Existence of solutions to the nonlinear equations characterizing the precise error of M-estimators

Takuya Koriyama

University of Chicago

September 4, 2024

Linear model and (regularized) M-estimator

- Consider the linear model
$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{z}$$
where $\boldsymbol{A} \in \mathbb{R}^{n \times p}$ is the design matrix, $\boldsymbol{x}_0 \in \mathbb{R}^p$ is the signal of interest, and $\boldsymbol{z} \in \mathbb{R}^n$ is a noise vector.
- Given data $(\boldsymbol{y}, \boldsymbol{A})$, we estimate $\boldsymbol{x}_0$ by the regularized M-estimator $\hat{\boldsymbol{x}}$
$$\hat{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^p} \Big\{ \sum_{i=1}^{n} \rho(y_i - \boldsymbol{e}_i^\top \boldsymbol{A}\boldsymbol{x}) + \sum_{j=1}^{p} f(x_j) \Big\},$$
where ρ is a convex loss and f is a convex regularizer.
- We are interested in the behavior of the risk $\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|^2$.

Numerical simulation 1: Unregularized robust M-estimator

- When $f = 0$, $\hat{\boldsymbol{x}}$ is the unregularized M-estimator

$$\hat{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^p} \sum_{i=1}^{n} \rho(y_i - \boldsymbol{e}_i^\top \boldsymbol{A}\boldsymbol{x}),$$

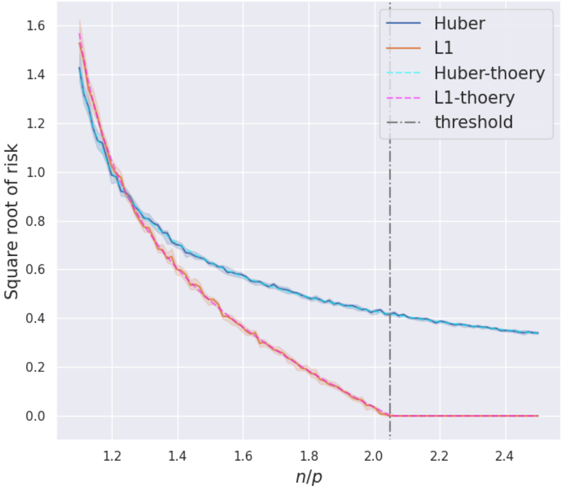
where ρ is a robust loss such as

$$\text{(L1) } \rho(x) = |x|, \qquad \text{(Huber) } \rho(x) = \begin{cases} x^2/2 & |x| \leq 1 \\ |x| - 1/2 & |x| \geq 1 \end{cases}$$

- The noise vector $\boldsymbol{z}$ is generated according to

$$\boldsymbol{z} = (z_i)_{i=1}^n \overset{\text{iid}}{\sim} 0.8\delta_0 + 0.2N(0,1).$$

Numerical simulation 1: Unregularized robust M-estimator



- $p = 300$, iterate=10.
- For L1 loss, perfect recovery $\hat{\boldsymbol{x}} = \boldsymbol{x}_0$ holds for sufficiently large n.
- For the Huber loss, the perfect recovery is impossible.

Numerical simulation 2: L1 loss and L1 penalty

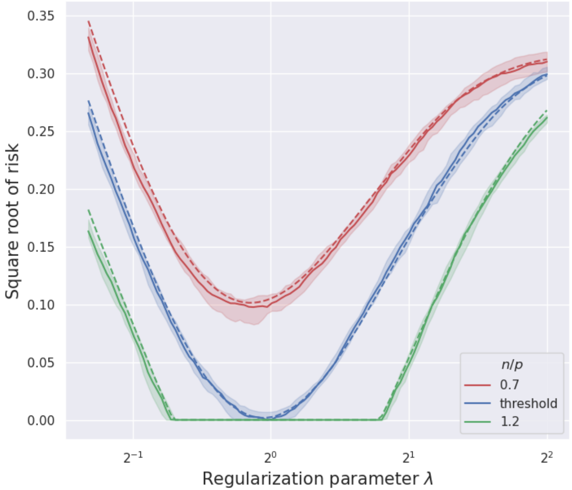
- Set $\rho(\cdot) = |\cdot|$ and $f(\cdot) = |\cdot|$ so that $\hat{\boldsymbol{x}}$ is the L1-penalized LAD estimator

$$\hat{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^p} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_1 + \lambda \|\boldsymbol{x}\|_1,$$

  where $\lambda > 0$ is a regularization parameter.
- The noise vector $\boldsymbol{z}$ and signal $\boldsymbol{x}_0$ are generated according to

$$\boldsymbol{z} = (z_i)_{i=1}^n \overset{\text{iid}}{\sim} 0.7\delta_0 + 0.3N(0,1).$$
$$\boldsymbol{x}_0 = (x_{0j})_{j=1}^p \overset{\text{iid}}{\sim} 0.9\delta_0 + 0.1N(0,10)$$

Numerical simulation 2: L1 loss and L1 penalty



- $p = 1000$, iterate=10.
- When $\delta = n/p$ is small, perfect recovery is impossible for any λ
- When $\delta = n/p$ is large, perfect recovery is possible for some λ.

Goal: characterization of risk behavior

- Goal: Characterization of the risk of the (regularized) M-estimator

$$\hat{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x}\in\mathbb{R}^p}\left\{\sum_{i=1}^{n}\rho(y_i - \boldsymbol{e}_i^\top \boldsymbol{A}\boldsymbol{x}) + \sum_{j=1}^{p} f(x_j)\right\}.$$

- More precisely, we are interested in the questions below

What is the condition under which perfect recovery is possible?

If the perfect recovery is possible, what is the minimum sample size n?

- To answer these questions, we consider the proportional asymptotic regime.

Proportional asymptotic regime

Assumption

- *The design matrix* $\boldsymbol{A} \in \mathbb{R}^{n \times p}$ *has iid entries* $N(0, 1/p)$.
- *Sample size* n *and feature dimension* p *are increasing such that*

$$n/p \to \delta \in (0, \infty) \quad (\textit{proportional regime})$$

- *The signal* $\boldsymbol{x}_0 \in \mathbb{R}^p$ *and the noise vector* $\boldsymbol{z} \in \mathbb{R}^n$ *have iid marginals*

$$\boldsymbol{x}_0 = (x_{01}, \ldots, x_{0p}) \stackrel{\text{iid}}{\sim} X, \qquad \boldsymbol{z} = (z_1, \ldots, z_n) \stackrel{\text{iid}}{\sim} Z,$$

for some scalar random variables (X, Z).

- Proportional regime has received attention from fields such as Deep Learning, Statistical Physics, Compressed sensing, etc.
- [Thrampoulidis et al., 2018] gives the precise asymptotics of risk behavior (next slide)

Asymptotic error of M-estimator in proportional regime

Theorem (Thrampoulidis et al. [2018])

Let α_* *be the solution to the nonlinear system below with positive unknown* $(\alpha, \beta, \kappa, \nu)$:

$$\begin{aligned} \alpha^2 &= \mathbb{E}\Big[\big(\mathrm{prox}[\nu^{-1}f](\nu^{-1}\beta H + X) - X\big)^2\Big] \\ \beta^2\kappa^2 &= \delta\,\mathbb{E}\Big[\big(\alpha G + Z - \mathrm{prox}[\kappa\rho](\alpha G + Z)\big)^2\Big] \\ \nu\alpha\kappa &= \delta\,\mathbb{E}\Big[G\cdot\big(\alpha G + Z - \mathrm{prox}[\kappa\rho](\alpha G + Z)\big)\Big] \\ \kappa\beta &= \mathbb{E}\Big[H\cdot\big(\mathrm{prox}[\nu^{-1}f](\nu^{-1}\beta H + X) - X\big)\Big] \end{aligned} \quad \text{for} \quad \begin{cases} G \sim \mathcal{N}(0,1), \\ H \sim \mathcal{N}(0,1), \\ X =^d x_{0j}, \\ Z =^d z_i. \end{cases},$$

where $\mathrm{prox}[f](u) = \arg\min_{x\in\mathbb{R}}(u-x)^2/2 + f(x)$ *is the proximal operator. Then, under the proportional regime* $p/n \to \delta$, *we have*

$$p^{-1}\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|^2 \to^P \alpha_*^2.$$

- The authors assumed that the nonlinear system admits a unique solution.
- The existence and uniqueness of solution to the nonlinear system are unknown for a general pair of loss and penalty.

Previous work on existence and uniqueness of solution is limited

When $\rho(x) = x^2/2$, the nonlinear system with four unknowns is reduced to a system with two unknowns (α, β).

- For $f(x) = x^2/2$ (Ridge), the solution has a closed form.
- For $f(x) = |x|$ (Lasso), Miolane and Montanari [2021] show the uniqueness and existence of the solution to the nonlinear system, but the proof is specific to the least square loss

When $f = 0$, the nonlinear system is reduced to a system with two unknowns.

- The uniqueness is proved by Donoho and Montanari [2016] for strongly convex losses ρ, which excludes a large family of robust losses such as L1 loss and Huber loss.
- The existence has not yet been shown.

Overview of main result

Theorem

1. *For given* (Z, X, ρ, f) *satisfying mild assumptions, there exists some* $\delta_0 \in (0, \infty]$ *s.t.*

$$\delta(=\lim \tfrac{n}{p}) > \delta_0 \Rightarrow \mathbb{P}(\textit{perfect recovery } \hat{\boldsymbol{x}} = \boldsymbol{x}_0) \to 1,$$
$$\delta < \delta_0 \Rightarrow \mathbb{P}(\textit{perfect recovery } \hat{\boldsymbol{x}} = \boldsymbol{x}_0) \to 0.$$

2. *When* $\delta < \delta_0$*, the nonlinear system admits a unique positive solution* $(\alpha_*, \beta_*, \kappa_*, \nu_*)$ *and* $p^{-1}\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|^2 \to^p \alpha_*^2 > 0$.

- The existence of threshold δ_0 is a consequence of the kinematic formula in Amelunxen et al. [2014] and Han and Ren [2022].
- The main contribution is to establish the second point. We prove it by constructing a 'dual' infinite-dimensional constrained optimization problem on a Hilbert space such that the existence and uniqueness of the solution to this dual optimization problem implies the existence and uniqueness of the solution to the nonlinear system.

Let us start with the unregularized case ($f = 0$)

Unregularized case ($f = 0$)

- In this case, $\hat{\boldsymbol{x}}$ is the unregularized M-estimator

$$\hat{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x}\in\mathbb{R}^p} \sum_{i=1}^{n} \rho(y_i - \boldsymbol{e}_i^\top \boldsymbol{A}\boldsymbol{x}).$$

- Assume $\delta = \lim \frac{n}{p} > 1$ otherwise $\hat{\boldsymbol{x}}$ is ill-defined.
- Using $\mathrm{prox}[f](x) = x$ for $f = 0$, the nonlinear system of interest is reduced to

$$\begin{aligned} \alpha^2 &= \delta\, \mathbb{E}\left[((\alpha G + Z) - \mathrm{prox}[\kappa\rho](\alpha G + Z))^2\right] \\ \alpha &= \delta\, \mathbb{E}\left[((\alpha G + Z) - \mathrm{prox}[\kappa\rho](\alpha G + Z))G\right] \end{aligned} \quad \text{for} \quad \begin{cases} G \sim \mathcal{N}(0,1), \\ Z =^d z_i \end{cases}$$

  with positive unknown (α, κ).
- $p^{-1}\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|^2 \to^p \alpha_*^2$ for the solution (α_*, κ_*) is proved by Donoho and Montanari [2016], El Karoui et al. [2013] among others, provided that such solution is uniquely exists.

Existence and uniqueness of solution in the unregularized case

Let δ_0 be the positive scalar

$$\delta_0 = \frac{1}{(1 - \inf_{t>0} \mathbb{E}[\mathrm{dist}(G, t\partial\rho(Z))^2])_+} \in (0, \infty],$$

where $\mathrm{dist}(\cdot, S) = \inf_{u \in \mathbb{R}} (\cdot - u)^2$ for any set $S \subset \mathbb{R}$.

Theorem (unregularized case)

Assume ρ is Lipschitz with $\{0\} = \arg\min_x \rho(x)$ and $\mathbb{P}(Z \neq 0) > 0$. Then,

$$\delta(= \lim \tfrac{n}{p}) > \delta_0 \Rightarrow \mathbb{P}(\textit{perfect recovery } \hat{\boldsymbol{x}} = \boldsymbol{x}_0) \to 1,$$

$$\delta < \delta_0 \Rightarrow \mathbb{P}(\textit{perfect recovery } \hat{\boldsymbol{x}} = \boldsymbol{x}_0) \to 0.$$

When $\delta < \delta_0$, the nonlinear system admits a unique positive solution (α_*, κ_*).

When $\delta_0 = \infty$, the perfect recovery is always impossible.
Q. When does $\delta_0 = \infty$ hold? (next slide)

When does the case $\delta_0 = \infty$ occur?

$$\begin{aligned} \delta > \delta_0 &\Rightarrow \mathbb{P}(\hat{\boldsymbol{x}} = \boldsymbol{x}_0) \to 1 \\ \delta < \delta_0 &\Rightarrow \mathbb{P}(\hat{\boldsymbol{x}} = \boldsymbol{x}_0) \to 0 \end{aligned} \quad \text{for} \quad \delta_0 = \frac{1}{(1 - \inf_{t>0} \mathbb{E}[\mathrm{dist}(G, t\partial\rho(Z))^2])_+}$$

Note $\delta_0 = +\infty$ iff $\inf_{t>0} \mathbb{E}[\mathrm{dist}(G, t\partial\rho(Z))^2] \geq 1$. For all $t > 0$,

$$\begin{aligned} \mathbb{E}[\mathrm{dist}(G, t\partial\rho(Z))^2] &\geq \mathbb{E}[I\{\rho \text{ is differentiable at } Z\}(G - t\rho'(Z))^2] \\ &\geq \mathbb{P}(\rho \text{ is differentiable at } Z) \quad (\text{by } G \sim N(0,1) \perp\!\!\!\perp Z), \end{aligned}$$

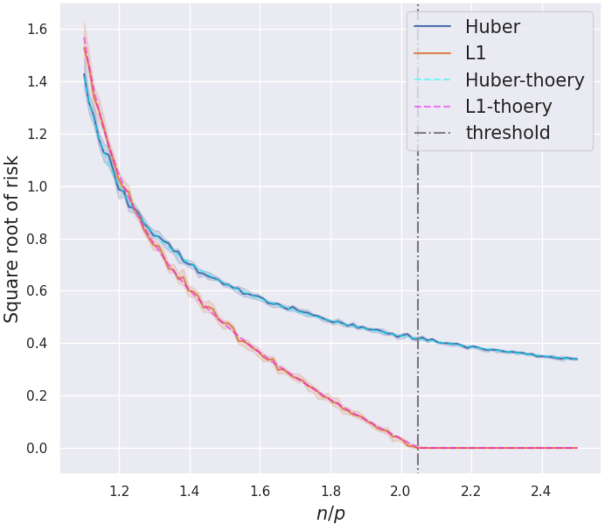
so the sufficient condition for $\delta_0 = +\infty$ is

$$\mathbb{P}(\rho \text{ is differentiable at } Z) = 1.$$

Example of (ρ, Z) satisfying this condition:

- ρ is differentiable such as the Huber loss
- ρ has a finite point of discontinuities, e.g., $\rho(x) = |x|$, and Z is continuous.

Numerical simulation



$Z = 0.8 \cdot \delta_0 + 0.2 \cdot N(0,1)$, $p = 300$, iterate=10, dash-line = theory

Proof outline: "Dual" optimization problem on Hilbert space

$(\alpha_*, \kappa_*) \in \mathbb{R}^2_{>0}$ solves the nonlinear system

⇅

$v_* \in \mathcal{H}$ solves $\begin{array}{l} \text{minimize}_{v \in \mathcal{H}} \ \mathcal{L}(v) \\ \text{subject to } \ \mathcal{G}(v) \leq 0 \end{array}$ for some $\begin{array}{l} \mathcal{L} : \mathcal{H} \to \mathbb{R} \\ \mathcal{G} : \mathcal{H} \to \mathbb{R} \end{array}$

where $\mathcal{H} = \{ v : \mathbb{R}^2 \to \mathbb{R}, \ \mathbb{E}[v(G, Z)^2] < +\infty \}$

Proof of existence of solution in the unregularized case

Let $\mathcal{H}$ be the Hilbert space

$$\mathcal{H} := \{v : \mathbb{R}^2 \to \mathbb{R} : \mathbb{E}[(G,Z)^2] < +\infty\},$$

and define $\mathcal{L} : \mathcal{H} \to \mathbb{R}$ and $\mathcal{G} : \mathcal{H} \to \mathbb{R}$ as

$$\mathcal{L} : \mathcal{H} \to \mathbb{R}, \quad v \mapsto \mathbb{E}[\rho(v(G,Z) + Z) - \rho(Z)],$$
$$\mathcal{G} : \mathcal{H} \to \mathbb{R}, \quad v \mapsto \mathbb{E}[v(G,Z)^2]^{1/2} - \mathbb{E}[v(G,Z)G]/\sqrt{1-\delta^{-1}}.$$

Theorem (Existence)

The constrained optimization $\min_{\mathcal{G}(v)\le 0} \mathcal{L}(v)$ *admits a solution* $v_* \in \mathcal{H}/\{0\}$*, and there exists an associated lagrange multiplier* $\mu_* > 0$ *such that* v_* *solves the unconstrained optimization* $\min_{v\in\mathcal{H}} \mathcal{L}(v) + \mu_*\mathcal{G}(v)$*. Then, the positive scalar* (α_*, κ_*) *defined as*

$$\alpha_* = \mathbb{E}[v_*^2]^{1/2}/\sqrt{1-\delta^{-1}}, \quad \kappa_* = \mathbb{E}[v_*^2]^{1/2}/\mu_*$$

is a solution to the nonlinear system.

Proof of uniqueness of solution in the unregularized case

Recall that $\mathcal{L} : \mathcal{H} \to \mathbb{R}$ and $\mathcal{G} : \mathcal{H} \to \mathbb{R}$ are defined as

$$\mathcal{L} : \mathcal{H} \to \mathbb{R}, \quad v \mapsto \mathbb{E}[\rho(v(G,Z) + Z) - \rho(Z)],$$
$$\mathcal{G} : \mathcal{H} \to \mathbb{R}, \quad v \mapsto \mathbb{E}[v(G,Z)^2]^{1/2} - \mathbb{E}[v(G,Z)G]/\sqrt{1-\delta^{-1}}.$$

Theorem (Uniqueness)

If $(\alpha_*, \kappa_*) \in \mathbb{R}^2_{>0}$ is a solution to the nonlinear system, then $v_* \in \mathcal{H}$ defined as

$$v_* : (G,Z) \mapsto \text{prox}[\kappa_* \rho](\alpha_* G + Z) - Z$$

solves the constrained optimization problem $\min_{v \in \mathcal{H}:\mathcal{G}(v) \leq 0} \mathcal{L}(v)$.

- *If $v_{**} \in \mathcal{H}$ is also a minimizer of $\min_{v \in \mathcal{H}:\mathcal{G}(v) \leq 0} \mathcal{L}(v)$, then v_{**} is necessarily proportional to v_*.*
- *If $(\alpha_{**}, \kappa_{**}) \in \mathbb{R}^2_{>0}$ is another solution to the nonlinear sysystem, then we must have $\alpha_{**} = \alpha_*$ and $\kappa_{**} = \kappa_*$.*

Summary of proof

$(\alpha_*, \kappa_*) \in \mathbb{R}^2_{>0}$ solves the nonlinear system

$$(\alpha_*, \kappa_*) = \left(\frac{\mathbb{E}[v_*^2]^{1/2}}{\sqrt{1-\delta^{-1}}}, \frac{\mathbb{E}[v_*^2]^{1/2}}{\mu_*} \right)$$

for Lagrange multiplier $\mu_* > 0$

$$v_* = \mathrm{prox}[\kappa_* \rho](\alpha_* G + Z) - Z$$

$v_* \in \mathcal{H}$ solves

$$\underset{v \in \mathcal{H}}{\text{minimize}} \ \mathbb{E}[\rho(v+Z) - \rho(Z)]$$

$$\text{subject to} \ \mathbb{E}[v^2]^{1/2} - \frac{\mathbb{E}[vG]}{\sqrt{1-\delta^{-1}}} \leq 0$$

where $\mathcal{H} = \{ v : \mathbb{R}^2 \to \mathbb{R}, \ \mathbb{E}[v(G,Z)^2] < +\infty \}$

When does the perfect recovery $\hat{\boldsymbol{x}} = \boldsymbol{x}_0$ happen?

- Assume that ρ is convex and Lipschitz with $\{0\} = \arg\min_x \rho(x)$. Let $\hat{\boldsymbol{x}}$ be the unregularized M-estimator

$$\hat{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^p} \sum_{i=1}^{n} \rho(y_i - \boldsymbol{e}_i^\top \boldsymbol{A}\boldsymbol{x})$$

  and assume $n > p$.

- By the KTT condition and $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{z}$, perfect recovery $(\hat{\boldsymbol{x}} = \boldsymbol{x}_0)$ holds iff

$$\boldsymbol{A}^\top \partial \boldsymbol{\rho}(\boldsymbol{z}) \ni \boldsymbol{0}_p,$$

  where $\boldsymbol{z} \in \mathbb{R}^n$ is the noise vector and $\partial \boldsymbol{\rho}(\boldsymbol{z}) = \bigtimes_{i=1}^{n} \partial \rho(z_i) \subset \mathbb{R}^n$.

- Since $\boldsymbol{z} = (z_1, \ldots, z_n) \overset{\text{iid}}{\sim} Z$, if $\mathbb{P}(Z = 0) = 1$ then $\partial \boldsymbol{\rho}(\boldsymbol{z}) \ni \boldsymbol{0}_p$ also holds with probability 1, so let us assume

$$\mathbb{P}(Z \neq 0) > 0$$

  to avoid this nontrivial situation.

When does the perfect recovery $\hat{\boldsymbol{x}} = \boldsymbol{x}_0$ happen?

- The iff condition $\boldsymbol{A}^\top \partial \boldsymbol{\rho}(\boldsymbol{z}) \ni \mathbf{0}_p$ for the perfect recovery is equivalent to

$$\ker(\boldsymbol{A}^\top) \cap C_{\boldsymbol{z}} \neq \emptyset \quad \text{for} \quad C_{\boldsymbol{z}} := \text{cone}(\partial \boldsymbol{\rho}(\boldsymbol{z})) = \cup_{t \geq 0} \partial(t \boldsymbol{\rho}(\boldsymbol{z}))$$

- Since $\boldsymbol{A} \in \mathbb{R}^{n \times p}$ has iid $N(0, 1/p)$ entries,

$$\ker(\boldsymbol{A}^\top) \sim \text{Unif}\{S \subset \mathbb{R}^n : \text{dim}(S) = n - p\},$$

so Amelunxen et al. [2014] implies

$$\mathrm{d}(C_{\boldsymbol{z}}) + (n - p) \gg n \Rightarrow \mathbb{P}(\text{Perfect recovery} | \boldsymbol{z}) \to 1$$
$$\mathrm{d}(C_{\boldsymbol{z}}) + (n - p) \ll n \Rightarrow \mathbb{P}(\text{Perfect recovery} | \boldsymbol{z}) \to 0$$

- Here, $\mathrm{d}(C_{\boldsymbol{z}})$ is the *statistical dimension* of the cone $C_{\boldsymbol{z}}$

$$\mathrm{d}(C_{\boldsymbol{z}}) := n - \mathbb{E}[\mathrm{dist}(\boldsymbol{g}, C_{\boldsymbol{z}})^2 | \boldsymbol{z}] \quad \text{for} \quad \boldsymbol{g} \sim N(\mathbf{0}_n, \boldsymbol{I}_n)$$

where $\mathrm{dist}(\cdot, S) = \inf_{\boldsymbol{u} \in S} \| \cdot - \boldsymbol{u} \|_2$ for any set $S \subset \mathbb{R}^n$

When does the perfect recovery $\hat{\boldsymbol{x}} = \boldsymbol{x}_0$ happen?

- Recall that for $\mathrm{d}(C_{\boldsymbol{z}}) = n - \mathbb{E}[\mathrm{dist}(\boldsymbol{g}, C_{\boldsymbol{z}})^2 | \boldsymbol{z}]$,

$$\mathrm{d}(C_{\boldsymbol{z}}) + (n-p) \gg n \Rightarrow \mathbb{P}(\text{P.R.}|\boldsymbol{z}) \to 1$$
$$\mathrm{d}(C_{\boldsymbol{z}}) + (n-p) \ll n \Rightarrow \mathbb{P}(\text{P.R.}|\boldsymbol{z}) \to 0$$

- The Gaussian Poincaré inequality and standard convex analysis yield

$$n^{-1}\,\mathrm{d}(C_{\boldsymbol{z}}) \to 1 - \inf_{t>0} \mathbb{E}[\mathrm{dist}(G, t\partial\rho(Z))^2].$$

- Combined with $n/p \to \delta$, we are left with

$$1 - \inf_{t>0} \mathbb{E}[\mathrm{dist}(G, t\partial\rho(Z))^2] - \delta^{-1} > 0 \Rightarrow \mathbb{P}(\text{perfect recovery}) \to 1,$$
$$1 - \inf_{t>0} \mathbb{E}[\mathrm{dist}(G, t\partial\rho(Z))^2] - \delta^{-1} < 0 \Rightarrow \mathbb{P}(\text{perfect recovery}) \to 0,$$

which gives the threshold $\delta_0 = 1/(1 - \inf_{t>0} \mathbb{E}[\mathrm{dist}(G, t\partial\rho(Z))^2])_+$.

Regularized case ($f \neq 0$)

Regularized case $(f \neq 0)$

- $\hat{\boldsymbol{x}}$ is the regularized M-estimator

$$\hat{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^p} \sum_{i=1}^{n} \rho(y_i - \boldsymbol{e}_i^\top \boldsymbol{A}\boldsymbol{x}) + \sum_{j=1}^{p} \lambda f(x_j)$$

and the nonlinear system of interest is

$$\begin{aligned} \alpha^2 &= \mathbb{E}\left[\left(\mathrm{prox}[\nu^{-1}\lambda f](\nu^{-1}\beta H + X) - X\right)^2\right] \\ \beta^2\kappa^2 &= \delta\,\mathbb{E}\left[\left(\alpha G + Z - \mathrm{prox}[\kappa\rho](\alpha G + Z)\right)^2\right] \\ \nu\alpha\kappa &= \delta\,\mathbb{E}\left[G \cdot \left(\alpha G + Z - \mathrm{prox}[\kappa\rho](\alpha G + Z)\right)\right] \\ \kappa\beta &= \mathbb{E}\left[H \cdot \left(\mathrm{prox}[\nu^{-1}\lambda f](\nu^{-1}\beta H + X) - X\right)\right] \end{aligned} \quad \text{for} \quad \begin{cases} G \sim \mathcal{N}(0,1), \\ H \sim \mathcal{N}(0,1), \\ X =^d x_{0j}, \\ Z =^d z_i. \end{cases},$$

with positive unknown $(\alpha, \beta, \kappa, \nu)$.

- $p^{-1}\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|^2 \to^p \alpha_*^2$ for the solution $(\alpha_*, \kappa_*, \beta_*, \nu_*)$ is proved by Thrampoulidis et al. [2018] provided that such solution uniquely exists.

Existence and uniqueness of solution in the regularized case

Let δ_0 be the positive scalar

$$\delta_0 = \frac{\inf_{t>0} \mathbb{E}[\mathrm{dist}(H, t\partial f(X))^2]}{(1 - \inf_{t>0} \mathbb{E}[\mathrm{dist}(G, t\partial \rho(Z))^2])_+} \in (0, \infty],$$

where $\mathrm{dist}(\cdot, S) = \inf_{u \in \mathbb{R}} |\cdot - u|$ for any set $S \subset \mathbb{R}$.

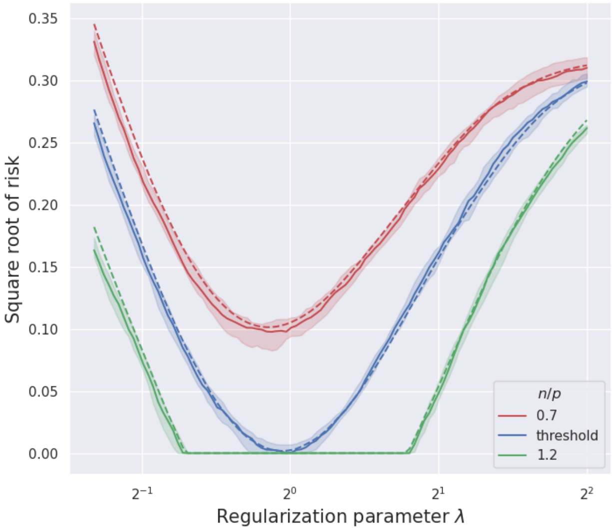
Theorem

Assume that ρ and f are Lipschitz, and satisfying mild conditions. Then,

$$\delta (= \lim \tfrac{n}{p}) > \delta_0 \Rightarrow \mathbb{P}(\textit{perfect recovery } \hat{\boldsymbol{x}} = \boldsymbol{x}_0) \to 1,$$
$$\delta < \delta_0 \Rightarrow \mathbb{P}(\textit{perfect recovery } \hat{\boldsymbol{x}} = \boldsymbol{x}_0) \to 0.$$

When $\delta < \delta_0$, the nonlinear system admits a unique positive solution $(\alpha_*, \beta_*, \kappa_*, \nu_*)$.

Numerical simulation: L1 loss and L1 penalty



$Z = 0.7 \cdot \delta_0 + 0.3 \cdot N(0,1)$, $X = 0.9 \cdot \delta_0 + 0.1 \cdot N(0,10)$,
$p = 500$, iterate=10, dash-line = theory

Proof of existence in the regularized case

Let $\mathcal{H}$ be the product of Hilbert spaces

$$\mathcal{H} := \mathcal{H}_Z \times \mathcal{H}_X, \quad \begin{cases} \mathcal{H}_Z := \{v : \mathbb{R}^2 \to \mathbb{R},\ \mathbb{E}[v(G,Z)^2] < +\infty\} \\ \mathcal{H}_X := \{w : \mathbb{R}^2 \to \mathbb{R},\ \mathbb{E}[w(H,X)^2] < +\infty\} \end{cases}$$

and define $\mathcal{L} : \mathcal{H} \to \mathbb{R}$ and $\mathcal{G} : \mathcal{H} \to \mathbb{R}$ as

$$\mathcal{L} : \mathcal{H} \to \mathbb{R}, \quad (v,w) \mapsto \delta\,\mathbb{E}[\rho(v+W) - \rho(W)] + \mathbb{E}[f(w+X) - f(X)],$$
$$\mathcal{G} : \mathcal{H} \to \mathbb{R}, \quad (v,w) \mapsto \mathcal{T}(v,w) - \delta^{-1/2}\,\mathbb{E}[Hw],$$

where $\mathcal{T}(v,w)$ is defined by

$$\mathcal{T}(v,w) := \sqrt{\left(\mathbb{E}[w^2]^{1/2} - \mathbb{E}[vG]\right)_+^2 + \mathbb{E}[\Pi_{\mathrm{G}}^{\perp}(v)^2]} \quad \text{with} \quad \Pi_{\mathrm{G}}^{\perp}(v) = v - \mathbb{E}[vG]G.$$

Proof of existence in the regularized case

$$\mathcal{L} : \mathcal{H} \to \mathbb{R}, \quad (v, w) \mapsto \delta\, \mathbb{E}[\rho(v + W) - \rho(W)] + \mathbb{E}[f(w + X) - f(X)],$$
$$\mathcal{G} : \mathcal{H} \to \mathbb{R}, \quad (v, w) \mapsto \mathcal{T}(v, w) - \delta^{-1/2}\, \mathbb{E}[Hw].$$

Theorem (Existence)

The constrained optimization $\min_{\mathcal{G}(v,w)\le 0} \mathcal{L}(v, w)$ *admits a nonzero solution* $(v_*, w_*) \in \mathcal{H}$*, and there exists an associated lagrange multiplier* $\mu_* > 0$ *such that* (v_*, w_*) *solves the unconstrained optimization* $\min_{(v,w)\in\mathcal{H}} \mathcal{L}(v, w) + \mu_* \mathcal{G}(v, w)$*. Then, the positive scalar* $(\alpha_*, \kappa_*, \beta_*, \nu_*)$ *defined as*

$$\alpha_* = \mathbb{E}[w_*^2]^{1/2}, \quad \beta_* = \frac{\mu_*}{\sqrt{\delta}}, \quad \kappa_* = \frac{\delta}{\mu_*}\mathcal{T}(v_*, w_*), \quad \nu_* = \mu_* \frac{1 - \mathbb{E}[w_*^2]^{-1/2}\, \mathbb{E}[v_* G]}{\mathcal{T}(v_*, w_*)}$$

provide a solution to the nonlinear system.

Proof of uniqueness in the regularized case

Theorem (Uniqueness)

If $(\alpha_*, \beta_*, \kappa_*, \nu_*) \in \mathbb{R}^4_{>0}$ *is a solution to the nonlinear system, then* $(v_*, w_*) \in \mathcal{H}$ *defined as*

$$v_* : (G, Z) \mapsto \mathrm{prox}[\kappa_* \rho](\alpha_* G + Z) - Z$$
$$w_* : (H, X) \mapsto \mathrm{prox}[\nu_*^{-1} f](\nu_*^{-1} \beta_* H + X) - X$$

solves the constrained optimization problem $\min_{(v,w) \in \mathcal{H} : \mathcal{G}(v,w) \leq 0} \mathcal{L}(v, w)$.

- *If* $(v_{**}, w_{**}) \in \mathcal{H}$ *is also a minimizer of* $\min_{(v,w) \in \mathcal{H} : \mathcal{G}(v,w) \leq 0} \mathcal{L}(v, w)$, *then* (v_{**}, w_{**}) *is necessarily proportional to* (v_*, w_*).
- *If* $(\alpha_{**}, \beta_{**}, \kappa_{**}, \nu_{**}) \in \mathbb{R}^4_{>0}$ *is another solution to the nonlinear sysystem, then we must have* $(\alpha_{**}, \beta_{**}, \kappa_{**}, \nu_{**}) = (\alpha_*, \beta_*, \kappa_*, \nu_*)$.

Summary

- We showed the existence and uniqueness of the solution to the nonlinear system characterizing the asymptotic error of (regularized) M-estimator in the proportional regime.
- The condition under which we derive this result is on the side of phase transition where perfect recovery is impossible.
- In the proof, we construct a "dual" convex optimization problem on a Hilbert space, which gives an explicit solution to the nonlinear system of interest.
- This proof technique can be applied to other settings (e.g. single index model)

Reference I

D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: Phase transitions in convex programs with random data. *Information and Inference: A Journal of the IMA*, 3(3):224–294, 2014.

D. Donoho and A. Montanari. High dimensional robust m-estimation: Asymptotic variance via approximate message passing. *Probability Theory and Related Fields*, 166:935–969, 2016.

N. El Karoui, D. Bean, P. J. Bickel, C. Lim, and B. Yu. On robust regression with high-dimensional predictors. *Proceedings of the National Academy of Sciences*, 110(36):14557–14562, 2013.

Q. Han and H. Ren. Gaussian random projections of convex cones: approximate kinematic formulae and applications. *arXiv preprint arXiv:2212.05545*, 2022.

L. Miolane and A. Montanari. The distribution of the lasso: Uniform control over sparse balls and adaptive parameter tuning. *The Annals of Statistics*, 49(4):2313–2335, 2021.

C. Thrampoulidis, E. Abbasi, and B. Hassibi. Precise error analysis of regularized m-estimators in high dimensions. *IEEE Transactions on Information Theory*, 64(8):5592–5628, 2018.

Derive potential $\mathcal{L}$ and constraint $\mathcal{G}$ from CGMT

Recall

$$\underset{(v,w)\in\mathcal{H}}{\text{minimize}}\ \mathcal{L}(v,w) \quad \text{subject to} \quad \mathcal{G}(v,w)\le 0,$$

where

$$\mathcal{L}(v,w) = \delta\,\mathbb{E}[\rho(v+W)-\rho(W)] + \mathbb{E}[f(w+X)-f(X)],$$
$$\mathcal{G}(v,w) = \sqrt{(\|w\| - \mathbb{E}[vG])_+^2 + \|\,\Pi_{\mathrm{G}}^{\perp}(v)\|^2} - \delta^{-1/2}\,\mathbb{E}[Hw].$$

We derive $\mathcal{L}$ and $\mathcal{G}$ from Convex Gaussian Min-Max Theorem (CGMT) (cf. Thrampoulidis et al. [2018]).

Derive potential $\mathcal{L}$ and constraint $\mathcal{G}$ from CGMT

By the change of variables $\boldsymbol{x} \mapsto \boldsymbol{w} = (\boldsymbol{x} - \boldsymbol{x}_0)/\sqrt{p}$, we have

$$\hat{\boldsymbol{x}} = \sqrt{p}\hat{\boldsymbol{w}} + \boldsymbol{x}_0 \quad \text{so that} \quad p^{-1}\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|^2 = \|\hat{\boldsymbol{w}}\|^2,$$

where $\hat{\boldsymbol{w}}$ solves

$$\min_{\boldsymbol{w}\in\mathbb{R}^p} \frac{1}{p}\sum_{i=1}^{n} \rho(z_i - \boldsymbol{e}_i^\top \sqrt{p}\boldsymbol{A}\boldsymbol{w}) + \frac{1}{p}\sum_{j=1}^{p} f(x_{0j} + \sqrt{p}w_j).$$

Introducing new variable $\sqrt{n}\boldsymbol{v} = -\sqrt{p}\boldsymbol{A}\boldsymbol{w}$, letting $\bar{\rho}(\boldsymbol{v}) := \frac{1}{n}\sum_{i=1}^{n} \rho(\sqrt{n}v_i)$ and $\bar{f}(\boldsymbol{w}) := \frac{1}{p}\sum_{j=1}^{p} f(\sqrt{p}w_j)$, the last display can be written as

$$\min_{\boldsymbol{w}\in\mathbb{R}^p, \boldsymbol{v}\in\mathbb{R}^n} \max_{\boldsymbol{u}\in\mathbb{R}^n} -\boldsymbol{u}^\top\boldsymbol{v} + \frac{1}{\sqrt{n}}\boldsymbol{u}^\top(\sqrt{p}\boldsymbol{A})(-\boldsymbol{w}) + \delta\bar{\rho}\Big(\frac{\boldsymbol{z}}{\sqrt{n}} + \boldsymbol{v}\Big) + \bar{f}\Big(\frac{\boldsymbol{x}_0}{\sqrt{p}} + \boldsymbol{w}\Big).$$

Derive potential $\mathcal{L}$ and constraint $\mathcal{G}$ from CGMT

$$\min_{\boldsymbol{w}\in\mathbb{R}^p,\boldsymbol{v}\in\mathbb{R}^n}\max_{\boldsymbol{u}\in\mathbb{R}^n} -\boldsymbol{u}^\top\boldsymbol{v}+\frac{1}{\sqrt{n}}\boldsymbol{u}^\top(\sqrt{p}\boldsymbol{A})(-\boldsymbol{w})+\delta\bar{\rho}(\frac{\boldsymbol{z}}{\sqrt{n}}+\boldsymbol{v})+\bar{f}(\frac{\boldsymbol{x}_0}{\sqrt{p}}+\boldsymbol{w}).$$

Noting $\boldsymbol{A}$ has iid $N(0,1/p)$ entries, using CGMT, the min-max problem is approximately equal to

$$\min_{\boldsymbol{w}\in\mathbb{R}^p,\boldsymbol{v}\in\mathbb{R}^n}\max_{\boldsymbol{u}\in\mathbb{R}^n} -\boldsymbol{u}^\top\boldsymbol{v}+\frac{\|\boldsymbol{u}\|(-\boldsymbol{w})^\top\boldsymbol{h}+\|\boldsymbol{w}\|\boldsymbol{u}^\top\boldsymbol{g}}{\sqrt{n}}+\delta\bar{\rho}(\frac{\boldsymbol{z}}{\sqrt{n}}+\boldsymbol{v})+\bar{f}(\frac{\boldsymbol{x}_0}{\sqrt{p}}+\boldsymbol{w})$$

for independent $\boldsymbol{h}\sim\mathcal{N}(\boldsymbol{0}_p,\boldsymbol{I}_p)$ and $\boldsymbol{g}\sim\mathcal{N}(\boldsymbol{0}_n,\boldsymbol{I}_n)$.
By simple algebra, the above display is reduced to

$$\min_{\boldsymbol{w}\in\mathbb{R}^p,\boldsymbol{v}\in\mathbb{R}^n}\delta\bar{\rho}(\frac{\boldsymbol{z}}{\sqrt{n}}+\boldsymbol{v})+\bar{f}(\frac{\boldsymbol{x}_0}{\sqrt{p}}+\boldsymbol{w})\quad\text{s.t.}\quad \|-\boldsymbol{v}+\|\boldsymbol{w}\|\frac{\boldsymbol{g}}{\sqrt{n}}\|\le\delta^{-1/2}\boldsymbol{w}^\top\frac{\boldsymbol{h}}{\sqrt{p}}$$

Derive potential $\mathcal{L}$ and constraint $\mathcal{G}$ from CGMT

$$\min_{\boldsymbol{w}\in\mathbb{R}^p,\boldsymbol{v}\in\mathbb{R}^n} \delta\bar{\rho}(\frac{\boldsymbol{z}}{\sqrt{n}}+\boldsymbol{v})+\bar{f}(\frac{\boldsymbol{x}_0}{\sqrt{p}}+\boldsymbol{w}) \quad \text{s.t.} \quad \|-\boldsymbol{v}+\|\boldsymbol{w}\|\frac{\boldsymbol{g}}{\sqrt{n}}\| \le \delta^{-1/2}\boldsymbol{w}^\top\frac{\boldsymbol{h}}{\sqrt{p}}$$

Replacing $(\frac{\boldsymbol{z}}{\sqrt{n}},\frac{\boldsymbol{x}_0}{\sqrt{p}},\frac{\boldsymbol{g}}{\sqrt{n}},\frac{\boldsymbol{h}}{\sqrt{p}})$ by (Z,X,G,H) and $(\bar{\rho},\bar{f})$ by (ρ,f), we are left with

$$\min_{v,w}\mathbb{E}[\delta\rho(v+Z)+f(w+X)] \quad \text{subject to} \quad \left\|-v+\|w\|G\right\| \le \delta^{-1/2}\,\mathbb{E}[Hw].$$

Here the constraint is equivalent to

$$\begin{aligned}
&\Leftrightarrow \sqrt{\|v\|^2+\|w\|^2-2\|w\|\,\mathbb{E}[vG]} \le \delta^{-1/2}\,\mathbb{E}[Hw],\\
&\Leftrightarrow \sqrt{(\|w\|-\mathbb{E}[vG])^2+\|v\|^2-\mathbb{E}[vG]^2} \le \delta^{-1/2}\,\mathbb{E}[Hw],\\
&\Leftrightarrow \sqrt{(\|w\|-\mathbb{E}[vG])^2+\|\,\Pi_{\mathrm{G}}^{\perp}(v)\|^2} \le \delta^{-1/2}\,\mathbb{E}[Hw].
\end{aligned}$$

Taking the positive part of $(\|w\|-\mathbb{E}[vG])^2$, we obtain the constraint $\mathcal{G}$.