# Mathematics of Data: From Theory to Computation

Prof. Volkan Cevher
*volkan.cevher@epfl.ch*

## *Supplementary Material: Time-Data tradeoff*

Laboratory for Information and Inference Systems (LIONS)
École Polytechnique Fédérale de Lausanne (EPFL)

**EE-556** (Fall 2025)

# License Information for Mathematics of Data Slides

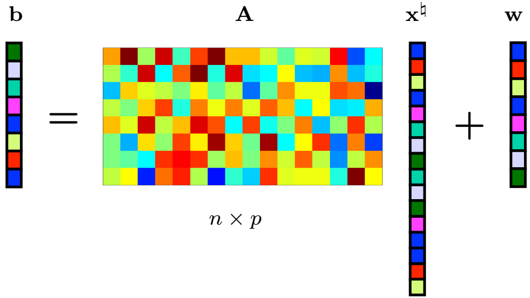
- This work is released under a Creative Commons License with the following terms:
- **Attribution**
  - The licensor permits others to copy, distribute, display, and perform the work. In return, licensees must give the original authors credit.
- **Non-Commercial**
  - The licensor permits others to copy, distribute, display, and perform the work. In return, licensees may not use the work for commercial purposes – unless they get the licensor's permission.
- **Share Alike**
  - The licensor permits others to distribute derivative works only under a license identical to the one that governs the licensor's work.
- Full Text of the License

# A simple *regression* model

$$\boxed{b_i = h_{\mathbf{x}^\natural}(\mathbf{a}_i)}$$

$\mathbf{x}^\natural$ : unknown function parameters
$\mathbf{a}_i$ : input
$\mathbf{b}_i$ : response / output

**Linear model:**



$$\boxed{\mathbf{b}_i = \langle \mathbf{a}_i, \mathbf{x}^\natural \rangle + \mathbf{w}_i}$$

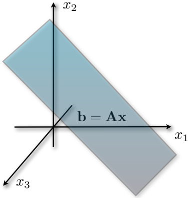
**Applications:** **Compressive sensing, machine learning, theoretical computer science...**

# A simple *regression* model and many *practical* questions

$$\boxed{\mathbf{b}_i = \langle \mathbf{a}_i, \mathbf{x}^\natural \rangle + \mathbf{w}_i}$$

$\mathbf{x}^\natural$ : unknown function parameters
$\mathbf{a}_i$ : input
$\mathbf{b}_i$ : response / output
$\mathbf{w}_i$ : perturbations / noise

- Estimation: find $\mathbf{x}^\star$ to minimize $\|\mathbf{x}^\star - \mathbf{x}^\natural\|$
- Prediction: find $\mathbf{x}^\star$ to minimize $L\left(\langle \mathbf{a}_i, \mathbf{x}^\star \rangle, \langle \mathbf{a}_i, \mathbf{x}^\natural \rangle\right)$
- Decision: choose $\mathbf{a}_i$ for estimation or prediction



***A difficult estimation challenge when $n < p$:***

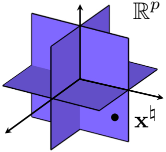
**Nullspace (null) of $\mathbf{A}$**: $\mathbf{x}^\natural + \boldsymbol{v} \to \mathbf{b}, \quad \forall \boldsymbol{v} \in \text{null}(\mathbf{A})$

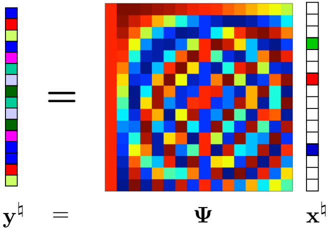
- **Needle in a haystack:** ***We need additional information on $\mathbf{x}^\natural$!***

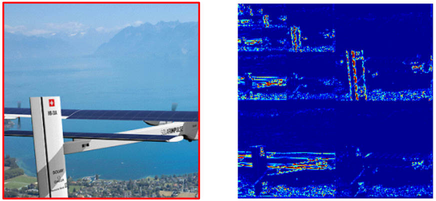
# A natural signal model

**Definition ($s$-sparse vector)**

A vector $\mathbf{x} \in \mathbb{R}^p$ is $s$-sparse if it has at most $s$ non-zero entries.



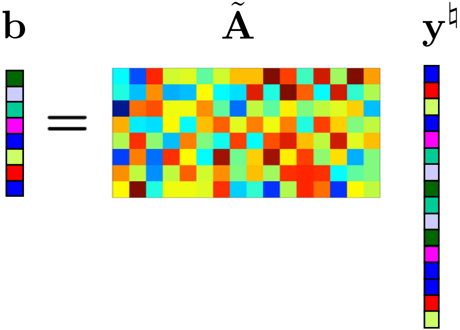




**Sparse representations**

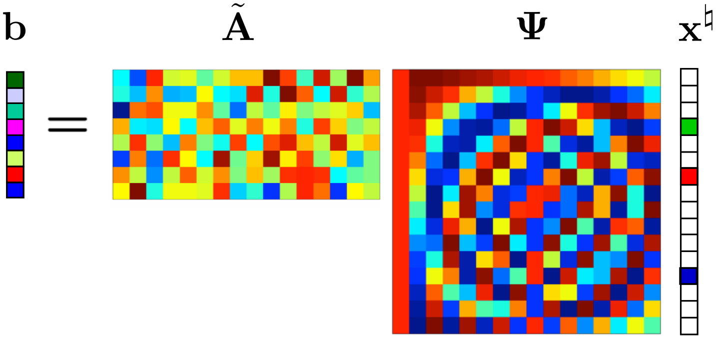
- $\mathbf{x}^\natural$: *sparse* transform coefficients
- Basis representations $\Psi \in \mathbb{R}^{p \times p}$
  - *Wavelets*, DCT, ...
- Frame representations $\Psi \in \mathbb{R}^{m \times p}$, $m > p$
  - Gabor, curvelets, shearlets, ...
- Other *dictionary* representations...

# Sparse representations strike back!



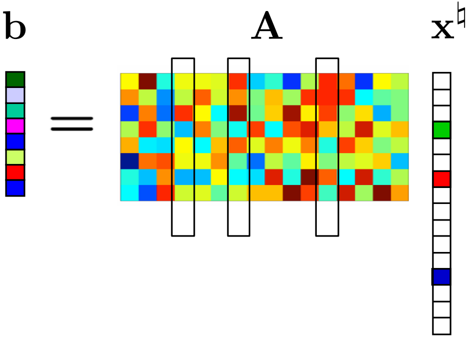
- $\mathbf{b} \in \mathbb{R}^n$, $\tilde{\mathbf{A}} \in \mathbb{R}^{n \times p}$, and $n < p$

# Sparse representations strike back!



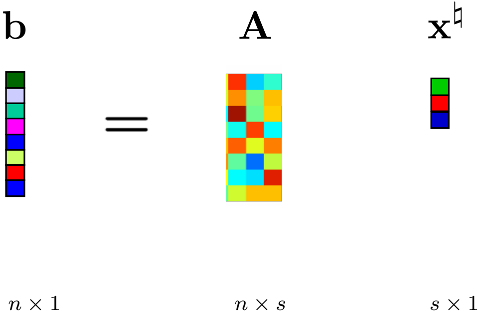
- $\mathbf{b} \in \mathbb{R}^n$, $\tilde{\mathbf{A}} \in \mathbb{R}^{n \times p}$, and $n < p$
- $\boldsymbol{\Psi} \in \mathbb{R}^{p \times p}$, $\mathbf{x}^\natural \in \mathbb{R}^p$, and $\|\mathbf{x}^\natural\|_0 \leq s < n$

# Sparse representations strike back!



- $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{n \times p}$, and $\mathbf{x}^\natural \in \mathbb{R}^p$, and $\|\mathbf{x}^\natural\|_0 \leq s < n < p$

# Sparse representations strike back!



**Observations:**
- ○ The matrix $\mathbf{A}$ effectively becomes *overcomplete*.
- ○ We could solve for $\mathbf{x}^\natural$ if we knew *the location of the non-zero entries of* $\mathbf{x}^\natural$.

# Enter sparsity

### A combinatorial approach for estimating $\mathbf{x}^\natural$ from $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$

We may consider the estimator with the least number of non-zero entries. That is,

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \{\|\,\mathbf{x}\,\|_0 : \|\,\mathbf{b} - \mathbf{A}\mathbf{x}\,\|_2 \leq \kappa\} \qquad (\mathcal{P}_0)$$

with some $\kappa \geq 0$. If $\kappa = \|\,\mathbf{w}\,\|_2$, then $\mathbf{x}^\natural$ is a feasible solution.

## Enter sparsity

### A combinatorial approach for estimating $\mathbf{x}^\natural$ from $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$

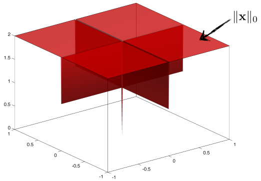
We may consider the estimator with the least number of non-zero entries. That is,

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \{\|\,\mathbf{x}\,\|_0 : \|\,\mathbf{b} - \mathbf{A}\mathbf{x}\,\|_2 \leq \kappa\} \qquad (\mathcal{P}_0)$$

with some $\kappa \geq 0$. If $\kappa = \|\,\mathbf{w}\,\|_2$, then $\mathbf{x}^\natural$ is a feasible solution.

- **$\mathcal{P}_0$ has the following characteristics:**
  - sample complexity: $\mathcal{O}(s)$
  - computational effort: NP-Hard
  - stability: No

$\|\,\mathbf{x}\,\|_0$ over the unit $\ell_\infty$-ball



# Enter sparsity

### A combinatorial approach for estimating $\mathbf{x}^\natural$ from $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$

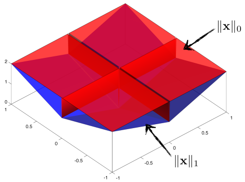
We may consider the estimator with the least number of non-zero entries. That is,

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \{\|\,\mathbf{x}\,\|_0 : \|\,\mathbf{b} - \mathbf{A}\mathbf{x}\,\|_2 \leq \kappa\} \qquad (\mathcal{P}_0)$$

with some $\kappa \geq 0$. If $\kappa = \|\,\mathbf{w}\,\|_2$, then $\mathbf{x}^\natural$ is a feasible solution.

- $\mathcal{P}_0$ **has the following characteristics:**
  - sample complexity: $\mathcal{O}(s)$
  - computational effort: NP-Hard
  - stability: No
- **Tightest convex relaxation:**
  - $\|\,\mathbf{x}\,\|_0^{**}$ is the biconjugate
  - i.e., Fenchel conjugate of Fenchel conjugate
- **Fenchel conjugate:**
  - $f^*(\mathbf{y}) := \sup_{\mathbf{x}:\mathrm{dom}(f)} \mathbf{x}^T\mathbf{y} - f(\mathbf{x})$.

$\|\,\mathbf{x}\,\|_1$ is the convex envelope of $\|\,\mathbf{x}\,\|_0$



**A technicality:** Restrict $\mathbf{x}^\natural \in [-1, 1]^p$.

# The role of convexity

## A convex candidate solution for $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_1 : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{w}\|_2, \|\mathbf{x}\|_\infty \leq 1 \right\}. \tag{SOCP}$$

## Theorem (A **model** recovery guarantee [8])

*Let* $\mathbf{A} \in \mathbb{R}^{n\times p}$ *be a matrix of i.i.d. Gaussian random variables with zero mean and variances* $1/n$*. For any* $t > 0$ *with probability at least* $1 - 6\exp\left(-t^2/26\right)$*, we have*

$$\|\mathbf{x}^\star - \mathbf{x}^\natural\|_2 \leq \left[ \frac{2\sqrt{2s\log(\frac{p}{s}) + \frac{5}{4}s}}{\sqrt{n} - \sqrt{2s\log(\frac{p}{s}) + \frac{5}{4}s} - t} \right] \|\mathbf{w}\|_2 := \varepsilon, \quad \textit{when } \|\mathbf{x}^\natural\|_0 \leq s.$$

**Observations:**

- ○ perfect recovery (i.e., $\varepsilon = 0$) with $n \geq 2s\log(\frac{p}{s}) + \frac{5}{4}s$ whp when $\mathbf{w} = 0$.
- ○ $\epsilon$-accurate solution in $k = \mathcal{O}\left(\sqrt{2p+1}\log(\frac{1}{\epsilon})\right)$ iterations via IPM with a total complexity of $\mathcal{O}(n^2 p^{1.5}\log(\frac{1}{\epsilon}))$ with each iteration requiring the solution of a structured $n \times 2p$ linear system.
- ○ robust to noise.

# A Time-Data conundrum — I

### A computational dogma

Running time of a learning algorithm increases with the size of the data.

# A Time-Data conundrum — I

### A computational dogma

Running time of a learning algorithm increases with the size of the data.

- ◦ Misaligned goals in the statistical and optimization disciplines

| Discipline | Goal | Metric |
|---|---|---|
| Optimization | reaching numerical $\epsilon$-accuracy | $\|\mathbf{x}^k - \mathbf{x}^\star\| \leq \epsilon$ |
| Statistics | learning $\varepsilon$-accurate model | $\|\mathbf{x}^\star - \mathbf{x}^\natural\| \leq \varepsilon$ |

- ◦ Main issue: **$\epsilon$ and $\varepsilon$ are NOT the same but should be treated jointly!**

# A Time-Data conundrum — II

## A stylized formalization of the time-data tradeoff

The goals of optimization and statistical modeling are tightly connected:

$$\underbrace{\|\mathbf{x}^k - \mathbf{x}^\natural\|}_{\text{learning quality}} \leq \underbrace{\|\mathbf{x}^k - \mathbf{x}^\star\|}_{\epsilon\text{: needs ``time'' } t(k)} + \underbrace{\|\mathbf{x}^\star - \mathbf{x}^\natural\|}_{\varepsilon\text{: needs ``data''} n},$$

$\mathbf{x}^\natural$: true model in $\mathbb{R}^p$
$\mathbf{x}^\star$: statistical model estimate
$\mathbf{x}^k$: numerical solution at iteration $k$

- As the number of data samples $n$ increases with a fixed optimization formulation,

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \{\|\,\mathbf{x}\,\|_1 : \|\,\mathbf{b} - \mathbf{A}\mathbf{x}\,\|_2 \leq \|\,\mathbf{w}\,\|_2, \|\mathbf{x}\|_\infty \leq 1\}$$

  - numerical methods take longer time $t$ to reach $\epsilon$-accuracy
    - e.g., per-iteration time to solve an $n \times 2p$ linear system
  - statistical model estimates $\varepsilon$ become more precise when $\|\mathbf{w}\|_2 = \mathcal{O}(\sqrt{n})$
    - $\varepsilon = \dfrac{2\sqrt{2s\log(\frac{p}{s}) + \frac{5}{4}s}}{\sqrt{n} - \sqrt{2s\log(\frac{p}{s}) + \frac{5}{4}s} - t}\|\,\mathbf{w}\,\|_2$, with probability $1 - 6\exp(-t^2/26)$.

# A Time-Data conundrum — II

## A stylized formalization of the time-data tradeoff

The goals of optimization and statistical modeling are tightly connected:

$$\underbrace{\|\mathbf{x}^k - \mathbf{x}^\natural\|}_{\leq \bar{\varepsilon}(t(k),n)} \leq \underbrace{\|\mathbf{x}^k - \mathbf{x}^\star\|}_{\epsilon\text{: needs ``time'' } t(k)} + \underbrace{\|\mathbf{x}^\star - \mathbf{x}^\natural\|}_{\varepsilon\text{: needs ``data''} n},$$

| | |
|---|---|
| $\mathbf{x}^\natural$: | true model in $\mathbb{R}^p$ |
| $\mathbf{x}^\star$: | statistical model estimate |
| $\mathbf{x}^k$: | numerical solution at iteration $k$ |
| $\bar{\varepsilon}(t(k),n)$: | actual learning quality at time $t(k)$ with $n$ samples |

- As the number of data samples $n$ increases with a fixed optimization formulation,

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \{\|\,\mathbf{x}\,\|_1 : \|\,\mathbf{b} - \mathbf{A}\mathbf{x}\,\|_2 \leq \|\,\mathbf{w}\,\|_2, \|\mathbf{x}\|_\infty \leq 1\}$$

  - numerical methods take longer time $t$ to reach $\epsilon$-accuracy
    - e.g., per-iteration time to solve an $n \times 2p$ linear system
  - statistical model estimates $\varepsilon$ become more precise when $\|\mathbf{w}\|_2 = \mathcal{O}(\sqrt{n})$
    - $\varepsilon = \frac{2\sqrt{2s\log(\frac{p}{s}) + \frac{5}{4}s}}{\sqrt{n} - \sqrt{2s\log(\frac{p}{s}) + \frac{5}{4}s} - t}\|\,\mathbf{w}\,\|_2$, with probability $1 - 6\exp(-t^2/26)$.

**"Time" effort has significant diminishing returns on $\varepsilon$ in the underdetermined case*** (cf., [6, 3, 9, 5, 4])

* "Data" effort also exhibits a similar behavior in the overdetermined case when a signal prior is used due to noise!

## Data as a computational resource

### A stylized formalization of the time-data tradeoff

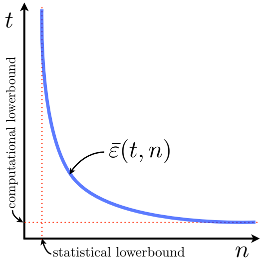
The goals of optimization and statistical modeling are tightly connected:

$$\underbrace{\|\mathbf{x}^{k(t)} - \mathbf{x}^\natural\|}_{\leq \bar{\varepsilon}(t,n)} \leq \underbrace{\|\mathbf{x}^{k(t)} - \mathbf{x}^\star\|}_{\epsilon\text{: needs ``time'' } t} + \underbrace{\|\mathbf{x}^\star - \mathbf{x}^\natural\|}_{\varepsilon\text{: needs ``data''} n},$$

$\mathbf{x}^\natural$: true model in $\mathbb{R}^p$
$\bar{\varepsilon}(t,n)$: actual model precision at time $t$ with $n$ samples

Rest of the lecture:

- ○ estimator formulation and sample complexity
- ○ a "continuous" time-data tradeoff
- ○ a different, algorithmic tradeoff with SGD



# Sample complexity analysis

### Convex optimization formulation for the estimator

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \{f(\mathbf{x}) : \mathbf{b} = \mathbf{A}\mathbf{x}\},$$

where $f : \mathbb{R}^p \to \mathbb{R} \cup \{-\infty, \infty\}$ is a convex function.

### Sample complexity

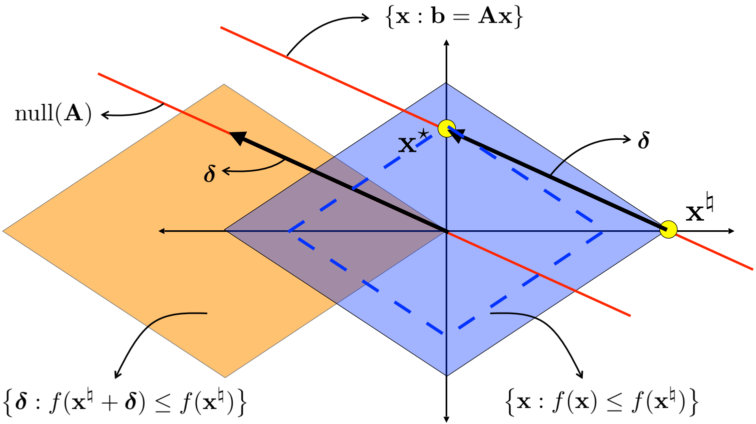
Assume that $A \in \mathbb{R}^{n\times p}$ is a matrix of independent identically distributed (i.i.d.) Gaussian random variables.

What is the minimum number of samples $n$ such that $\mathbf{x}^\star = \mathbf{x}^\natural$ with high probability?

# Characterization of the error vector

$$\boxed{\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \{f(\mathbf{x}) : \mathbf{b} = \mathbf{A}\mathbf{x}\}}$$

Define the error vector $\boldsymbol{\delta} := \mathbf{x}^\star - \mathbf{x}^\natural$.

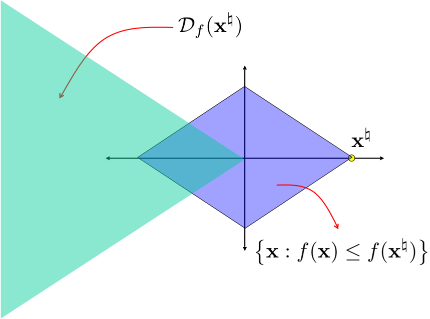


# Descent cone

## Definition (Descent cone)

*Let $f : \mathbb{R}^p \to \mathbb{R} \cup \{-\infty, \infty\}$ be a proper lower-semicontinuous function. The descent cone of $f$ at $\mathbf{x}^\natural$ is defined as*

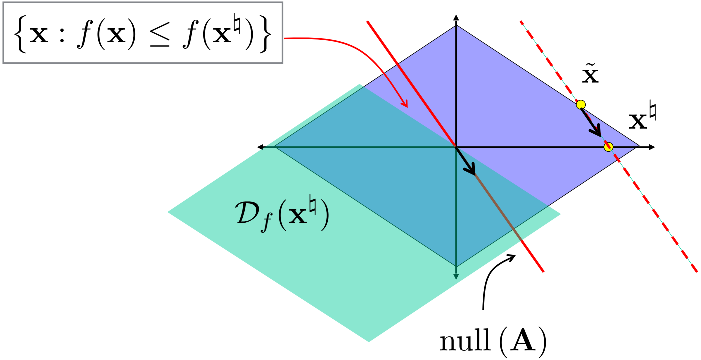
$$\mathcal{D}_f(\mathbf{x}^\natural) := \text{cone}\left(\left\{\delta : f(\mathbf{x}^\natural + \delta) \leq f(\mathbf{x}^\natural)\right\}\right).$$



# Condition for exact recovery in the *noiseless* case

**Proposition (Condition for exact recovery)**

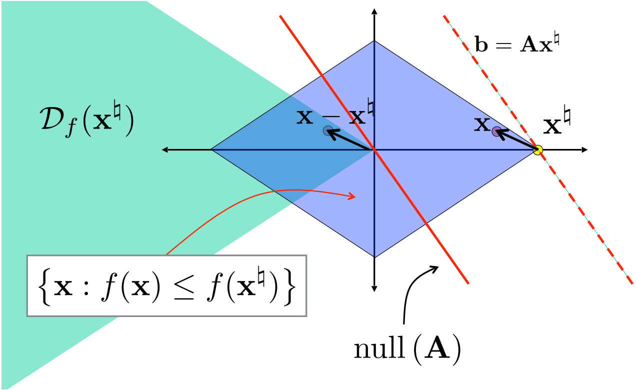
*We have successful recovery, i.e.,* $\boldsymbol{\delta} := \mathbf{x}^\star - \mathbf{x}^\natural = 0$ *with* $\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \{f(\mathbf{x}) : \mathbf{b} = \mathbf{A}\mathbf{x}\}$*, if and only if* $\mathrm{null}(\mathbf{A}) \cap \mathcal{D}_f(\mathbf{x}^\natural) = \{0\}$.



# Condition for exact recovery in the *noiseless* case

## Proposition (Condition for exact recovery)

*We have successful recovery, i.e.,* $\boldsymbol{\delta} := \mathbf{x}^\star - \mathbf{x}^\natural = 0$ *with* $\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \{f(\mathbf{x}) : \mathbf{b} = \mathbf{A}\mathbf{x}\}$, *if and only if* $\mathrm{null}(\mathbf{A}) \cap \mathcal{D}_f(\mathbf{x}^\natural) = \{0\}$.



# Statistical dimension and approximate kinematic formula

Now we have

$$\mathbb{P}\left\{\mathbf{x}^\star = \mathbf{x}^\natural\right\} = \mathbb{P}\left\{\mathrm{null}(\mathbf{A}) \cap \mathcal{D}_f(\mathbf{x}^\natural) = \{0\}\right\}.$$

## Definition (Statistical dimension [1][1])

*Let $\mathcal{C} \subseteq \mathbb{R}^p$ be a closed convex cone. The statistical dimension of $\mathcal{C}$ is defined as*

$$d(\mathcal{C}) := \mathbb{E}\left[\|\mathrm{proj}_{\mathcal{C}}(\mathbf{g})\|_2^2\right].$$

## Theorem (Approximate kinematic formula [1])

*Let $A \in \mathbb{R}^{n\times p}$, $n < p$, be a matrix of i.i.d. standard Gaussian random variables, and let $\mathcal{C} \subseteq \mathbb{R}^p$ be a closed convex cone. Let $\eta \in (0,1)$ Then*

$$\begin{aligned} n \geq d(\mathcal{C}) + c_\eta\sqrt{p} \quad &\Rightarrow \quad \mathbb{P}\left\{\mathrm{null}(\mathbf{A}) \cap \mathcal{C} = \{0\}\right\} \geq 1-\eta; \\ n \leq d(\mathcal{C}) - c_\eta\sqrt{p} \quad &\Rightarrow \quad \mathbb{P}\left\{\mathrm{null}(\mathbf{A}) \cap \mathcal{C} = \{0\}\right\} \leq \eta, \end{aligned}$$

*where* $c_\eta := \sqrt{8\log(4/\eta)}$.

---

[1] The statistical dimension is closely related to the Gaussian complexity [2], Gaussian width [7], and Gaussian squared complexity [6].

## Probability of exact recovery

### Corollary

*For any* $\eta \in (0,1)$,

$$\begin{aligned} n &\geq d(\mathcal{D}_f(\mathbf{x}^\natural)) + c_\eta \sqrt{p} \quad &\Rightarrow \quad \mathbb{P}\left\{\mathbf{x}^\star = \mathbf{x}^\natural\right\} \geq 1 - \eta; \\ n &\leq d(\mathcal{D}_f(\mathbf{x}^\natural)) - c_\eta \sqrt{p} \quad &\Rightarrow \quad \mathbb{P}\left\{\mathbf{x}^\star = \mathbf{x}^\natural\right\} \leq \eta, \end{aligned}$$

*where* $c_\eta := \sqrt{8\log(4/\eta)}$.

- ○ There is a *phase transition* at $n \approx d(\mathcal{D}_f(\mathbf{x}^\natural))$.

### Examples ([1])

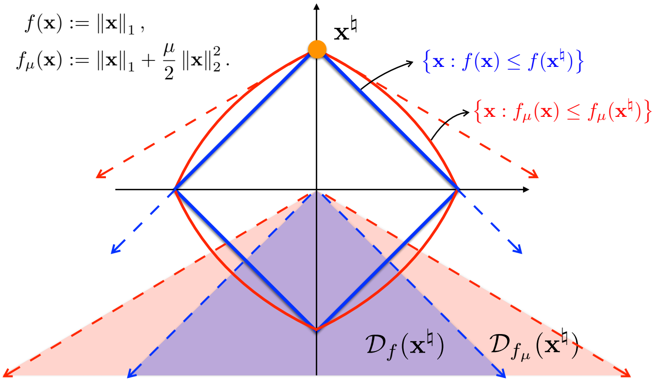
- ○ Let $f(\mathbf{x}) := \|\,\mathbf{x}\,\|_1$, and let $\mathbf{x}^\natural \in \mathbb{R}^p$ be $s$-sparse. Then $d(\mathcal{D}_f(\mathbf{x}^\natural)) \leq 2s\log(p/s) + (5/4)s$.
- ○ Let $f(\mathbf{x}) := \|\,\mathbf{X}\,\|_*$, and let $\mathbf{X}^\natural \in \mathbb{R}^{p\times p}$ of rank $r$. Then $d(\mathcal{D}_f(\mathbf{x}^\natural)) \leq 3r(2p - r)$.

# Smoothing increases the statistical dimension

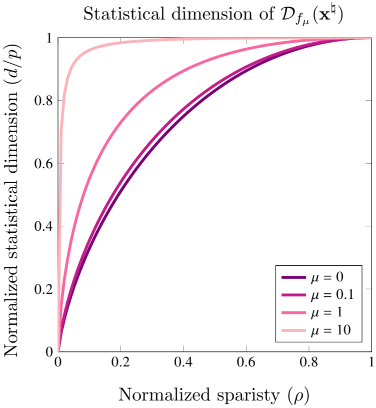
### Key properties of the statistical dimension [1]

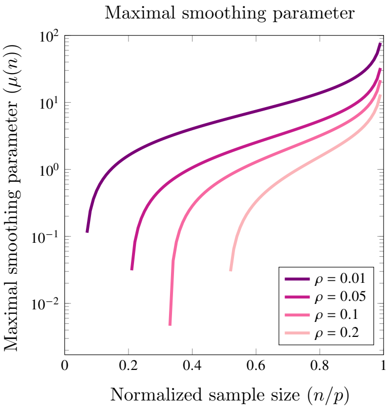
- The statistical dimension is invariant under unitary transformations (rotations).
- Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be closed convex cones. If $\mathcal{C}_1 \subseteq \mathcal{C}_2$, then $d(\mathcal{C}_1) \leq d(\mathcal{C}_2)$.

**The larger the statistical dimension is, the more number of observations is required.**



# Numerical results for the statistical dimension and $\mu(n)$





# Smoothing decreases the computational cost

- Consider the estimator,

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ f_\mu(\mathbf{x}) : \mathbf{b} = \mathbf{A}\mathbf{x}, \|\mathbf{x}\|_\infty \leq \|\mathbf{x}^\natural\|_\infty \right\}, \quad \mu \in [0,\infty).$$

### Proposition

*Let $\mu > 0$ and $f(\mathbf{x}) = \|\mathbf{x}\|_1$. Consider solving (1) with a primal-dual method as in [4, 5]. The output after the $k$-th iteration, $\mathbf{x}^k$, satisfies*

$$\|\,\mathbf{x}^\star - \mathbf{x}^k\,\|_2 \leq \frac{4p\kappa(\mathbf{A})\left[\rho(1+\mu\|\,\mathbf{x}^\star\,\|_\infty)^2 + (1-\rho)\right]}{\mu k} \propto \left.\frac{1}{\mu k}\right|_{\rho\ll 1},$$

*where $\rho := s/p$, $s$ being the number of non-zero entries in $\mathbf{x}^\star$, and $\kappa(\mathbf{A})$ denotes the restricted condition number of $\mathbf{A}$.*

**Observations:**

- When $\rho \ll 1$, the number of iterations $k$ to achieve the required precision decreases.
- In fact, we need $1/(\mu\varepsilon)$ iterations to have an error bound $\|\,\mathbf{x}^\star - \mathbf{x}^k\,\|_2 \leq \varepsilon$ for a fixed $\epsilon > 0$.

# Time-data tradeoff

- Define the maximal smoothing parameter

$$\mu(n) := \arg\max_{\mu > 0} \left\{ \mu : d\left(\mathcal{D}_{f_\mu}(\mathbf{x}^\natural)\right) \leq n \right\}.$$

- Consider the "conservative" estimator in probability,

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x} \in \mathbb{R}^p} \left\{ f_\mu(\mathbf{x})\big|_{\mu = \frac{1}{4}\mu(n)} : \mathbf{b} = \mathbf{A}\mathbf{x} \right\}.$$

## Corollary

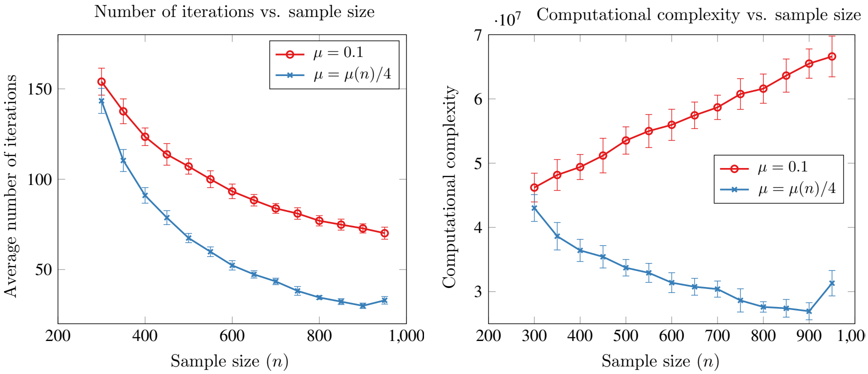
*Let* $\rho := s/p \ll 1$*. Then we have, with high probability,* $\mathbf{x}^\star = \mathbf{x}^\natural$*, and*

$$\| \mathbf{x}^\natural - \mathbf{x}^k \|_2 \propto \frac{1}{\mu(n) k}.$$

*Therefore, to achieve the error bound,* $\| \mathbf{x}^\natural - \mathbf{x}^k \|_2 \leq \varepsilon$ *for a fixed* $\varepsilon > 0$*, it suffices to choose*

$$k = O\left(\frac{1}{\mu(n)}\right).$$

# A numerical result for the time-data tradeoff



# References I

[1] Dennis Amelunxen, Martin Lotz, Michael B. McCoy, and Joel A. Tropp.
Living on the edge: Phase transitions in convex programs with random data.
*Information and Inference*, 3:224–294, 2014.
arXiv:1303.6672v2 [cs.IT].
(Cited on pages 24, 25, and 26.)

[2] Peter L. Barlett and Shahar Mendelson.
Rademacher and Gaussian complexities: Risk bounds and structural results.
*J. Mach. Learn. Res.*, 3, 2002.
(Cited on page 24.)

[3] Léon Bottou and Oliver Bousquet.
The tradeoffs of large scale learning.
In *Adv. Neur. Inf. Proc. Sys. (NIPS)*, 2007.
(Cited on pages 16 and 17.)

[4] John J Bruer, Joel A Tropp, Volkan Cevher, and Stephen Becker.
Time–data tradeoffs by aggressive smoothing.
*Advances in Neural Information Processing Systems*, 27, 2014.
(Cited on pages 16, 17, and 28.)

## References II

[5] John J Bruer, Joel A Tropp, Volkan Cevher, and Stephen R Becker.
Designing statistical estimators that balance sample size, risk, and computational cost.
*IEEE Journal of Selected Topics in Signal Processing*, 9(4):612–624, 2015.
(Cited on pages 16, 17, and 28.)

[6] Venkat Chandrasekaran and Michael I. Jordan.
Computational and statistical tradeoffs via convex relaxation.
*Proc. Nat. Acad. Sci.*, 110(13):E1181–E1190, 2013.
(Cited on pages 16, 17, and 24.)

[7] Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan S. Willsky.
The convex geometry of linear inverse problems.
*Found. Comp. Math.*, 12:805–849, 2012.
(Cited on page 24.)

[8] Samet Oymak, Christos Thrampoulidis, and Babak Hassibi.
The squared-error of generalized lasso: A precise analysis.
In *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1002–1009. IEEE, 2013.
(Cited on page 13.)

# References III

[9] Shai Shalev-Shwartz and Nathan Srebro.
Svm optimization: inverse dependence on training set size.
In *Proceedings of the 25th international conference on Machine learning*, pages 928–935, 2008.
(Cited on pages 16 and 17.)