- Scaling ambiguity:

$$(\boldsymbol{x}^*, \boldsymbol{d}^*) \in \mathcal{S} \quad \leftrightarrow \quad \forall \alpha \neq 0,\ (\tfrac{1}{\alpha}\,\boldsymbol{x}^*, \alpha\boldsymbol{d}^*) \in \mathcal{S}\,!$$

Intrinsic ambiguity (in noiseless case)

- Let $\mathcal{S} := \{(\boldsymbol{x}', \boldsymbol{d}') : \mathrm{diag}(\boldsymbol{d}')\boldsymbol{A}_l \boldsymbol{x}' = \mathrm{diag}(\boldsymbol{d})\boldsymbol{A}_l \boldsymbol{x} = \boldsymbol{y}_l, 1 \leqslant l \leqslant p\}$
- Scaling ambiguity:

$$(\boldsymbol{x}^*, \boldsymbol{d}^*) \in \mathcal{S} \quad \leftrightarrow \quad \forall \alpha \neq 0,\ (\tfrac{1}{\alpha}\boldsymbol{x}^*, \alpha \boldsymbol{d}^*) \in \mathcal{S}\,!$$

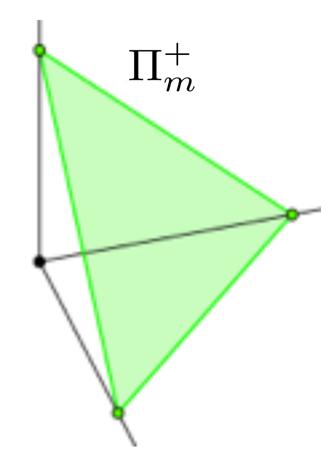
Our context:

- Gain calibration: $0 \leqslant d_i \approx 1,\ 1 \leqslant i \leqslant m$
- Let's assume (wlog):

$\sum_i d_i = m,$

or $\boldsymbol{d} \in \Pi_m^+ = \{\boldsymbol{w} \in \mathbb{R}_+^m : \mathbf{1}_m^\top \boldsymbol{w} = \sum_i w_i = m\}$

(Scaled) probability simplex



Intrinsic ambiguity (in noiseless case)

- Let $\mathcal{S} := \{(\boldsymbol{x}', \boldsymbol{d}') : \mathrm{diag}(\boldsymbol{d}')\boldsymbol{A}_l\boldsymbol{x}' = \mathrm{diag}(\boldsymbol{d})\boldsymbol{A}_l\boldsymbol{x} = \boldsymbol{y}_l, 1 \leqslant l \leqslant p\}$
- Scaling ambiguity:

$$(\boldsymbol{x}^*, \boldsymbol{d}^*) \in \mathcal{S} \quad \leftrightarrow \quad \forall \alpha \neq 0,\ (\tfrac{1}{\alpha}\boldsymbol{x}^*, \alpha\boldsymbol{d}^*) \in \mathcal{S}\,!$$

Our context:

- Gain calibration: $0 \leqslant d_i \approx 1,\ 1 \leqslant i \leqslant m$
- Let's assume (wlog):

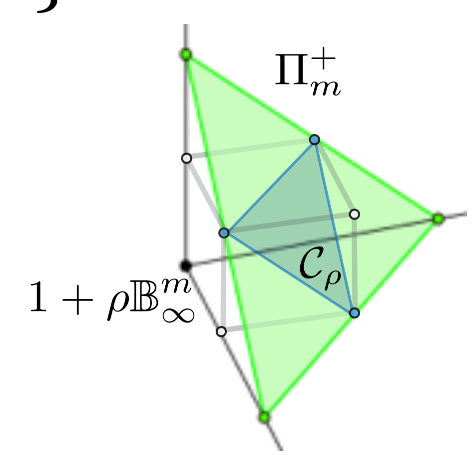
$\sum_i d_i = m,$

or $\boldsymbol{d} \in \Pi_m^+ = \{\boldsymbol{w} \in \mathbb{R}_+^m : \mathbf{1}_m^\top \boldsymbol{w} = \sum_i w_i = m\}$

+ perturbation analysis: $|d_i - 1| \leqslant \rho < 1$

(for some $0 \leqslant \rho < 1$)

$\Rightarrow \boldsymbol{d} \in \mathbf{1} + \rho\mathbb{B}_\infty^m$



$\Rightarrow$ We define $\mathcal{C}_\rho := \Pi_m^+ \cap (\mathbf{1} + \rho\mathbb{B}_\infty^m)$

our optimization space!

A Non-Convex Optimisation Problem

- Blind Calibration Problem:

$$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{d}}) = \underset{\boldsymbol{\xi} \in \mathbb{R}^n, \boldsymbol{\gamma} \in \mathcal{C}_\rho}{\operatorname{argmin}} \ \tfrac{1}{2mp} \textstyle\sum_{l=1}^{p} \| \underbrace{\operatorname{diag}(\boldsymbol{d}) \boldsymbol{A}_l \boldsymbol{x}}_{\boldsymbol{y}_l} - \operatorname{diag}(\boldsymbol{\gamma}) \boldsymbol{A}_l \boldsymbol{\xi} \|_2^2$$

- 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])

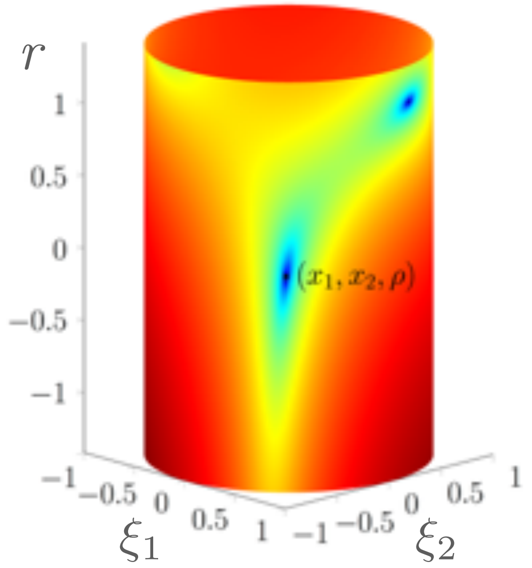
Geometric Analysis

- Low-dimensional intuitive example:

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

We study the variations of $f(\boldsymbol{\xi}, \boldsymbol{\gamma}) := \frac{1}{2mp} \sum_{l=1}^{p} \|\boldsymbol{y}_l - \operatorname{diag}(\boldsymbol{\gamma}) \boldsymbol{A}_l \boldsymbol{\xi}\|^2$
around $(x_1, x_2, \rho) = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0.08)$

$\log f(\boldsymbol{\xi}, \boldsymbol{\gamma}), \ p = 1$

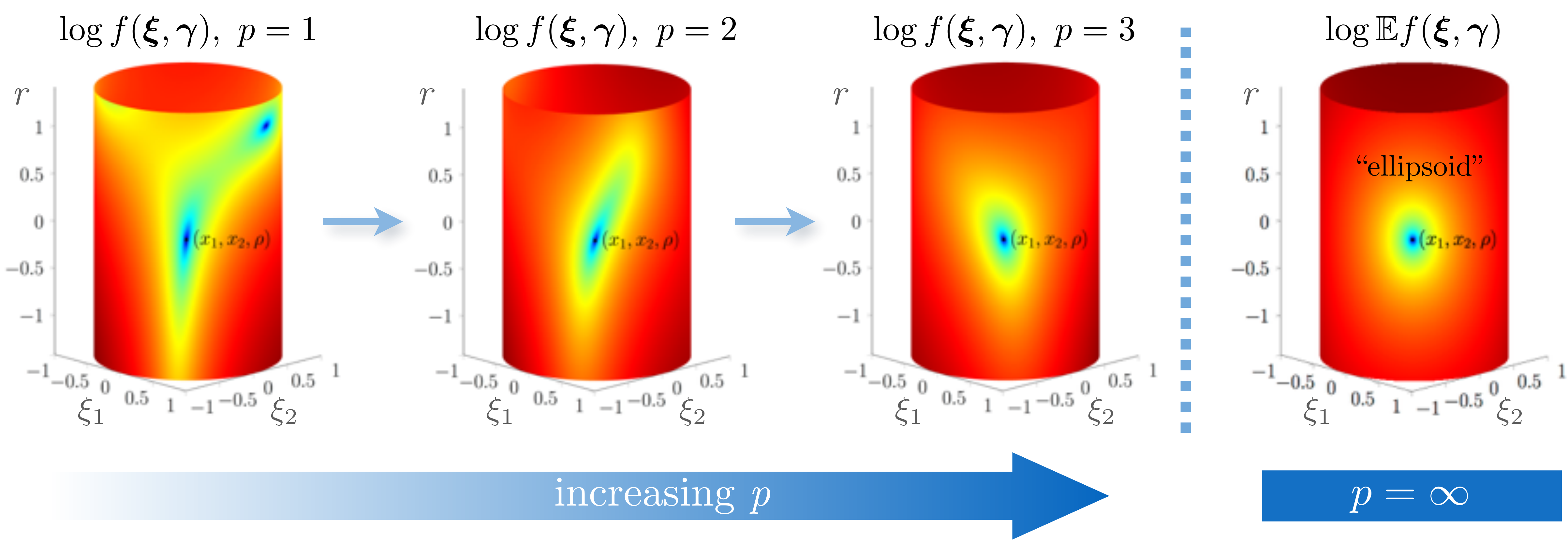


Geometric Analysis

- Low-dimensional intuitive example:

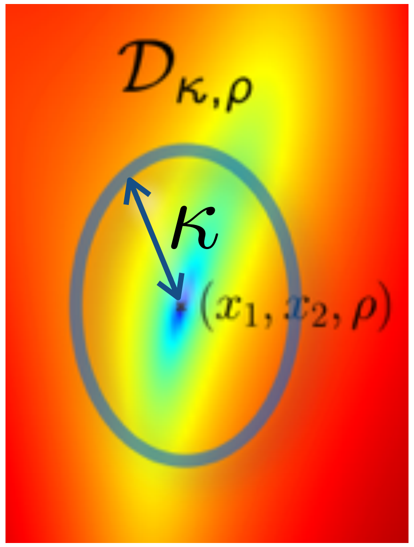
$$\boldsymbol{\gamma}, \boldsymbol{\xi} \in \mathbb{R}^2, \ i.e., \ n = m = 2,$$
$$\|\boldsymbol{\xi}\| = 1, \ \boldsymbol{\gamma} = (1+r, 1-r) \in \Pi_2^+, \ r \in \mathbb{R}$$
$\rightarrow$ Optimization space: (ξ_1, ξ_2, r) on a cylinder.

We study the variations of $f(\boldsymbol{\xi}, \boldsymbol{\gamma}) := \frac{1}{2mp} \sum_{l=1}^{p} \|\boldsymbol{y}_l - \mathrm{diag}(\boldsymbol{\gamma}) \boldsymbol{A}_l \boldsymbol{\xi}\|^2$ around $(x_1, x_2, \rho) = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0.08)$



Geometric Analysis

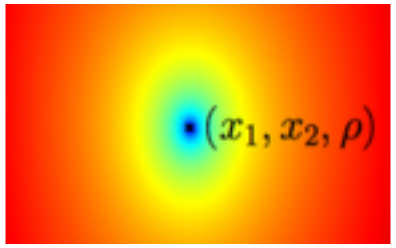
- Conclusion:
  - Hope for *local convexity*
    in a neighborhood (an "ellipsoid" of radius κ)



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

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

$\approx_\rho 2\,\mathbb{E}f(\boldsymbol{\xi},\boldsymbol{\gamma})$ if $\boldsymbol{\gamma} \in \mathcal{C}_\rho$.

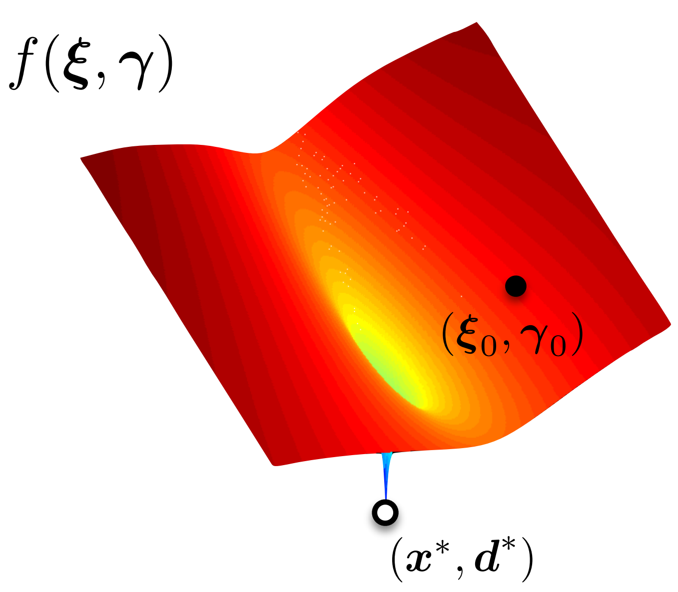


Solution by Projected Gradient Descent

$f(\boldsymbol{\xi}, \boldsymbol{\gamma}) := \frac{1}{2mp} \sum_{l=1}^{p} \|\boldsymbol{y}_l - \mathrm{diag}(\boldsymbol{\gamma}) \boldsymbol{A}_l \boldsymbol{\xi}\|^2$

- Algorithm:

1: Initialize $\boldsymbol{\xi}_0 := \frac{1}{mp} \sum_{l=1}^{p} (\boldsymbol{A}_l)^\top \boldsymbol{y}_l$, $\boldsymbol{\gamma}_0 := \mathbf{1}_m$, $k := 0$. (almost dumb ...



Solution by Projected Gradient Descent

- Algorithm:

$f(\boldsymbol{\xi}, \boldsymbol{\gamma}) := \frac{1}{2mp} \sum_{l=1}^{p} \|\boldsymbol{y}_l - \mathrm{diag}(\boldsymbol{\gamma}) \boldsymbol{A}_l \boldsymbol{\xi}\|^2$

1: Initialize $\boldsymbol{\xi}_0 := \frac{1}{mp} \sum_{l=1}^{p} (\boldsymbol{A}_l)^\top \boldsymbol{y}_l$, $\boldsymbol{\gamma}_0 := \mathbf{1}_m$, $k := 0$. (almost dumb ...

... but not so bad initialization!)

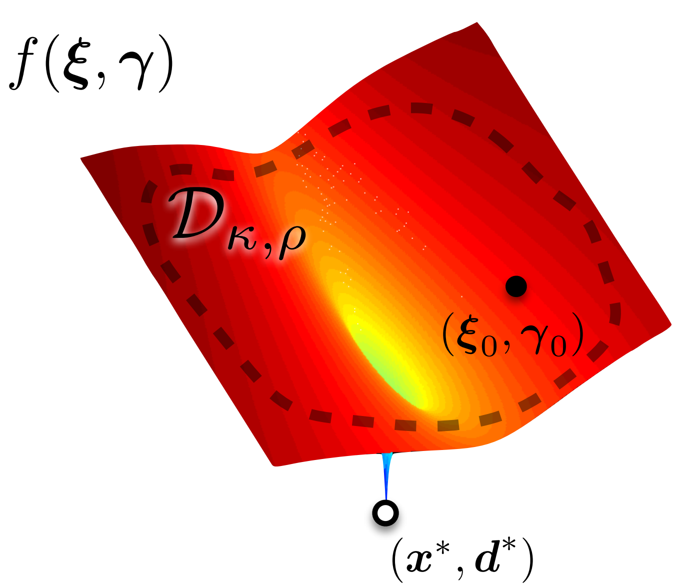
Prop. Let $0 < \delta < 1$, $t > 0$, and define $\kappa^2 := \delta^2 + \rho^2$. If

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

then

$$(\boldsymbol{\xi}_0, \boldsymbol{\gamma}_0) \in \mathcal{D}_{\kappa,\rho},$$

with prob. failure $\lesssim e^{-c\delta^2 mp} + (mp)^{-t}$ $(c > 0)$.



$$\Rightarrow \|\boldsymbol{\xi}_0 - \boldsymbol{x}^\star\|^2 + \frac{\|\boldsymbol{x}^\star\|^2}{m}\|\boldsymbol{\gamma}_0 - \boldsymbol{d}^\star\|^2 \leqslant \kappa^2 \|\boldsymbol{x}^*\|^2$$

Solution by Projected Gradient Descent

$f(\boldsymbol{\xi},\boldsymbol{\gamma}) := \frac{1}{2mp}\sum_{l=1}^{p}\|\boldsymbol{y}_l - \mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{A}_l\boldsymbol{\xi}\|^2$

- Algorithm:

1: Initialize $\boldsymbol{\xi}_0 := \frac{1}{mp}\sum_{l=1}^{p}(\boldsymbol{A}_l)^\top \boldsymbol{y}_l$, $\boldsymbol{\gamma}_0 := \mathbf{1}_m$, $k := 0$.

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

2: **while** stop criteria not met **do**

4: $\boldsymbol{\xi}_{k+1} := \boldsymbol{\xi}_k - \mu_{\boldsymbol{\xi}}\nabla_{\boldsymbol{\xi}} f(\boldsymbol{\xi}_k,\boldsymbol{\gamma}_k)$ {Signal Update}

5: $\underline{\boldsymbol{\gamma}}_{k+1} := \boldsymbol{\gamma}_k - \mu_{\boldsymbol{\gamma}}\nabla^{\perp}_{\boldsymbol{\gamma}} f(\boldsymbol{\xi}_k,\boldsymbol{\gamma}_k)$ {Gain Update}

6: $\boldsymbol{\gamma}_{k+1} := P_{\mathcal{C}_\rho}\underline{\boldsymbol{\gamma}}_{k+1}$ {Projection on C_ρ}

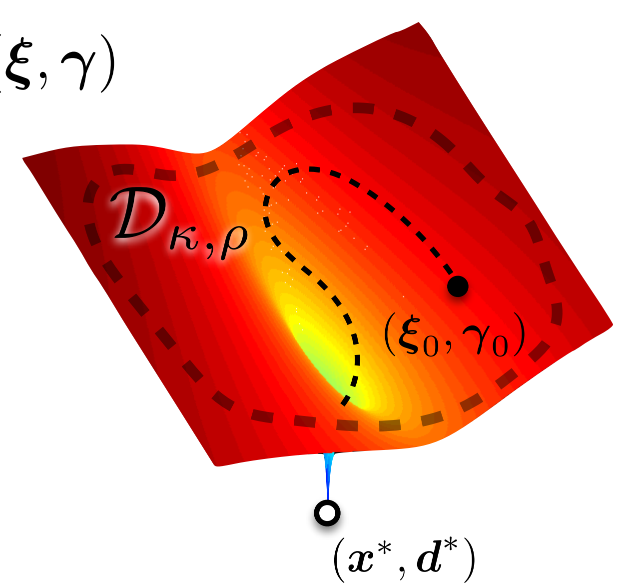
7: $k := k+1$

8: **end while**

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

technical requirement for proofs
(not required in experiments)

$f(\boldsymbol{\xi},\boldsymbol{\gamma})$



Solution by Projected Gradient Descent

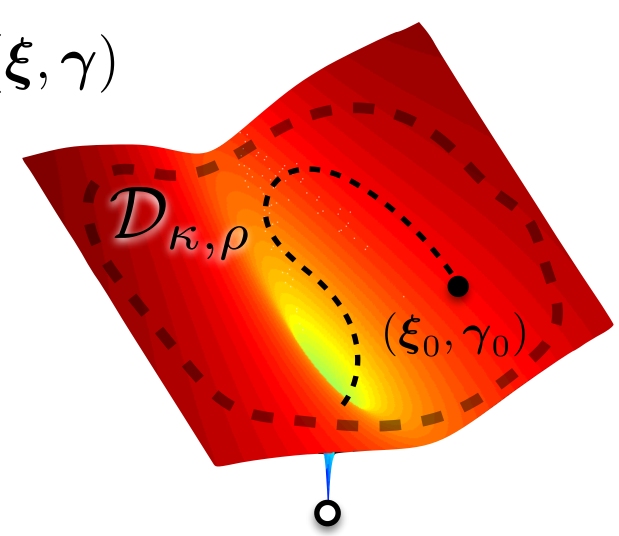
$f(\boldsymbol{\xi}, \boldsymbol{\gamma}) := \frac{1}{2mp} \sum_{l=1}^{p} \|\boldsymbol{y}_l - \mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{A}_l\boldsymbol{\xi}\|^2$

- <u>Algorithm</u>:

Initialize $\boldsymbol{\xi}_0 := \frac{1}{mp} \sum_{l=1}^{p} (\boldsymbol{A}_l)^\top \boldsymbol{y}_l$, $\boldsymbol{\gamma}_0 := \mathbf{1}_m$, $k := 0$.
while stop criteria not met **do**
$$\begin{cases} \mu_{\boldsymbol{\xi}} := \mathrm{argmin}_{\upsilon \in \mathbb{R}} f(\boldsymbol{\xi}_k - \upsilon \nabla_{\boldsymbol{\xi}} f(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k), \boldsymbol{\gamma}_k) \\ \mu_{\boldsymbol{\gamma}} := \mathrm{argmin}_{\upsilon \in \mathbb{R}} f(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k - \upsilon \nabla^{\perp}_{\boldsymbol{\gamma}} f(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k)) \end{cases}$$ {Line search in $\boldsymbol{\xi}, \boldsymbol{\gamma}$}
$\boldsymbol{\xi}_{k+1} := \boldsymbol{\xi}_k - \mu_{\boldsymbol{\xi}} \nabla_{\boldsymbol{\xi}} f(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k)$ {Signal Update}
$\underline{\boldsymbol{\gamma}}_{k+1} := \boldsymbol{\gamma}_k - \mu_{\boldsymbol{\gamma}} \nabla^{\perp}_{\boldsymbol{\gamma}} f(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k)$ {Gain Update}
$\boldsymbol{\gamma}_{k+1} := P_{\mathcal{C}_\rho} \underline{\boldsymbol{\gamma}}_{k+1}$ {Projection on C_ρ}
$k := k + 1$
end while

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

technical requirement for proofs
(not required in experiments)



Solution by Projected Gradient Descent

- Algorithm:

$f(\boldsymbol{\xi}, \boldsymbol{\gamma}) := \frac{1}{2mp} \sum_{l=1}^{p} \|\boldsymbol{y}_l - \mathrm{diag}(\boldsymbol{\gamma}) \boldsymbol{A}_l \boldsymbol{\xi}\|^2$

1: Initialize $\boldsymbol{\xi}_0 := \frac{1}{mp} \sum_{l=1}^{p} (\boldsymbol{A}_l)^\top \boldsymbol{y}_l$, $\boldsymbol{\gamma}_0 := \boldsymbol{1}_m$, $k := 0$.

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

2: **while** stop criteria not met **do**

4: $\boldsymbol{\xi}_{k+1} := \boldsymbol{\xi}_k - \mu_{\boldsymbol{\xi}} \nabla_{\boldsymbol{\xi}} f(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k)$ {Signal Update}

5: $\underline{\boldsymbol{\gamma}}_{k+1} := \boldsymbol{\gamma}_k - \mu_{\boldsymbol{\gamma}} \nabla^{\perp}_{\boldsymbol{\gamma}} f(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k)$ {Gain Update}

6: $\boldsymbol{\gamma}_{k+1} := P_{\mathcal{C}_\rho} \underline{\boldsymbol{\gamma}}_{k+1}$ {Projection on $\mathcal{C}_\rho$}

7: $k := k + 1$

8: **end while**

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

Convergence to $(\boldsymbol{x}^*, \boldsymbol{d}^*)$?

$$\Delta(\boldsymbol{\xi}_{k+1}, \boldsymbol{\gamma}_{k+1}) \overset{\text{(from } P_{\mathcal{C}_\rho})}{\leq} \Delta(\boldsymbol{\xi}_{k+1}, \underline{\boldsymbol{\gamma}}_{k+1}) \underset{?}{<} \Delta(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k)? \quad \text{(at any } k \geq 0)$$

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

... we need to prove regularity on *angles* and *magnitudes*!

Solution by Projected Gradient Descent

- Algorithm:

$f(\boldsymbol{\xi},\boldsymbol{\gamma}) := \frac{1}{2mp}\sum_{l=1}^{p}\|\boldsymbol{y}_l - \mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{A}_l\boldsymbol{\xi}\|^2$

1: Initialize $\boldsymbol{\xi}_0 := \frac{1}{mp}\sum_{l=1}^{p}(\boldsymbol{A}_l)^\top \boldsymbol{y}_l$, $\boldsymbol{\gamma}_0 := \mathbf{1}_m$, $k := 0$.

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

2: **while** stop criteria not met **do**

4: $\boldsymbol{\xi}_{k+1} := \boldsymbol{\xi}_k - \mu_{\boldsymbol{\xi}}\nabla_{\boldsymbol{\xi}} f(\boldsymbol{\xi}_k,\boldsymbol{\gamma}_k)$ {Signal Update}

5: $\underline{\boldsymbol{\gamma}}_{k+1} := \boldsymbol{\gamma}_k - \mu_{\boldsymbol{\gamma}}\nabla^{\perp}_{\boldsymbol{\gamma}} f(\boldsymbol{\xi}_k,\boldsymbol{\gamma}_k)$ {Gain Update}

6: $\boldsymbol{\gamma}_{k+1} := P_{\mathcal{C}_\rho}\underline{\boldsymbol{\gamma}}_{k+1}$ {Projection on C_ρ}

7: $k := k+1$

8: **end while**

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

Convergence to $(\boldsymbol{x}^*, \boldsymbol{d}^*)$?

$$\Delta(\boldsymbol{\xi}_{k+1},\underline{\gamma}_{k+1}) = \Delta(\boldsymbol{\xi}_k,\gamma_k) - 2\overbrace{\Big(\mu_{\boldsymbol{\xi}}\langle\nabla_{\boldsymbol{\xi}} f(\boldsymbol{\xi}_k,\gamma_k),\boldsymbol{\xi}_k - \boldsymbol{x}^\star\rangle + \mu_{\gamma}\tfrac{\|\boldsymbol{x}^\star\|_2^2}{m}\langle\nabla^{\perp}_{\gamma} f(\boldsymbol{\xi}_k,\gamma_k),\gamma_k - \boldsymbol{g}^\star\rangle\Big)}^{\text{(1) Gradient Angle Part (must be > 0)}}$$

$$+ \underbrace{\mu_{\boldsymbol{\xi}}^2\|\nabla_{\boldsymbol{\xi}} f(\boldsymbol{\xi}_k,\gamma_k)\|_2^2 + \mu_{\gamma}^2\tfrac{\|\boldsymbol{x}^\star\|_2^2}{m}\|\nabla^{\perp}_{\gamma} f(\boldsymbol{\xi}_k,\gamma_k)\|_2^2}_{\text{(2) Gradient Magnitude Part (must bounded)}} \overset{?}{<} \Delta(\boldsymbol{\xi}_k,\gamma_k)$$

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

... we need to prove regularity on *angles* and *magnitudes*!

Convergence to $(\boldsymbol{x}^*, \boldsymbol{d}^*)$?

- Regularity condition in $\mathcal{D}_{\kappa,\rho}$

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

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

then, $\exists\, \eta, L > 0$ (only depending on δ and ρ) such that, $\forall\, (\boldsymbol{\xi}, \boldsymbol{\gamma}) \in \mathcal{D}_{\kappa,\rho}$,

$$① \quad \left\langle \boldsymbol{\nabla}^{\perp} f(\boldsymbol{\xi}, \boldsymbol{\gamma}), \begin{bmatrix} \boldsymbol{\xi}-\boldsymbol{x}^* \\ \boldsymbol{\gamma}-\boldsymbol{d}^* \end{bmatrix} \right\rangle \;\geq\; \tfrac{1}{2}\eta\, \Delta(\boldsymbol{\xi}, \boldsymbol{\gamma}) \qquad \text{(Bounded angle)}$$

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

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 $\boldsymbol{x}^*$ and $\boldsymbol{d}^*$.

Convergence to $(\boldsymbol{x}^*, \boldsymbol{d}^*)$?

- Regularity condition in $\mathcal{D}_{\kappa,\rho}$

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

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

then, $\exists\, \eta, L > 0$ (only depending on δ and ρ) such that, $\forall\, (\boldsymbol{\xi}, \boldsymbol{\gamma}) \in \mathcal{D}_{\kappa,\rho}$,

$$\left\langle \boldsymbol{\nabla}^\perp f(\boldsymbol{\xi},\boldsymbol{\gamma}), \begin{bmatrix} \boldsymbol{\xi}-\boldsymbol{x}^* \\ \boldsymbol{\gamma}-\boldsymbol{d}^* \end{bmatrix} \right\rangle \geq \tfrac{1}{2}\eta\, \Delta(\boldsymbol{\xi},\boldsymbol{\gamma}) \qquad \text{(Bounded angle)}$$

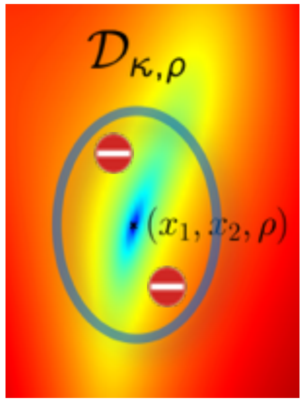
$$\|\boldsymbol{\nabla}^\perp f(\boldsymbol{\xi},\boldsymbol{\gamma})\|^2 \leq L^2\, \Delta(\boldsymbol{\xi},\boldsymbol{\gamma}) \qquad \text{(Lipschitz gradient)}$$

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

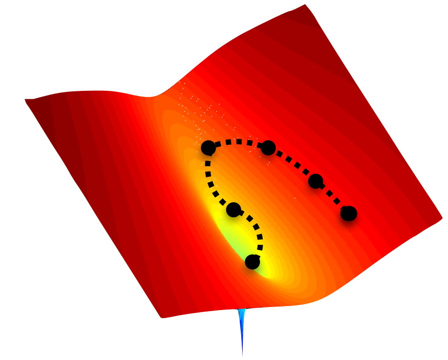
$\Rightarrow \|\boldsymbol{\nabla}^\perp f(\boldsymbol{\xi},\boldsymbol{\gamma})\| \neq 0$ but on the solution in $\mathcal{D}_{\kappa,\rho}$!

$\Rightarrow$ allows convergence for $\mu_{\boldsymbol{\gamma}} = \mu_{\boldsymbol{\xi}} \frac{m}{\|\boldsymbol{x}^*\|}$.

(no spurious minima)



Convergence to $(\boldsymbol{x}^*, \boldsymbol{d}^*)$?



Combining previous propositions gives then ...

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

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

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

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

for some $C, c > 0$, our descent algorithm initialized on $(\boldsymbol{\xi}_0, \boldsymbol{\gamma}_0)$ with $\mu_{\boldsymbol{\xi}} = \mu$ and $\mu_{\boldsymbol{\gamma}} = \mu \frac{m}{\|\boldsymbol{x}^\star\|_2^2}$ gives jointly, at each iteration k,

$$(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k) \in \mathcal{D}_{\kappa,\rho} \quad \text{and} \quad \Delta(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k) \leq \big(\overbrace{1 - \eta\mu + \tfrac{L^2}{\tau}\mu^2}^{<1}\big)^k \kappa^2 \|\boldsymbol{x}^\star\|_2^2,$$

provided $\mu \in \big(0, \frac{\eta\|\boldsymbol{x}^\star\|^2}{mL^2 + \|\boldsymbol{x}^\star\|^2 L^2}\big)$. Hence, $\Delta(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k) \underset{k\to\infty}{\longrightarrow} 0$.

Roughly speaking, for ρ small enough,
we need n and $p > 1$ large enough,
and $mp \gtrsim (m+n)\log(n/\delta)$ observations.

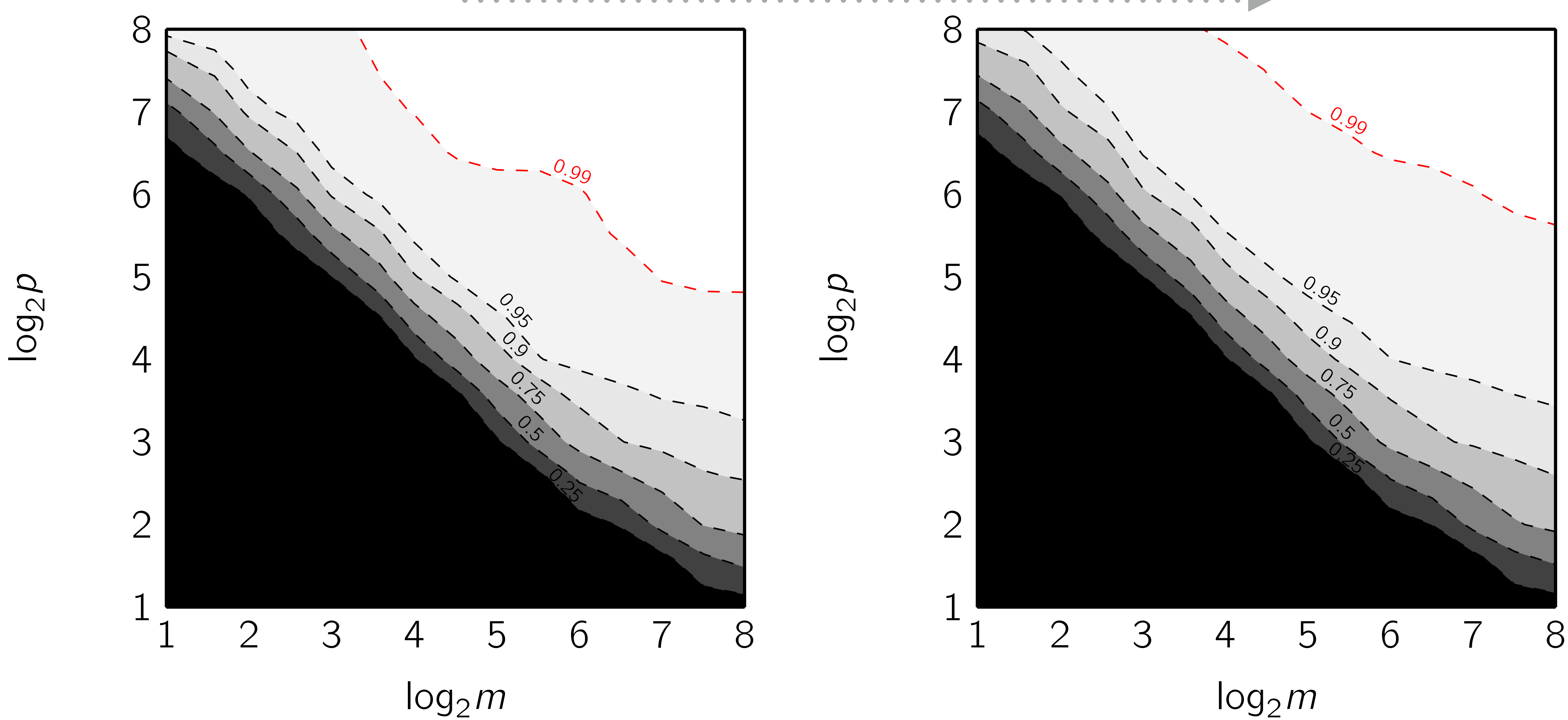
Empirical Phase Transition

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

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

for 256 randomly generated problem instances (per point).

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

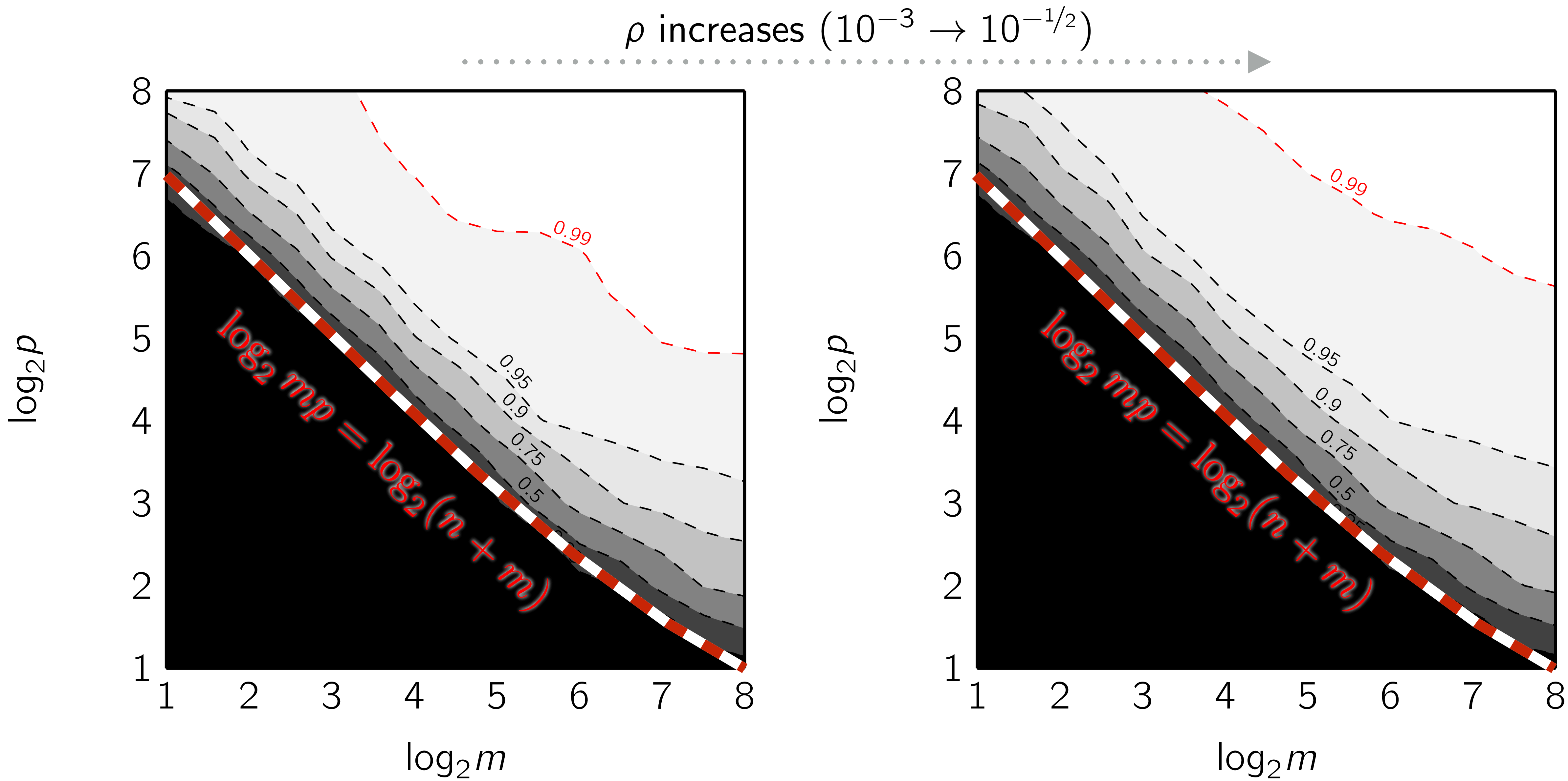


Empirical Phase Transition

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

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

for 256 randomly generated problem instances (per point).



(Randomized) Computational Imaging

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

Fixed signal $\boldsymbol{x}$

e.g.,

Pluto

(NewHorizon 2015)

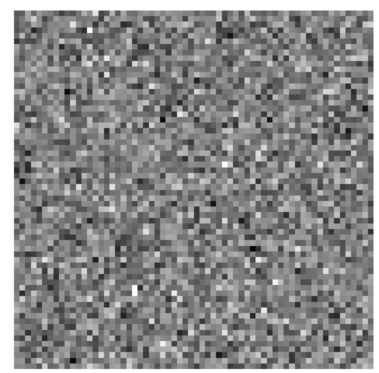
(Randomized) Computational Imaging

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

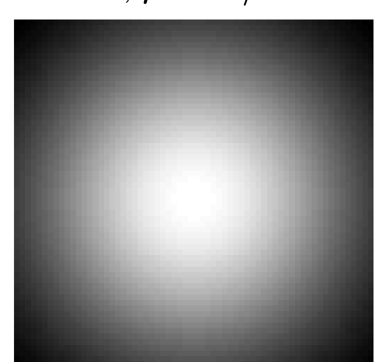
- LS SNR: 5.5 dB on signal

Fixed signal $\boldsymbol{x}$

Unstructured $\boldsymbol{d}, \rho = {}^1/_2$

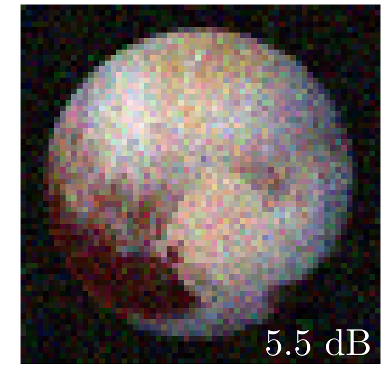


Structured $\boldsymbol{d}, \rho = {}^9/_{10}$



(about 2')

$\hat{\boldsymbol{x}}$ with LS, unstructured $\boldsymbol{d}$



$\hat{\boldsymbol{x}}$ with LS,structured $\boldsymbol{d}$

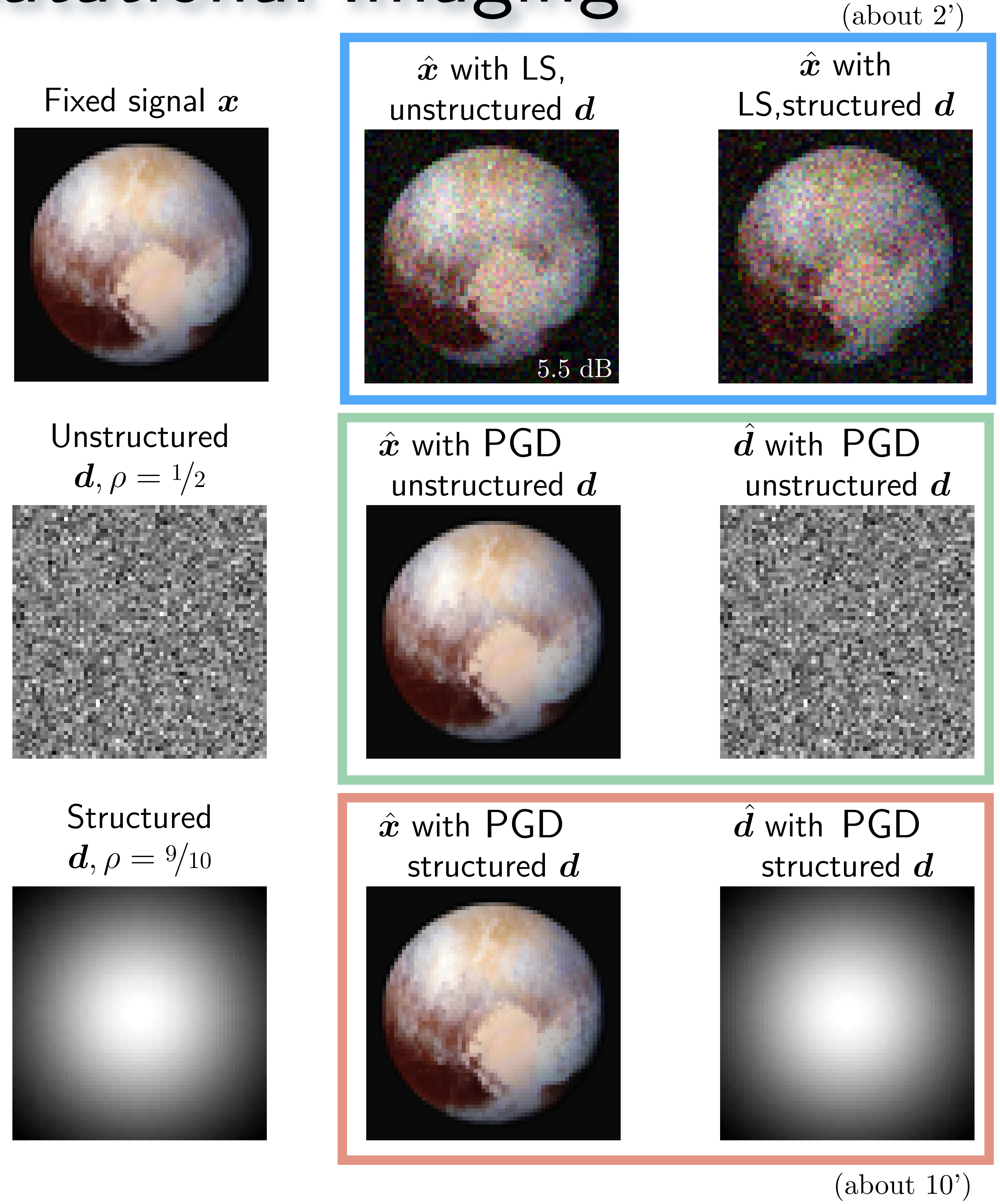
(Randomized) Computational Imaging

- 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 $\boldsymbol{A}_l$ (e.g., *subsampled random convolution, spread-spectrum)* (not covered by current theory).



Preliminary extension to sparse signals (ICASSP'17)

- Assumption: $\boldsymbol{x}$ is k-sparse in an ONB $\boldsymbol{\Psi}$

  $(i.e., |\operatorname{supp} \boldsymbol{\Psi}^\top \boldsymbol{x}| =: \|\boldsymbol{\Psi}^\top \boldsymbol{x}\|_0 \leqslant k)$

- Hard thresholding "projection":

$$\mathcal{H}_k(\boldsymbol{u}) := \operatorname{argmin}_{\boldsymbol{v}} \|\boldsymbol{u} - \boldsymbol{v}\| \text{ s.t. } \|\boldsymbol{v}\|_0 = k$$

Preliminary extension to sparse signals (ICASSP'17)

- Assumption: $\boldsymbol{x}$ is k-sparse in an ONB $\boldsymbol{\Psi}$
  $$(i.e., |\operatorname{supp} \boldsymbol{\Psi}^\top \boldsymbol{x}| =: \|\boldsymbol{\Psi}^\top \boldsymbol{x}\|_0 \leqslant k)$$
- Hard thresholding "projection":
  $$\mathcal{H}_k(\boldsymbol{u}) := \operatorname{argmin}_{\boldsymbol{v}} \|\boldsymbol{u} - \boldsymbol{v}\| \text{ s.t. } \|\boldsymbol{v}\|_0 = k$$
- Objective: (under similar assumptions: $\boldsymbol{d} \in \Pi_m^+$, $\|\boldsymbol{d} - \mathbf{1}\|_\infty \leqslant \rho$)
  $$\text{Recover } (\boldsymbol{x}, \boldsymbol{d}) \text{ from } \{\boldsymbol{y}_l = \operatorname{diag}(\boldsymbol{d}) \boldsymbol{A}_l \boldsymbol{x} : 1 \leqslant p \leqslant p\}$$
- Hope: sample complexity should be like
  $$mp \gtrsim m + k \quad \Rightarrow \quad m(p-1) \gtrsim k$$

Preliminary extension to sparse signals (ICASSP'17)

- Assumption: $\boldsymbol{x}$ is k-sparse in an ONB $\boldsymbol{\Psi}$

$$(i.e., \, |\operatorname{supp} \boldsymbol{\Psi}^\top \boldsymbol{x}| =: \|\boldsymbol{\Psi}^\top \boldsymbol{x}\|_0 \leqslant k)$$

- Hard thresholding "projection":

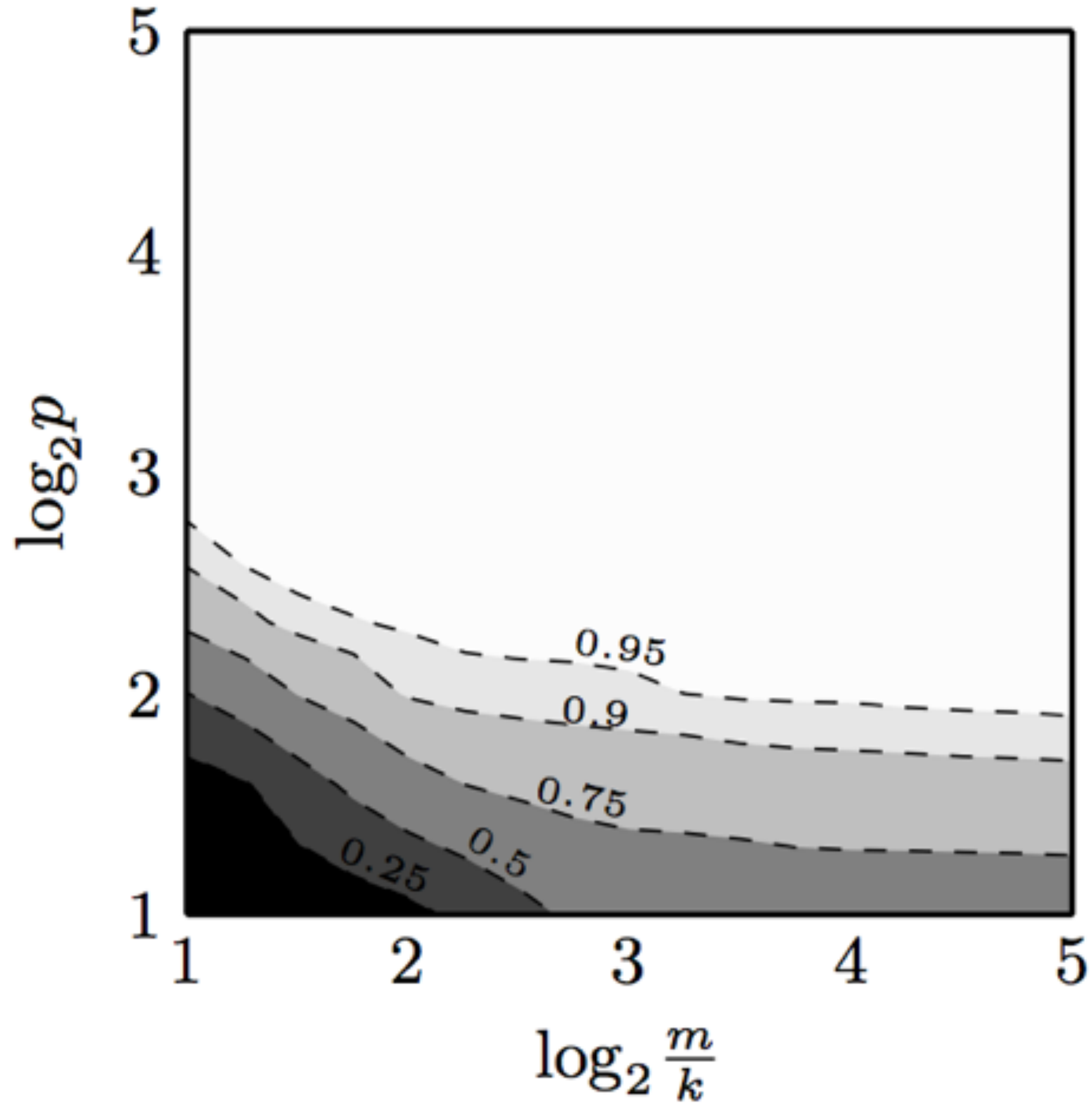
$$\mathcal{H}_k(\boldsymbol{u}) := \operatorname{argmin}_{\boldsymbol{v}} \|\boldsymbol{u} - \boldsymbol{v}\| \text{ s.t. } \|\boldsymbol{v}\|_0 = k$$

- <u>Blind Calibration with Iterative Hard Thresholding</u>

Initialize $\boldsymbol{\xi}_0 := \frac{1}{mp} \sum_{l=1}^{p} (\boldsymbol{A}_l)^\top \boldsymbol{y}_l$, $\boldsymbol{\gamma}_0 := \mathbf{1}_m$, $j := 0$.
while stop criteria not met **do**
  $\boldsymbol{\xi}_{j+1} := \boldsymbol{\Psi}\, \mathcal{H}_k[\boldsymbol{\Psi}^\top(\boldsymbol{\xi}_j - \mu_{\boldsymbol{\xi}} \boldsymbol{\nabla}_{\boldsymbol{\xi}} f(\boldsymbol{\xi}_j, \boldsymbol{\gamma}_j)]$ {Signal Update}
  $\underline{\boldsymbol{\gamma}}_{j+1} := \boldsymbol{\gamma}_j - \mu_{\gamma}\, \boldsymbol{\nabla}^{\perp}_{\boldsymbol{\gamma}} f(\boldsymbol{\xi}_j, \boldsymbol{\gamma}_j)$ {Gain Update}
  $\boldsymbol{\gamma}_{j+1} := P_{\mathcal{C}_\rho} \underline{\boldsymbol{\gamma}}_{j+1}$ {Projection on C_ρ}
  $j := j + 1$
end while

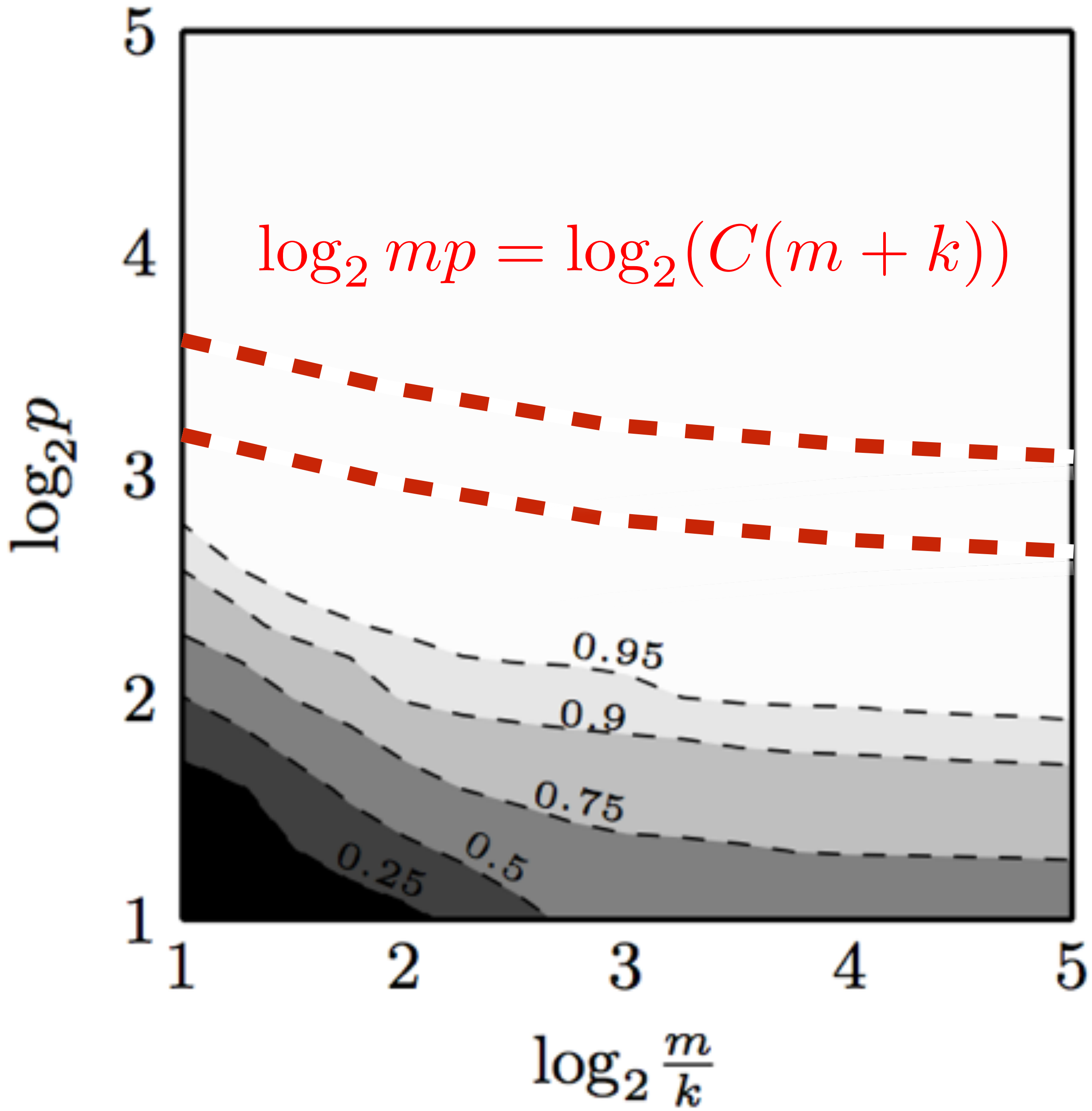
Empirical Phase Transition (bis)

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



Empirical Phase Transition (bis)

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

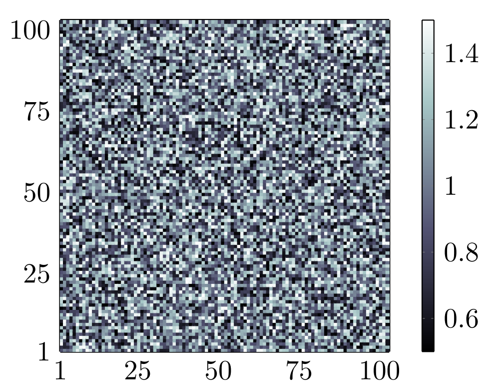


"Tous les jours", by René Magritte

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

(sparsified) true signal

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



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

"Tous les jours", René Magritte, 1966 (Charly Herscovici 2011, Wikiart.org)

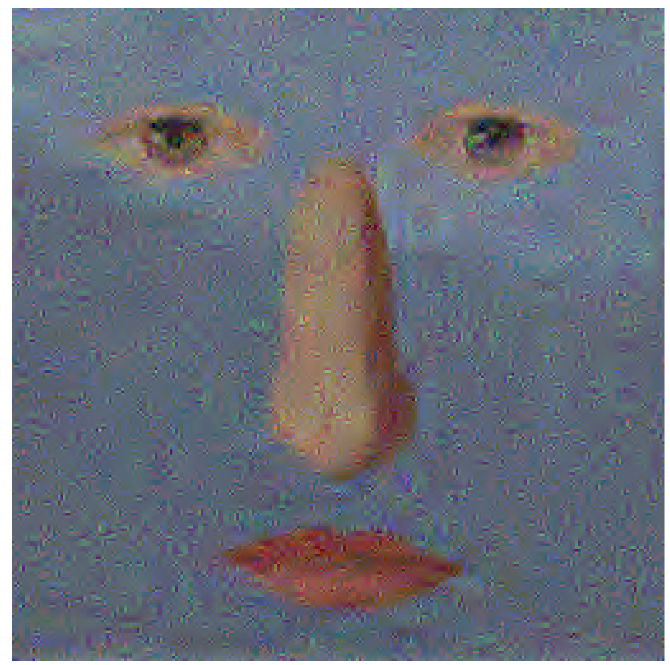
“Tous les jours”, by René Magritte

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

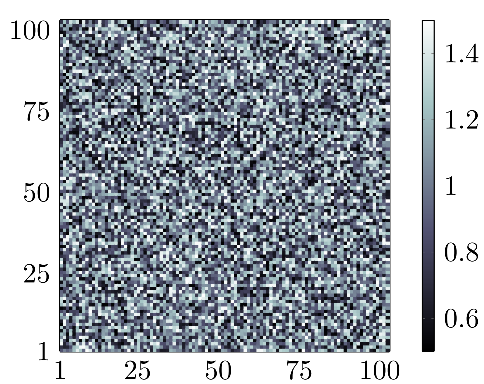
(sparsified) true signal

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

IHT



(b) Recovery $\hat{\boldsymbol{x}}$ provided by IHT, $\mathrm{RSNR}_{\boldsymbol{x},\hat{\boldsymbol{x}}} = 17.83\,\mathrm{dB}$



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

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

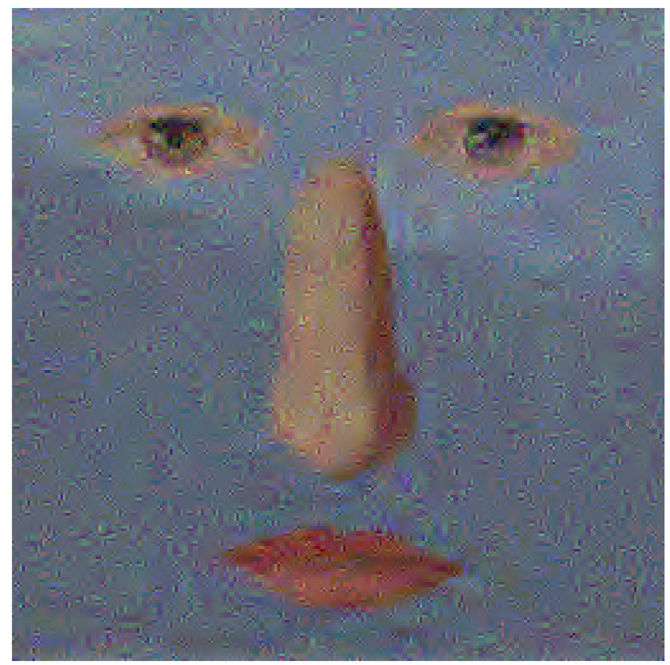
"Tous les jours", by René Magritte

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

(sparsified) true signal

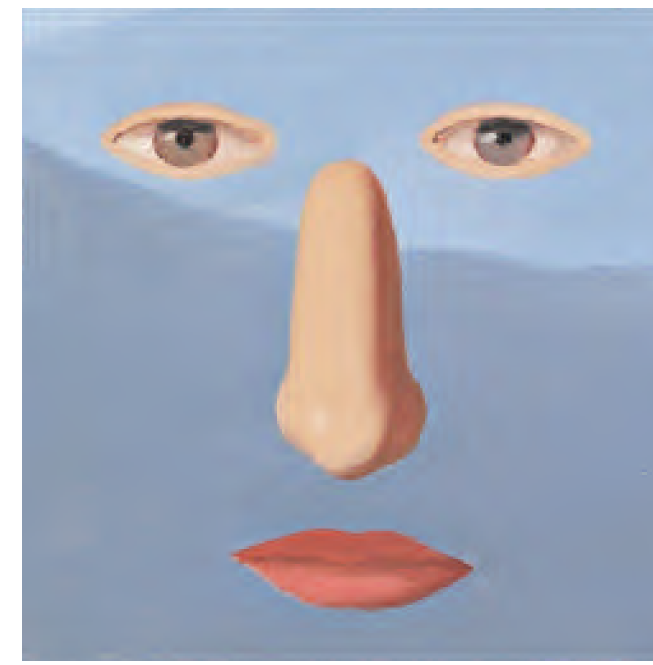
(a) True signal $\boldsymbol{x}$, $n = 256 \times 256$ px

IHT

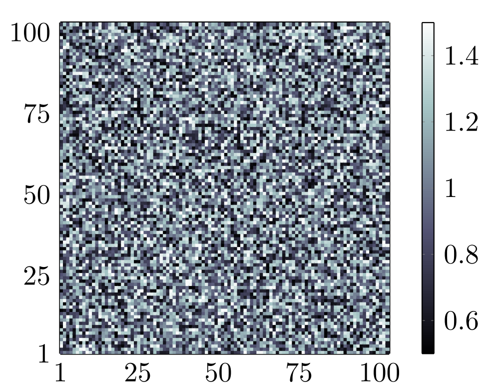


(b) Recovery $\hat{\boldsymbol{x}}$ provided by IHT, $\mathrm{RSNR}_{\boldsymbol{x},\hat{\boldsymbol{x}}} = 17.83\,\mathrm{dB}$

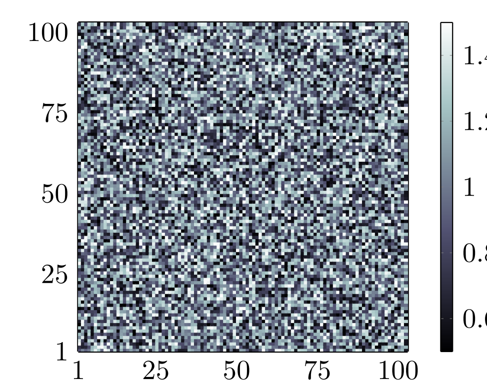
BC-IHT



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



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



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

"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:

$$mp \underset{*}{\gtrsim} (n+m)\log n,\ \ p \gtrsim \log m,\ \ n \gtrsim \log mp$$

(*: 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:

$$\Delta(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k) \underset{k\to\infty}{\longrightarrow} C\|\text{noise}\|^2$$

  - 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:

$$mp \underset{*}{\gtrsim} (n+m)\log n,\ \ p \gtrsim \log m,\ \ n \gtrsim \log mp$$

(*: 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:

$$\Delta(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k) \underset{k\to\infty}{\longrightarrow} C\|\text{noise}\|^2$$

  - 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

1. M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," IEEE Journal of Selected Topics in Signal Processing, vol. 4, no. 2, pp. 342–349, 2010.
2. R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," in Compressed Sensing: Theory and Applications. Cambridge University Press, 2012, pp. 210–268.
3. C. Bilen, G. Puy, R. Gribonval, and L. Daudet, "Convex Optimisation Approaches for Blind Sensor Calibration Using Sparsity," IEEE Transactions on Signal Processing, vol. 62, no. 18, pp. 4847–4856, Sep. 2014.
4. B. Friedlander and T. Strohmer, "Bilinear compressed sensing for array self-calibration," in 2014 48th Asilomar Conference on Signals, Systems and Computers, Nov. 2014, pp. 363–367.
5. S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," Inverse Problems, vol. 31, no. 11, p. 115002, 2015.
6. E. Candès, X. Li, and M. Soltanolkotabi, "Phase Retrieval via Wirtinger Flow: Theory and Algorithms," IEEE Transactions on Information Theory, vol. 61, no. 4, pp. 1985– 2007, Apr. 2015.
7. C. D. White, S. Sanghavi, and R. Ward, "The local convexity of solving systems of quadratic equations," arXiv:1506.07868 [math, stat], Jun. 2015, arXiv: 1506.07868.
8. V. Cambareri and L. Jacques, "A Non-Convex Blind Calibration Method for Randomised Sensing Strategies," in IEEE 2016 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa).
9. V. Cambareri and L. Jacques, "Through the Haze: A Non-Convex Approach to Blind Calibration for Linear Random Sensing Models," 2016, in preparation.
10. X.Li, S.Ling, T.Strohmer, K.Wei, "Rapid, Robust, and Reliable Blind Deconvolution via Nonconvex Optimization," ArXiv Preprint, 2016.