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

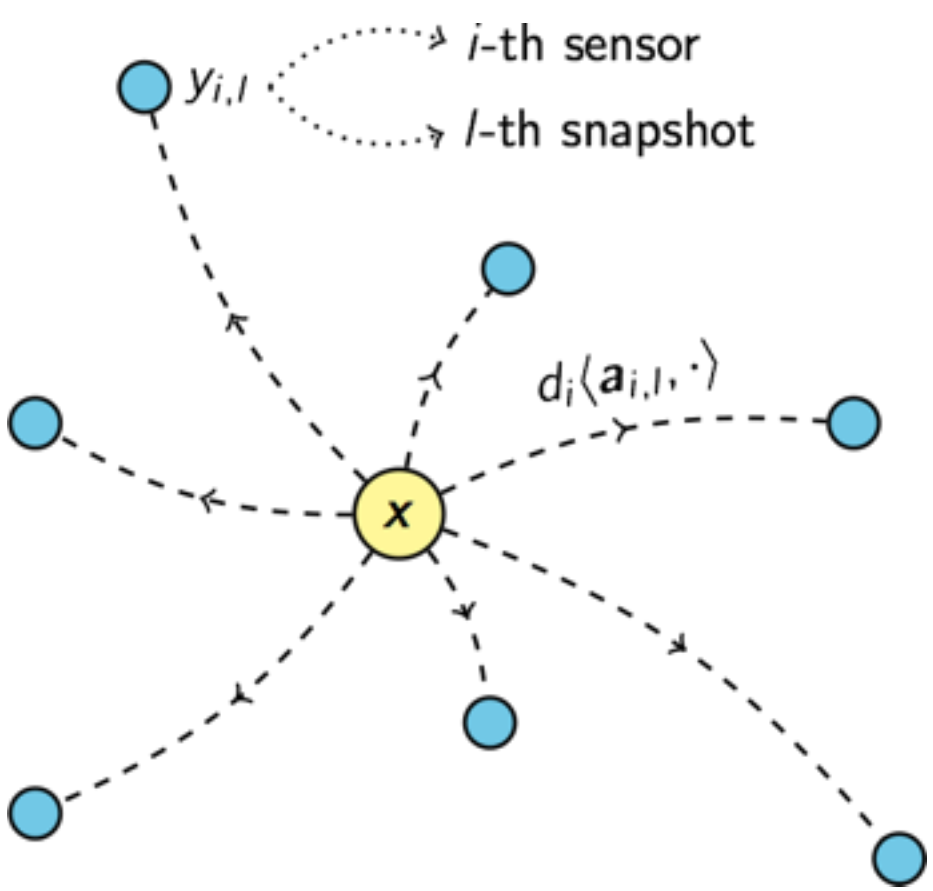
- **Introduction:** Blind Calibration and Random Linear Models
- **A Non-Convex Approach to Blind Calibration**
- **Solution by Projected Gradient Descent**
- **Global Convergence Guarantees** (even if non-convex!)
- **Experiments I:** Empirical Phase Transition
- **Experiments II:** Computational Imaging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

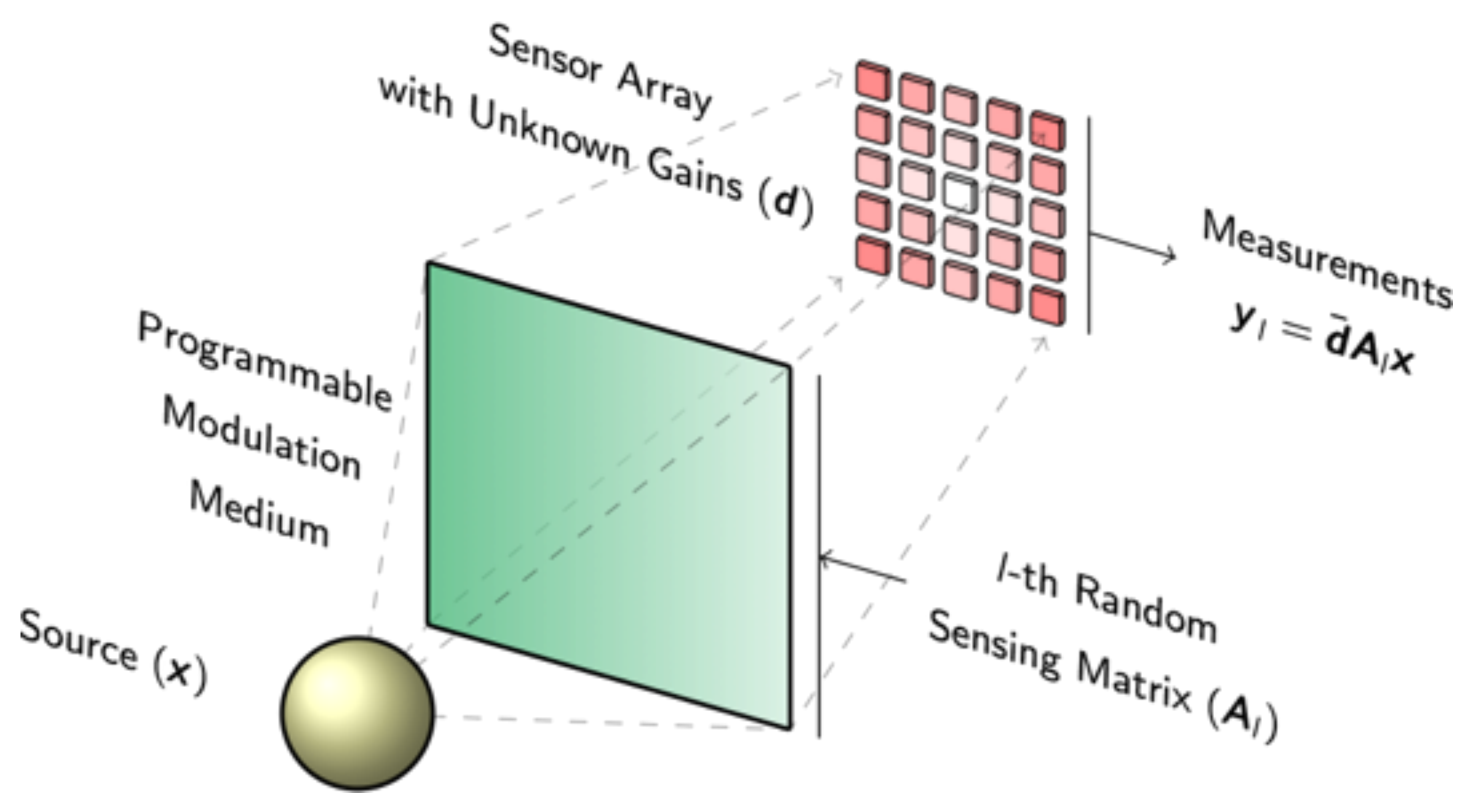
Random Linear Sensing Model

$y_{i,l} = d_i \langle \boldsymbol{a}_{i,l}, \boldsymbol{x} \rangle, d_i > 0, \boldsymbol{x} \in \mathbb{R}^n$
$i = 1, \ldots, m, l = 1, \ldots, p$

$\equiv$

$\boldsymbol{y}^{(l)} = \mathrm{diag}\,(\boldsymbol{d}) \boldsymbol{A}^{(l)} \boldsymbol{x}, \boldsymbol{d} \in \mathbb{R}^m_+, \boldsymbol{x} \in \mathbb{R}^n$
$l = 1, \ldots, p, mp \geq n + m$



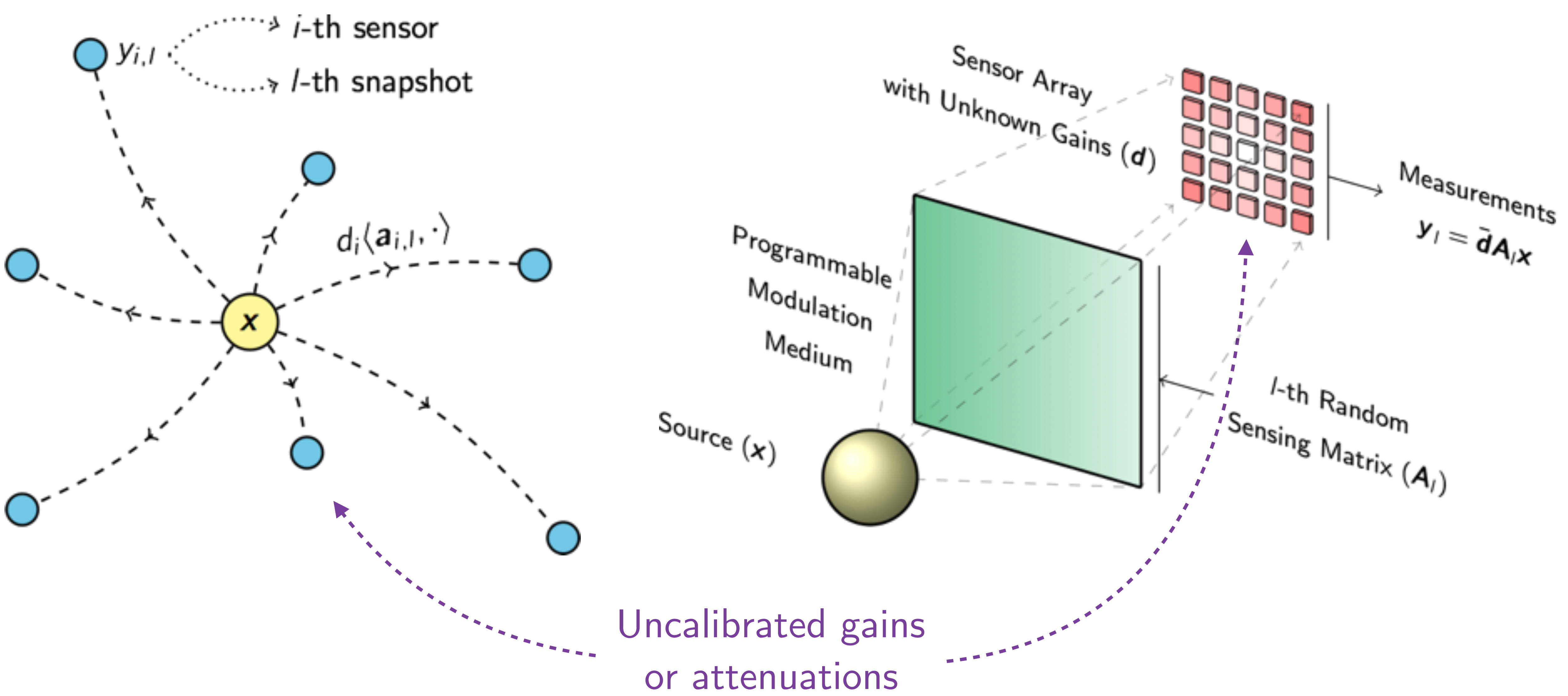


Random Linear Sensing Model

$y_{i,l} = d_i \langle \boldsymbol{a}_{i,l}, \boldsymbol{x} \rangle, d_i > 0, \boldsymbol{x} \in \mathbb{R}^n$
$i = 1, \ldots, m, l = 1, \ldots, p$

$\equiv$

$\boldsymbol{y}^{(l)} = \mathrm{diag}(\boldsymbol{d}) \boldsymbol{A}^{(l)} \boldsymbol{x}, \boldsymbol{d} \in \mathbb{R}^m_+, \boldsymbol{x} \in \mathbb{R}^n$
$l = 1, \ldots, p, mp \geq n + m$



Random Linear Sensing Model

$y_{i,l} = d_i \langle \boldsymbol{a}_{i,l}, \boldsymbol{x} \rangle, d_i > 0, \boldsymbol{x} \in \mathbb{R}^n$
$i = 1, \ldots, m, l = 1, \ldots, p$

$\equiv$

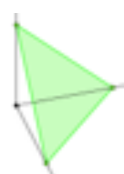
$\boldsymbol{y}^{(l)} = \operatorname{diag}(\boldsymbol{d}) \boldsymbol{A}^{(l)} \boldsymbol{x}, \boldsymbol{d} \in \mathbb{R}^m_+, \boldsymbol{x} \in \mathbb{R}^n$
$l = 1, \ldots, p, mp \geq n + m$

- Which *algorithm* to *jointly* recover $\boldsymbol{x}$ and $\boldsymbol{d}$? 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 $\boldsymbol{a}_{i,l}$?

A Non-Convex Optimisation Problem

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

$$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{d}}) = \underset{\boldsymbol{\xi}\in\mathbb{R}^n, \boldsymbol{\gamma}\in\Pi_+^m}{\operatorname{argmin}} \frac{1}{2mp} \sum_{l=1}^{p} \| \underbrace{\operatorname{diag}(\boldsymbol{d})\boldsymbol{A}^{(l)}\boldsymbol{x}}_{\boldsymbol{y}^{(l)}} - \operatorname{diag}(\boldsymbol{\gamma})\boldsymbol{A}^{(l)}\boldsymbol{\xi}\|_2^2$$



(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: $\left\{(\boldsymbol{\xi},\boldsymbol{\gamma}) \in \mathbb{R}^n \times \mathbb{R}^m : \boldsymbol{\xi} = \frac{1}{\alpha}\boldsymbol{x}, \boldsymbol{\gamma} = \alpha\boldsymbol{d}, \alpha \in \mathbb{R}\setminus\{0\}\right\}$
- The constraint fixes one global minimiser: $(\boldsymbol{x}^\star, \boldsymbol{d}^\star) := \left(\frac{\|\boldsymbol{d}\|_1}{m}\boldsymbol{x}, \frac{m}{\|\boldsymbol{d}\|_1}\boldsymbol{d}\right)$
- Since the gains are *positive* and *bounded*, we let:

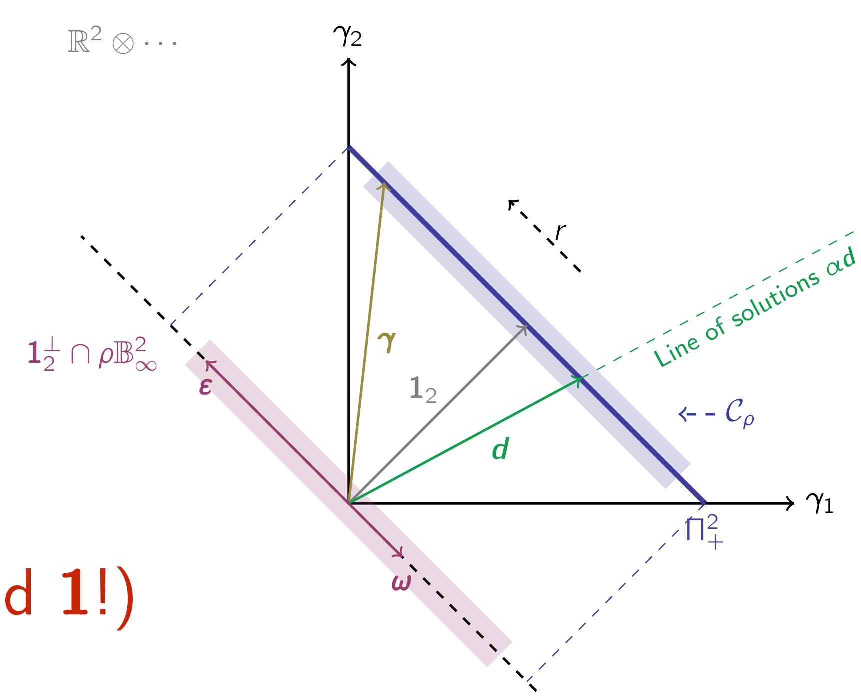
$$\boldsymbol{\gamma}, \boldsymbol{d}^\star \in \mathcal{C}_\rho \subset \Pi_+^m,\ \mathcal{C}_\rho := \mathbf{1}_m + \mathbf{1}_m^\perp \cap \rho\,\mathbb{B}_\infty^m$$

$$\boldsymbol{d}^\star = \mathbf{1}_m + \boldsymbol{\omega},\ \boldsymbol{\omega} \in \mathbf{1}_m^\perp \cap \rho\,\mathbb{B}_\infty^m$$

$$\boldsymbol{\gamma} = \mathbf{1}_m + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \in \mathbf{1}_m^\perp \cap \rho\,\mathbb{B}_\infty^m$$

where:

$\rho > \|\boldsymbol{d}^\star - \mathbf{1}_m\|_\infty, \rho < 1$ (Perturbation analysis around **1**!)



Solution by Projected Gradient Descent

- The solution is obtained by *projected gradient descent*:

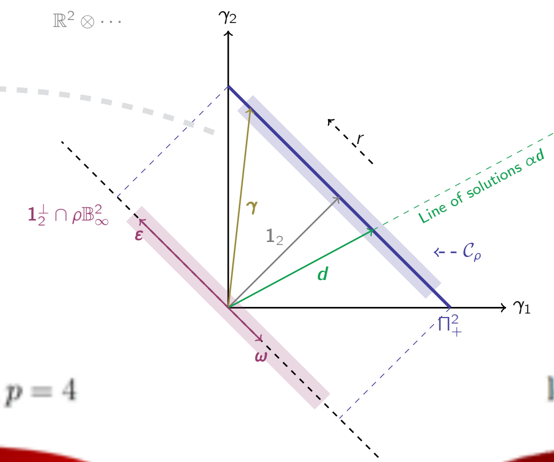
Initialise $\boldsymbol{\xi}_0 := \frac{1}{mp}\sum_{l=1}^{p} (\boldsymbol{A}_l)^\top \boldsymbol{y}_l$, $\boldsymbol{\gamma}_0 := \boldsymbol{1}_m$, $k := 0$.
while stop criteria not met **do**
    $\begin{cases} \mu_{\boldsymbol{\xi}} := \mathrm{argmin}_{\upsilon \in \mathbb{R}} f(\boldsymbol{\xi}_k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

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

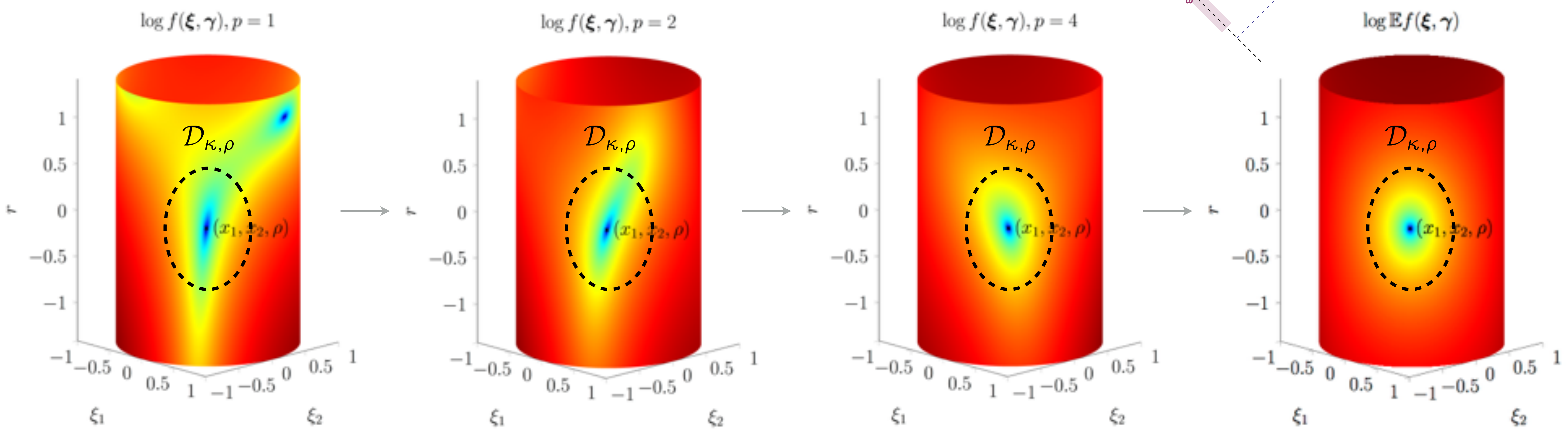
- The chosen *initialisation* is *crucial*: in expectation (asymptotic p) it yields $\boldsymbol{x}$ (unbiased estimator). For finite p it can be shown to lie *close* to the *global minimum*.

- Projection on $\mathcal{C}_\rho$ is only a technical requirement for proofs (not required in experiments).

Geometric Analysis



- Consider a low-dimensional intuitive example for a random instance of the problem, at $n = 2, m = 2, \|\boldsymbol{\xi}\|_2 = 1$, parametrising $\boldsymbol{\gamma}(r)$.



- To measure distances, we adopt the pre-metric:

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

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

$$\mathcal{D}_{\kappa,\rho} := \{(\boldsymbol{\xi}, \boldsymbol{\gamma}) \in \mathbb{R}^n \times \mathcal{C}_\rho : \Delta(\boldsymbol{\xi}, \boldsymbol{\gamma}) \leq \kappa^2 \|\boldsymbol{x}^\star\|_2^2\}, \ \rho \in [0, 1).$$

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, $(\boldsymbol{\xi}_0, \boldsymbol{\gamma}_0) \in \mathcal{D}_{\kappa,\rho}$, $\rho \in [0, 1)$
  2. **Regularity Condition**: developing the distance at iterate $k{+}1$,

$$\Delta(\boldsymbol{\xi}_{k+1}, \underline{\boldsymbol{\gamma}}_{k+1}) = \Delta(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k) - 2\overbrace{\left(\mu_{\boldsymbol{\xi}} \langle \nabla_{\boldsymbol{\xi}} f(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k), \boldsymbol{\xi}_k - \boldsymbol{x}^\star\rangle + \mu_{\boldsymbol{\gamma}} \tfrac{\|\boldsymbol{x}^\star\|_2^2}{m} \langle \nabla_{\boldsymbol{\gamma}}^{\perp} f(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k), \boldsymbol{\gamma}_k - \boldsymbol{g}^\star\rangle\right)}^{\text{Gradient Angle Part}}$$
$$+ \underbrace{\mu_{\boldsymbol{\xi}}^2 \|\nabla_{\boldsymbol{\xi}} f(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k)\|_2^2 + \mu_{\boldsymbol{\gamma}}^2 \tfrac{\|\boldsymbol{x}^\star\|_2^2}{m} \|\nabla_{\boldsymbol{\gamma}}^{\perp} f(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k)\|_2^2}_{\text{Gradient Magnitude Part}} < \Delta(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k)$$

  3. **Projection** on convex set $\mathcal{C}_\rho$ ensures $\Delta(\boldsymbol{\xi}_{k+1}, \boldsymbol{\gamma}_{k+1}) \leq \Delta(\boldsymbol{\xi}_{k+1}, \underline{\boldsymbol{\gamma}}_{k+1})$

- Let's have a look at the gradient:

$$\nabla^{\perp} f(\boldsymbol{\xi}, \boldsymbol{\gamma}) = \frac{1}{mp} \sum_{l=1}^{p} \begin{bmatrix} \boldsymbol{A}_l^\top \operatorname{diag}(\boldsymbol{\gamma}) \left(\operatorname{diag}(\boldsymbol{\gamma}) \boldsymbol{A}_l \boldsymbol{\xi} - \operatorname{diag}(\boldsymbol{d}) \boldsymbol{A}_l \boldsymbol{x}\right) \\ \boldsymbol{P}_{1_m^\perp} \operatorname{diag}(\boldsymbol{A}_l \boldsymbol{\xi}) \left(\operatorname{diag}(\boldsymbol{\gamma}) \boldsymbol{A}_l \boldsymbol{\xi} - \operatorname{diag}(\boldsymbol{d}) \boldsymbol{A}_l \boldsymbol{x}\right) \end{bmatrix} \xrightarrow[p \to \infty]{} \frac{1}{m} \begin{bmatrix} \|\boldsymbol{\gamma}\|_2^2 \boldsymbol{\xi} - (\boldsymbol{\gamma}^\top \boldsymbol{d}) \boldsymbol{x} \\ \|\boldsymbol{\xi}\|_2^2 \boldsymbol{\varepsilon} - (\boldsymbol{\xi}^\top \boldsymbol{x}) \boldsymbol{\omega} \end{bmatrix}$$

Initialisation Proximity and Gradient Regularity

- 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*

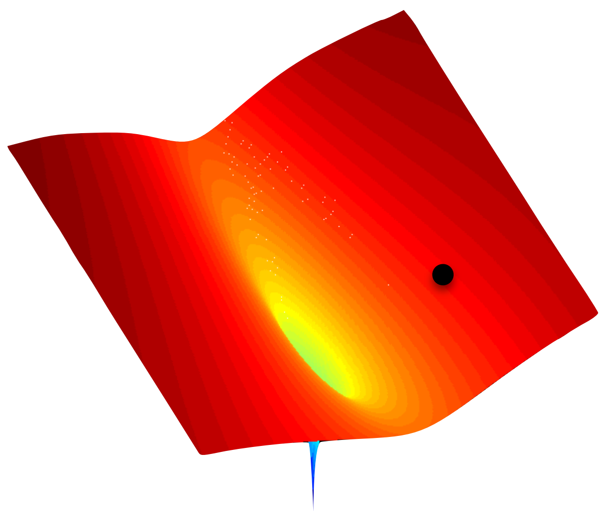
$$1 - Ce^{-c\epsilon^2 mp} - (mp)^{-t}$$

for some $C, c > 0$, *that* $\|\boldsymbol{\xi}_0 - \boldsymbol{x}^\star\|_2 \leq \epsilon \|\boldsymbol{x}^\star\|_2$ *provided* $n \gtrsim t \log(mp)$ *and*

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

Since $\boldsymbol{\gamma}_0 = 1_m$ *we also have* $\|\boldsymbol{\gamma}_0 - \boldsymbol{d}^\star\|_\infty \leq \rho < 1$. *Thus* $(\boldsymbol{\xi}_0, \boldsymbol{\gamma}_0) \in \mathcal{D}_{\kappa,\rho}$ *with the same probability and* $\kappa := \sqrt{\epsilon^2 + \rho^2} \leq \sqrt{2}$.

Initialise . . .



Initialisation Proximity and Gradient Regularity

Proposition (Regularity condition in $\mathcal{D}_{\kappa,\rho}$). *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* $\sqrt{mp} \gtrsim \delta^{-2}(n+m)\log(\frac{n}{\delta})$, *with probability exceeding*

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

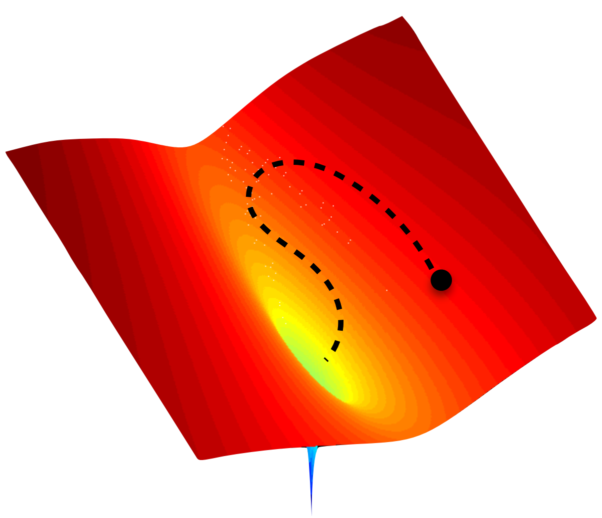
for some $C, c > 0$, *we have that for all* $(\boldsymbol{\xi}, \boldsymbol{\gamma}) \in \mathcal{D}_{\kappa,\rho}$,

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

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

for $\eta := 2(1-9\rho-2\delta) > 0$, $L := 4\sqrt{2}[1+\rho+(1+\kappa)\|\boldsymbol{x}^\star\|_2]$.

Initialise . . .

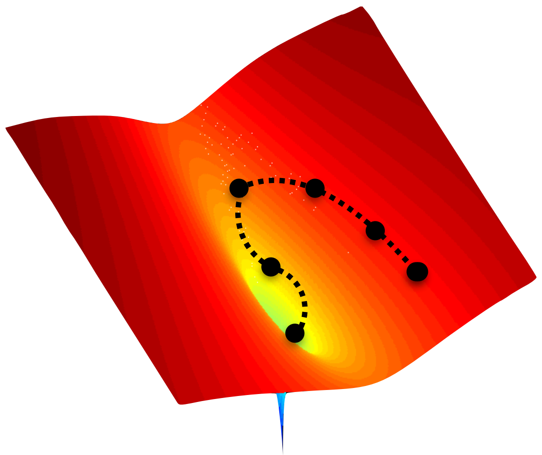


. . . and Converge

$\sqrt{mp} \gtrsim \delta^{-2}(n+m)\log\left(\frac{n}{\delta}\right)$

A Convergence Guarantee

- Under the previous conditions, we can bound the *error decay* of projected gradient descent when run in a neighbourhood of the global minimiser.
- The projection step serves to ensure *theoretically* that the neighbourhood does not change (*i.e.*, for the regularity condition)



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

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

for some $C, c > 0$*, our descent algorithm with* $\mu_{\boldsymbol{\xi}} := \mu, \mu_{\boldsymbol{\gamma}} := \mu \frac{m}{\|\boldsymbol{x}^\star\|_2^2}$ *has error decay*

$$\Delta(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k) \leq \left(1 - \eta\mu + \tfrac{L^2}{\tau}\mu^2\right)^k (\epsilon^2 + \rho^2)\|\boldsymbol{x}^\star\|_2^2, (\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k) \in \mathcal{D}_{\kappa,\rho}$$

at any iteration $k > 0$ *provided* $\mu \in \left(0, {}^{\tau\eta}/{}_{L^2}\right)$, $\tau := \min\{1, {}^{\|\boldsymbol{x}^\star\|_2^2}/{}_{m}\}$. *Hence,* $\Delta(\boldsymbol{\xi}_k, \boldsymbol{\gamma}_k) \underset{k\to\infty}{\longrightarrow} 0$.

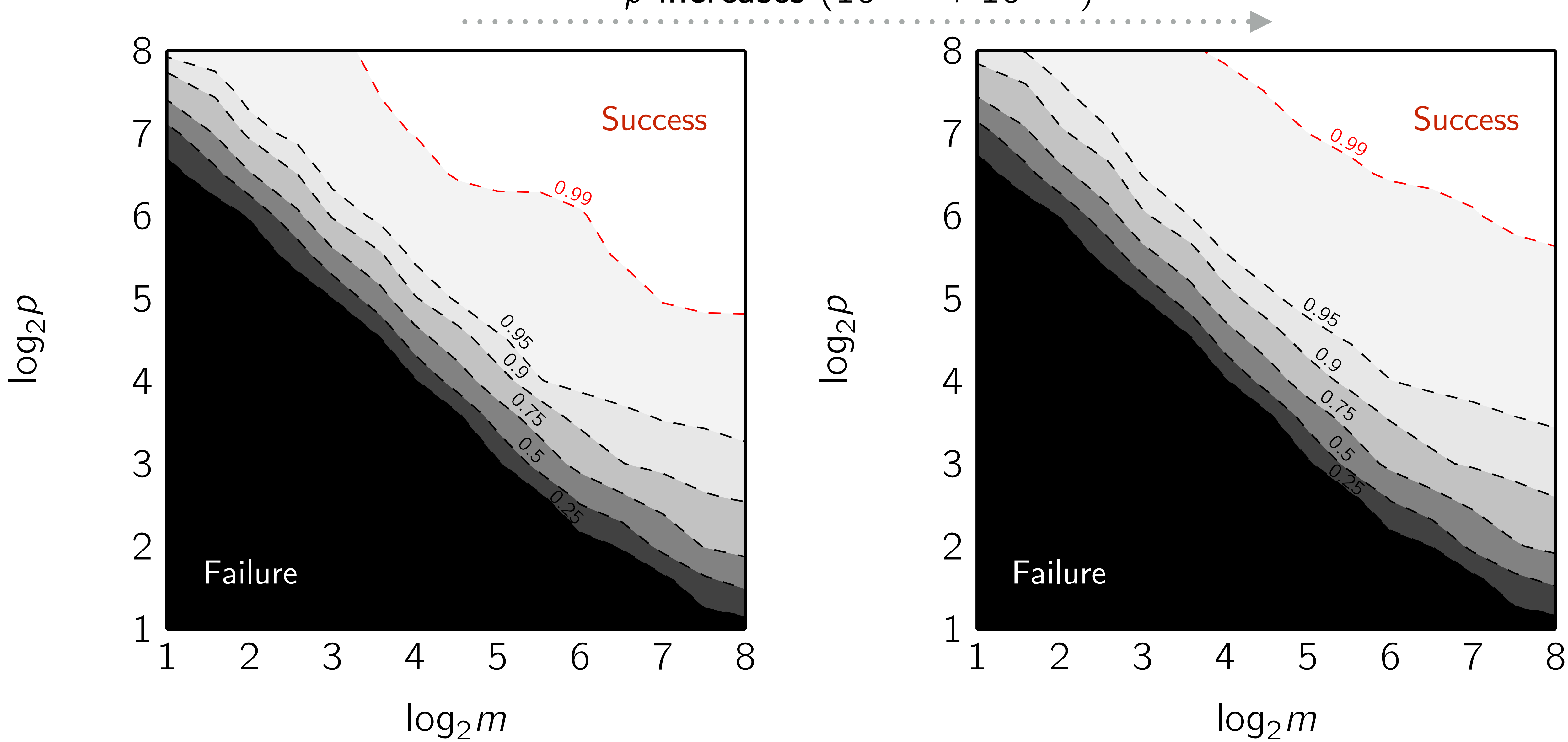
Empirical Phase Transition

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

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

for 256 randomly generated problem instances (per point).

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

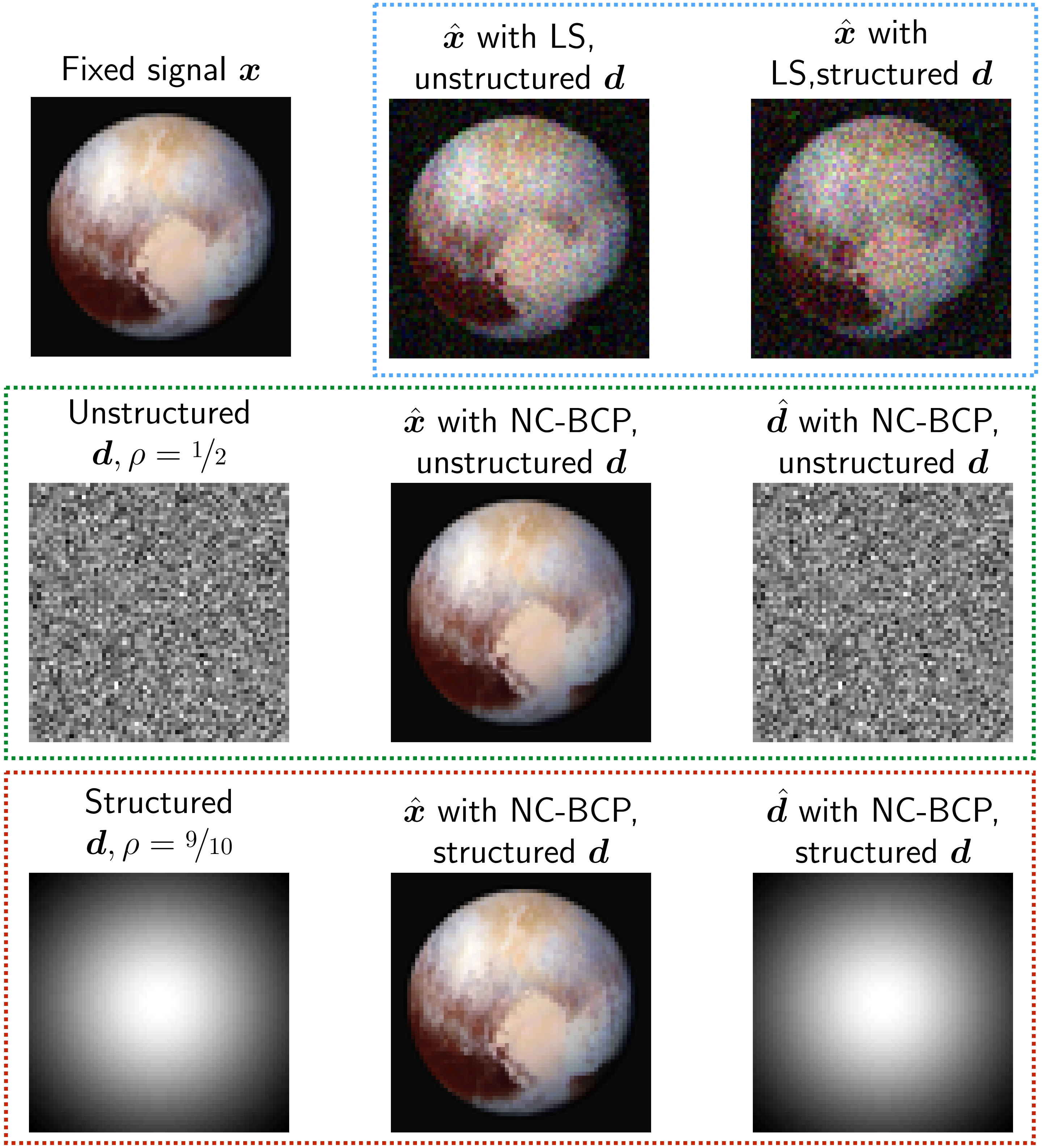


(Randomised) Computational Imaging

- 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:

$$\max\left\{\frac{\|\hat{\boldsymbol{d}}-\boldsymbol{d}^\star\|_2}{\|\boldsymbol{d}^\star\|_2}, \frac{\|\hat{\boldsymbol{x}}-\boldsymbol{x}^\star\|_2}{\|\boldsymbol{x}^\star\|_2}\right\} \approx -147.38\,\mathrm{dB}$$

- 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* $\boldsymbol{A}_l$ (not covered by current theory).



Conclusion

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

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

  - **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.

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.

Thank you for your attention.

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