Empirical risk minimization in linear regression and phase recovery

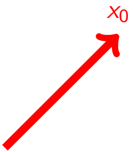
Guillaume Lecué

CNRS, centre de mathématiques appliquées, Ecole Polytechnique.

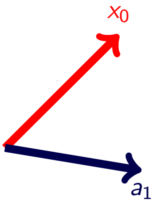
12th November 2013 - Göttingen

joint works with Shahar Mendelson

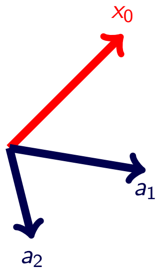
Two frameworks : linear regression and phase recovery



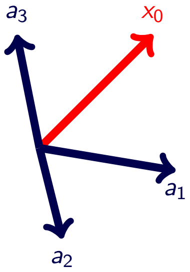
Two frameworks : linear regression and phase recovery



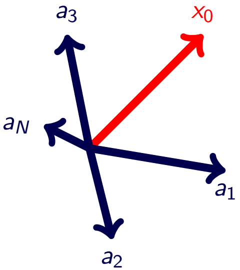
Two frameworks : linear regression and phase recovery



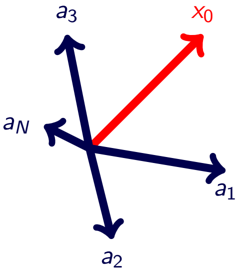
Two frameworks : linear regression and phase recovery



Two frameworks : linear regression and phase recovery

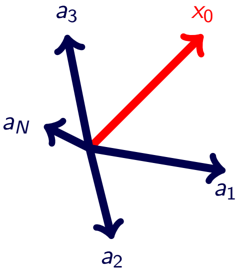


Two frameworks : linear regression and phase recovery



$T \subset \mathbb{R}^d$ and $x_0 \in T$

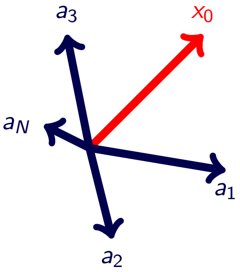
Two frameworks : linear regression and phase recovery



$T \subset \mathbb{R}^d$ and $x_0 \in T$

1) Linear regression :

Two frameworks : linear regression and phase recovery

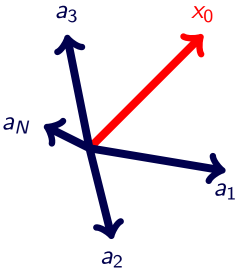


$T \subset \mathbb{R}^d$ and $x_0 \in T$

1) Linear regression :

$$\langle a_i, x_0 \rangle$$

Two frameworks : linear regression and phase recovery

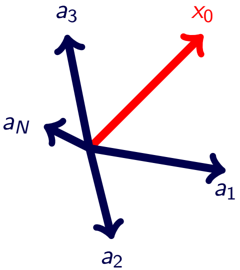


$T \subset \mathbb{R}^d$ and $x_0 \in T$

1) Linear regression :

$$\langle a_i, x_0 \rangle + \sigma g_i$$

Two frameworks : linear regression and phase recovery

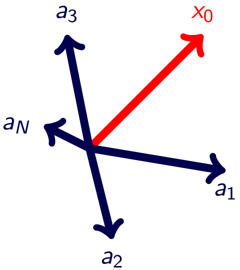


$T \subset \mathbb{R}^d$ and $x_0 \in T$

1) Linear regression :

$$y_i = \langle a_i, x_0 \rangle + \sigma g_i$$

Two frameworks : linear regression and phase recovery



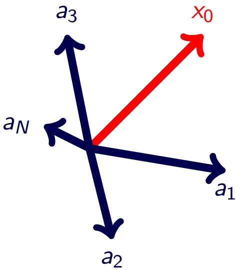
$T \subset \mathbb{R}^d$ and $x_0 \in T$

1) Linear regression :

$$y_i = \langle a_i, x_0 \rangle + \sigma g_i$$

aim : estimate x_0 from $(a_i, y_i)_{i=1}^N$

Two frameworks : linear regression and phase recovery



$T \subset \mathbb{R}^d$ and $x_0 \in T$

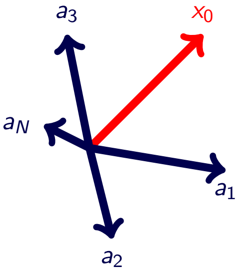
1) Linear regression :

$$y_i = \langle a_i, x_0 \rangle + \sigma g_i$$

aim : estimate x_0 from $(a_i, y_i)_{i=1}^N$

2) Phase recovery :

Two frameworks : linear regression and phase recovery



$T \subset \mathbb{R}^d$ and $x_0 \in T$

1) Linear regression :

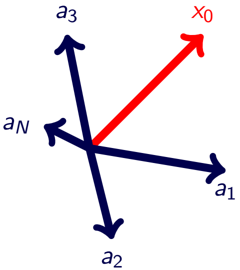
$$y_i = \langle a_i, x_0 \rangle + \sigma g_i$$

aim : estimate x_0 from $(a_i, y_i)_{i=1}^N$

2) Phase recovery :

$$\langle a_i, x_0 \rangle^2$$

Two frameworks : linear regression and phase recovery



$T \subset \mathbb{R}^d$ and $x_0 \in T$

1) Linear regression :

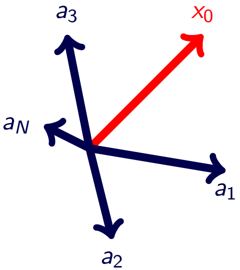
$$y_i = \langle a_i, x_0 \rangle + \sigma g_i$$

aim : estimate x_0 from $(a_i, y_i)_{i=1}^N$

2) Phase recovery :

$$\langle a_i, x_0 \rangle^2 + \sigma g_i$$

Two frameworks : linear regression and phase recovery



$T \subset \mathbb{R}^d$ and $x_0 \in T$

1) Linear regression :

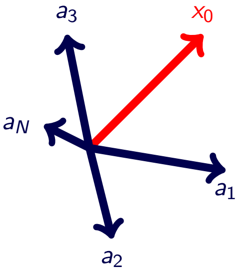
$$y_i = \langle a_i, x_0 \rangle + \sigma g_i$$

aim : estimate x_0 from $(a_i, y_i)_{i=1}^N$

2) Phase recovery :

$$y_i = \langle a_i, x_0 \rangle^2 + \sigma g_i$$

Two frameworks : linear regression and phase recovery



$T \subset \mathbb{R}^d$ and $x_0 \in T$

1) Linear regression :

$$y_i = \langle a_i, x_0 \rangle + \sigma g_i$$

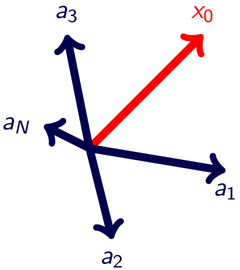
aim : estimate x_0 from $(a_i, y_i)_{i=1}^N$

2) Phase recovery :

$$y_i = \langle a_i, x_0 \rangle^2 + \sigma g_i$$

aim : estimate x_0 or $-x_0$

Two frameworks : linear regression and phase recovery



$T \subset \mathbb{R}^d$ and $x_0 \in T$

1) Linear regression :

$$y_i = \langle a_i, x_0 \rangle + \sigma g_i$$

aim : estimate x_0 from $(a_i, y_i)_{i=1}^N$

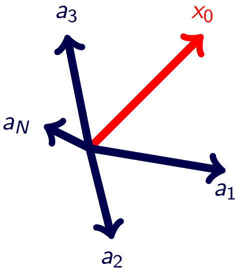
2) Phase recovery :

$$y_i = \langle a_i, x_0 \rangle^2 + \sigma g_i$$

aim : estimate x_0 or $-x_0$

- the noise g_i are independent Gaussian (noisy case $\sigma > 0$ - noise free case $\sigma = 0$)

Two frameworks : linear regression and phase recovery



$T \subset \mathbb{R}^d$ and $x_0 \in T$

1) Linear regression :

$$y_i = \langle a_i, x_0 \rangle + \sigma g_i$$

aim : estimate x_0 from $(a_i, y_i)_{i=1}^N$

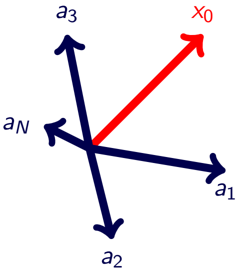
2) Phase recovery :

$$y_i = \langle a_i, x_0 \rangle^2 + \sigma g_i$$

aim : estimate x_0 or $-x_0$

- the noise g_i are independent Gaussian (noisy case $\sigma > 0$ - noise free case $\sigma = 0$)
- the measurement vectors are isotropic : $\mathbb{E}\langle a_i, x \rangle^2 = \|x\|_2^2$

Two frameworks : linear regression and phase recovery



$T \subset \mathbb{R}^d$ and $x_0 \in T$

1) Linear regression :

$$y_i = \langle a_i, x_0 \rangle + \sigma g_i$$

aim : estimate x_0 from $(a_i, y_i)_{i=1}^N$

2) Phase recovery :

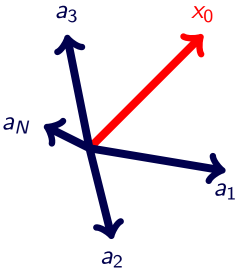
$$y_i = \langle a_i, x_0 \rangle^2 + \sigma g_i$$

aim : estimate x_0 or $-x_0$

- the noise g_i are independent Gaussian (noisy case $\sigma > 0$ - noise free case $\sigma = 0$)
- the measurement vectors are isotropic : $\mathbb{E}\langle a_i, x \rangle^2 = \|x\|_2^2$
- the measurement vectors are L-subgaussian :

$$\mathbb{P}[|\langle a_i, x \rangle| \geq tL\|x\|_2] \leq \exp(-t^2/2), \forall t > 0$$

Two frameworks : linear regression and phase recovery



$T \subset \mathbb{R}^d$ and $x_0 \in T$

1) Linear regression :

$$y_i = \langle a_i, x_0 \rangle + \sigma g_i$$

aim : estimate x_0 from $(a_i, y_i)_{i=1}^N$

2) Phase recovery :

$$y_i = \langle a_i, x_0 \rangle^2 + \sigma g_i$$

aim : estimate x_0 or $-x_0$

- the noise g_i are independent Gaussian (noisy case $\sigma > 0$ - noise free case $\sigma = 0$)
- the measurement vectors are isotropic : $\mathbb{E}\langle a_i, x \rangle^2 = \|x\|_2^2$
- the measurement vectors are L-subgaussian :

$$\mathbb{P}[|\langle a_i, x \rangle| \geq tL\|x\|_2] \leq \exp(-t^2/2), \forall t > 0$$

ex. : Gaussian measurements, Rademacher measurements.

Linear regression

Empirical risk minimization in Linear regression

- **Data :** $y_i = \left\langle a_i, x_0 \right\rangle + \sigma g_i, i = 1, \ldots, N$

Empirical risk minimization in Linear regression

- **Data :** $y_i = \langle a_i, x_0 \rangle + \sigma g_i, i = 1, \ldots, N$
- **Model :** $x_0 \in T \subset \mathbb{R}^d$

Empirical risk minimization in Linear regression

- **Data :** $y_i = \langle a_i, x_0 \rangle + \sigma g_i, i = 1, \ldots, N$
- **Model :** $x_0 \in T \subset \mathbb{R}^d$
- **Aim :** We want to construct $\hat{x}$ such that with high probability (w.h.p.) :

$$\|\hat{x} - x_0\|_2 \leq rate.$$

Empirical risk minimization in Linear regression

- **Data :** $y_i = \langle a_i, x_0 \rangle + \sigma g_i, i = 1, \ldots, N$
- **Model :** $x_0 \in T \subset \mathbb{R}^d$
- **Aim :** We want to construct $\hat{x}$ such that with high probability (w.h.p.) :

$$\|\hat{x} - x_0\|_2 \leq rate.$$

- **Estimator :** Empirical risk minimization

$$\hat{x} \in \underset{x \in T}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \big(y_i - \langle a_i, x \rangle\big)^2$$

Empirical risk minimization in Linear regression

- **Data :** $y_i = \langle a_i, x_0 \rangle + \sigma g_i, i = 1, \ldots, N$
- **Model :** $x_0 \in T \subset \mathbb{R}^d$
- **Aim :** We want to construct $\hat{x}$ such that with high probability (w.h.p.) :

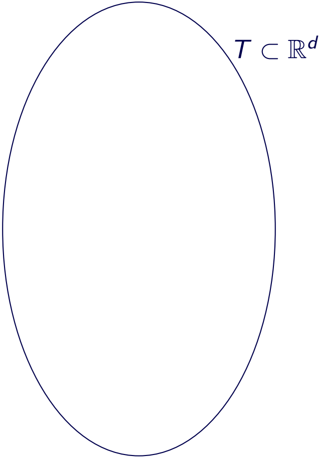
$$\|\hat{x} - x_0\|_2 \leq rate.$$

- **Estimator :** Empirical risk minimization

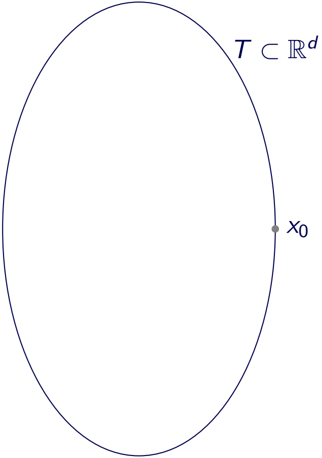
$$\hat{x} \in \operatorname*{argmin}_{x \in T} \frac{1}{N} \sum_{i=1}^{N} \big(y_i - \langle a_i, x \rangle\big)^2$$

Ordinary least square estimator - Maximum likelihood estimator

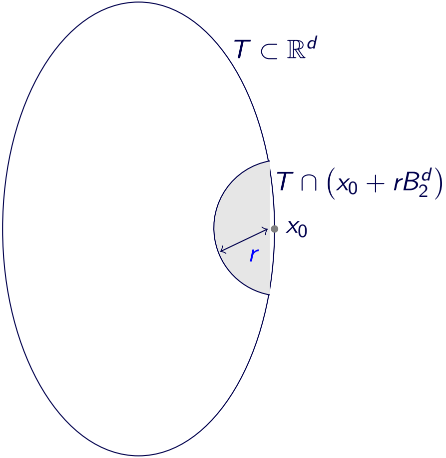
Where is localized ERM ?



Where is localized ERM ?

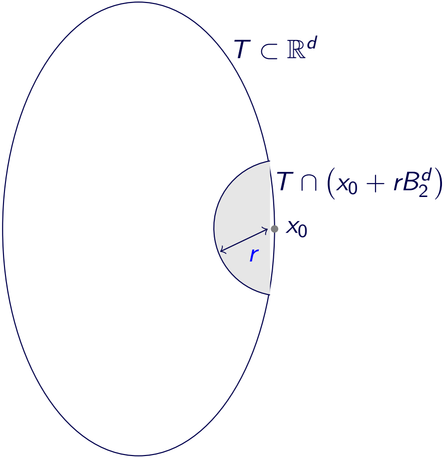


Where is localized ERM ?

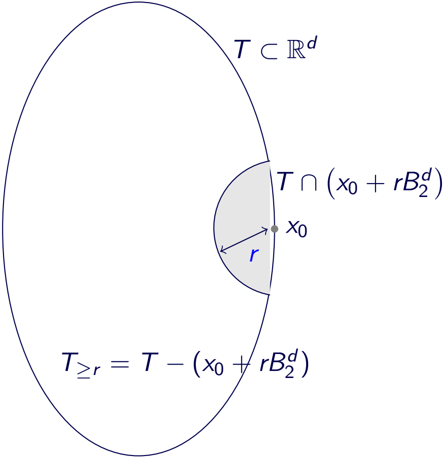


Where is localized ERM ?

$\text{bias} = r$

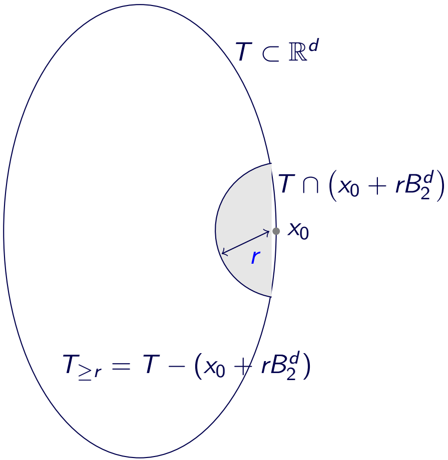


Where is localized ERM ?



bias $= r$
variance $=$ complexity of $T_{\geq r}$

Where is localized ERM ?



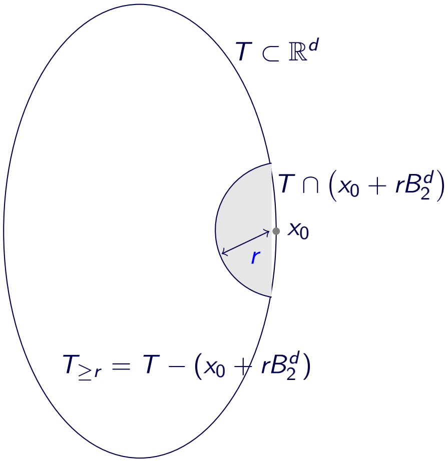
bias $= r$
variance $=$ complexity of $T_{\geq r}$
bias/variance trade-off :

$$\|\hat{x} - x_0\|_2 \sim rate_N^*$$

where

$$rate_N^* = \inf(r : \mathrm{comp}(T_{\geq r}) \leq r)$$

Where is localized ERM ?



bias $= r$
variance $=$ complexity of $T_{\geq r}$
bias/variance trade-off :

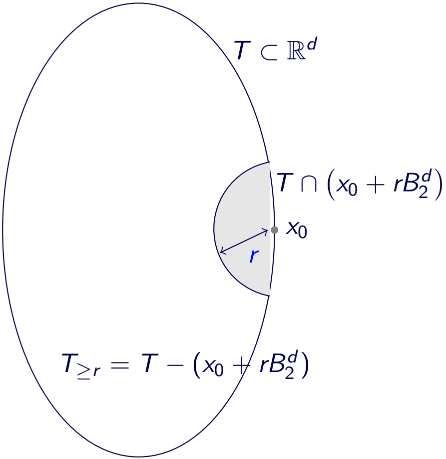
$$\|\hat{x} - x_0\|_2 \sim rate_N^*$$

where

$$rate_N^* = \inf(r : \mathrm{comp}(T_{\geq r}) \leq r)$$

Measure of complexity of $T_{\geq r}$?

Where is localized ERM ?



bias $= r$
variance $=$ complexity of $T_{\geq r}$
bias/variance trade-off :

$$\|\hat{x} - x_0\|_2 \sim rate_N^*$$

where

$$rate_N^* = \inf(r : \mathrm{comp}(T_{\geq r}) \leq r)$$

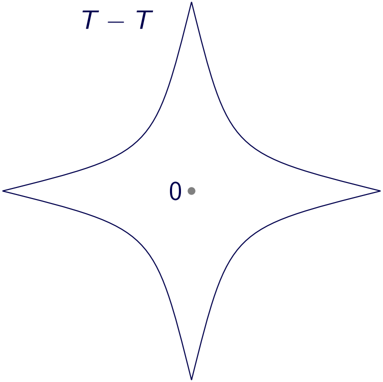
Measure of complexity of $T_{\geq r}$?
$\ell(V) = \mathbb{E}\sup_{v\in V}\left|\sum_{j=1}^d g_j v_j\right|$: Gaussian mean width

Regularity on the complexity structure : the star-shaped assumption

$T - T$ is **star-shaped in** $0 : \forall u, v \in T, \quad [u - v, 0] \subset T - T$

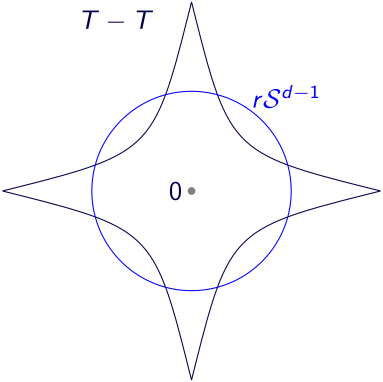
Regularity on the complexity structure : the star-shaped assumption

$T - T$ is **star-shaped in** 0 : $\forall u, v \in T, \quad [u - v, 0] \subset T - T$



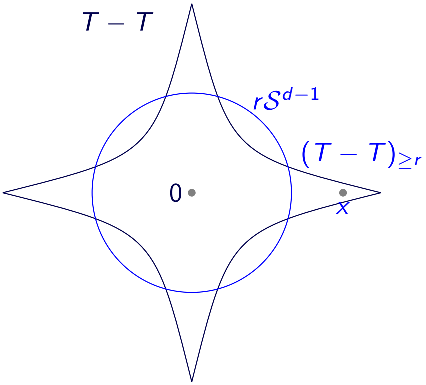
Regularity on the complexity structure : the star-shaped assumption

$T - T$ is **star-shaped in** 0 : $\forall u, v \in T, \quad [u - v, 0] \subset T - T$



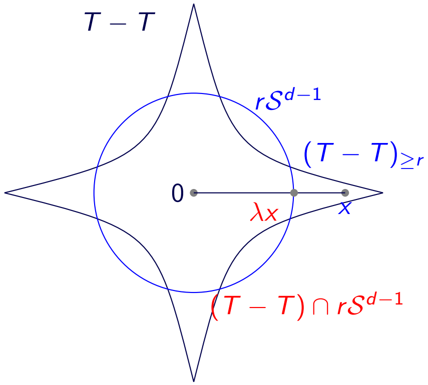
Regularity on the complexity structure : the star-shaped assumption

$T - T$ is **star-shaped in** 0 : $\forall u, v \in T, \quad [u - v, 0] \subset T - T$



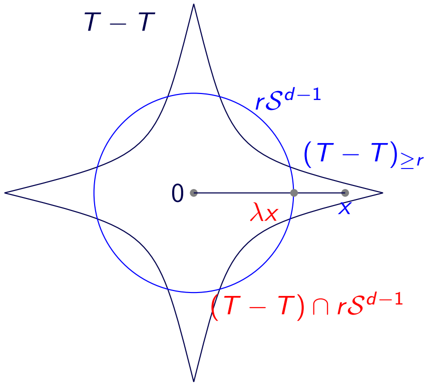
Regularity on the complexity structure : the star-shaped assumption

$T - T$ is **star-shaped in** 0 : $\forall u, v \in T, \quad [u - v, 0] \subset T - T$



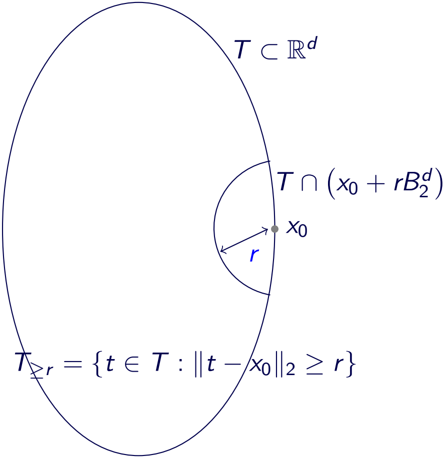
Regularity on the complexity structure : the star-shaped assumption

$T - T$ is **star-shaped in** 0 : $\forall u, v \in T, \quad [u - v, 0] \subset T - T$

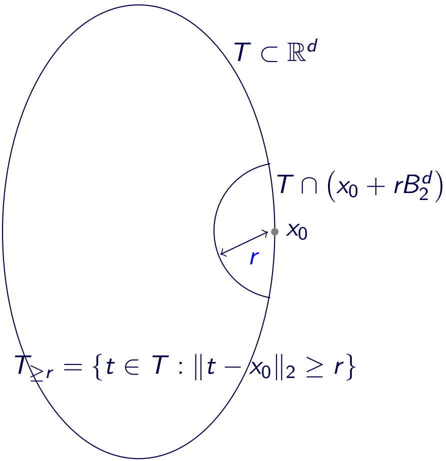


Complexity of localized sets : $(T - T) \cap r\mathcal{S}^{d-1}$
Other ways to study the complexity of $(T - T) \cap r\mathcal{S}^{d-1}$ via "peeling" cf. S. van de Geer, Cambridge University Press

Isomorphic method (Bartlett and Mendelson, PTRF06)

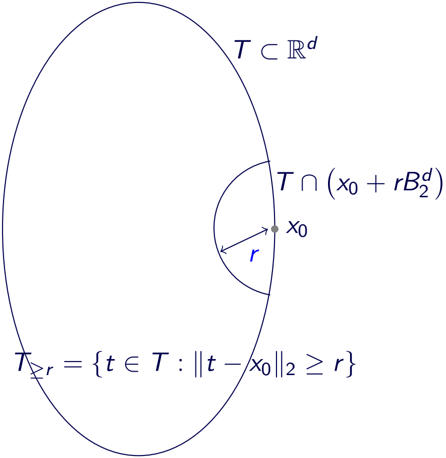


Isomorphic method (Bartlett and Mendelson, PTRF06)



ERM : $\hat{x} \in \mathrm{argmin}_{x \in T} P_N \ell_x$,
loss function :
$\ell_x(a, y) = \big(y - \langle a, x \rangle\big)^2$

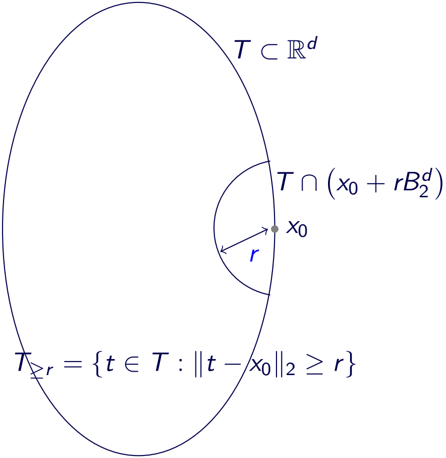
Isomorphic method (Bartlett and Mendelson, PTRF06)



ERM : $\hat{x} \in \mathrm{argmin}_{x \in T} P_N \ell_x$,
loss function :
$\ell_x(a, y) = \big(y - \langle a, x \rangle\big)^2$

excess loss function :
$\mathcal{L}_x = \ell_x - \ell_{x_0}$

Isomorphic method (Bartlett and Mendelson, PTRF06)

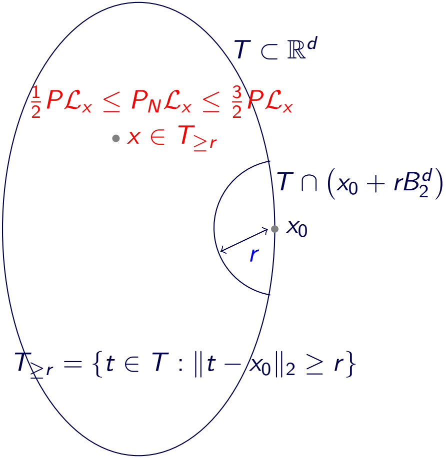


ERM : $\hat{x} \in \operatorname{argmin}_{x \in T} P_N \ell_x$,
loss function :
$\ell_x(a, y) = \big(y - \langle a, x \rangle\big)^2$

excess loss function :
$\mathcal{L}_x = \ell_x - \ell_{x_0}$

$P_N \mathcal{L}_{\hat{x}} \leq P_N \mathcal{L}_{x_0} = 0$

Isomorphic method (Bartlett and Mendelson, PTRF06)



ERM : $\hat{x} \in \text{argmin}_{x \in T} P_N \ell_x$,

loss function :

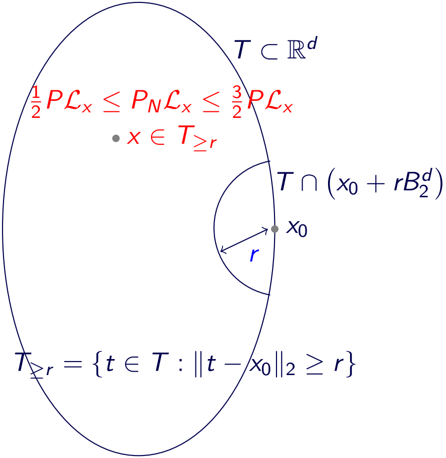
$\ell_x(a, y) = \big(y - \langle a, x \rangle\big)^2$

excess loss function :

$\mathcal{L}_x = \ell_x - \ell_{x_0}$

$P_N \mathcal{L}_{\hat{x}} \leq P_N \mathcal{L}_{x_0} = 0$

Isomorphic method (Bartlett and Mendelson, PTRF06)



ERM : $\hat{x} \in \operatorname{argmin}_{x \in T} P_N \ell_x$,
loss function :
$\ell_x(a, y) = \big(y - \langle a, x\rangle\big)^2$

excess loss function :
$\mathcal{L}_x = \ell_x - \ell_{x_0}$

$P_N \mathcal{L}_{\hat{x}} \le P_N \mathcal{L}_{x_0} = 0$

Isomorphic property over $T_{\ge r}$ implies that $\hat{x} \notin T_{\ge r}$

$$\Longrightarrow \|\hat{x} - x_0\|_2 \le r$$

Isomorphic property over $T_{\geq r}$

We want : w.h.p. for any $x \in T_{\geq r}$,

$$\left| P_N \mathcal{L}_x - P\mathcal{L}_x \right| \leq \frac{1}{2} P\mathcal{L}_x$$

Isomorphic property over $T_{\geq r}$

We want : w.h.p. for any $x \in T_{\geq r}$,

$$\left| P_N \mathcal{L}_x - P\mathcal{L}_x \right| \leq \frac{1}{2} P\mathcal{L}_x$$

Study of the ratio process (cf. Koltchinskii, Saint-Flour) :

$$\sup_{x \in T_{\geq r}} \left| 1 - P_N \Big(\frac{\mathcal{L}_x}{P\mathcal{L}_x} \Big) \right| \leq \frac{1}{2}$$

Isomorphic property over $T_{\geq r}$

We want : w.h.p. for any $x \in T_{\geq r}$,

$$\left|P_N\mathcal{L}_x - P\mathcal{L}_x\right| \leq \frac{1}{2}P\mathcal{L}_x$$

Study of the ratio process (cf. Koltchinskii, Saint-Flour) :

$$\sup_{x\in T_{\geq r}}\left|1 - P_N\left(\frac{\mathcal{L}_x}{P\mathcal{L}_x}\right)\right| \leq \frac{1}{2}$$

Here : ratio process via decomposition of the excess risk quadratic + multiplier

$$\begin{aligned}\mathcal{L}_x(a,y) &= (\ell_x - \ell_{x_0})(a,y) = (y - \langle a,x\rangle)^2 - (y - \langle a,x_0\rangle)^2\\ &= \langle a, x - x_0\rangle^2 + 2\sigma g\langle a, x - x_0\rangle\end{aligned}$$

$\Rightarrow$ study of the isomorphic structure over $x - x_0 \in (T - T) \cap r\mathcal{S}^{n-1}$

$$P\mathcal{L}_x(a,y) = P\left\langle a, x - x_0 \right\rangle^2 + 2\sigma P[g\left\langle a, x - x_0 \right\rangle] = P\left\langle a, x - x_0 \right\rangle^2 = \|x - x_0\|_2^2 = r^2.$$

$\Rightarrow$ study of the isomorphic structure over $x - x_0 \in (T - T) \cap r\mathcal{S}^{n-1}$

$$P\mathcal{L}_x(a,y) = P\big\langle a, x - x_0\big\rangle^2 + 2\sigma P[g\big\langle a, x - x_0\big\rangle] = P\big\langle a, x - x_0\big\rangle^2 = \|x - x_0\|_2^2 = r^2.$$

Estimation of x_0 using observations $(a_i, \big\langle a_i, x_0\big\rangle + \sigma g_i)_{i=1}^N$: 2 sources of statistical complexity :

$\Rightarrow$ study of the isomorphic structure over $x - x_0 \in (T - T) \cap r\mathcal{S}^{n-1}$

$$P\mathcal{L}_x(a,y) = P\big\langle a, x-x_0\big\rangle^2 + 2\sigma P[g\big\langle a, x-x_0\big\rangle] = P\big\langle a, x-x_0\big\rangle^2 = \|x-x_0\|_2^2 = r^2.$$

Estimation of x_0 using observations $(a_i, \langle a_i, x_0\rangle + \sigma g_i)_{i=1}^N$: 2 sources of statistical complexity :

- the projection $P_{\mathbb{A}} : x_0 \in \mathbb{R}^d \mapsto \big(\langle a_i, x_0\rangle\big)_{i=1}^N$ is a source of complexity.

$\Rightarrow$ study of the isomorphic structure over $x - x_0 \in (T - T) \cap r\mathcal{S}^{n-1}$

$$P\mathcal{L}_x(a,y) = P\left\langle a, x-x_0\right\rangle^2 + 2\sigma P[g\left\langle a, x-x_0\right\rangle] = P\left\langle a, x-x_0\right\rangle^2 = \|x-x_0\|_2^2 = r^2.$$

Estimation of x_0 using observations $(a_i, \left\langle a_i, x_0\right\rangle + \sigma g_i)_{i=1}^N$: 2 sources of statistical complexity :

- the projection $P_{\mathbb{A}} : x_0 \in \mathbb{R}^d \mapsto \left(\left\langle a_i, x_0\right\rangle\right)_{i=1}^N$ is a source of complexity. Main source of complexity when $\sigma \lesssim r_N^*$.

$\Rightarrow$ study of the isomorphic structure over $x - x_0 \in (T - T) \cap r\mathcal{S}^{n-1}$

$$P\mathcal{L}_x(a,y) = P\left\langle a, x - x_0\right\rangle^2 + 2\sigma P[g\left\langle a, x-x_0\right\rangle] = P\left\langle a, x-x_0\right\rangle^2 = \|x-x_0\|_2^2 = r^2.$$

Estimation of x_0 using observations $(a_i, \left\langle a_i, x_0\right\rangle + \sigma g_i)_{i=1}^N$: 2 sources of statistical complexity :

- the projection $P_{\mathbb{A}} : x_0 \in \mathbb{R}^d \mapsto \left(\left\langle a_i, x_0\right\rangle\right)_{i=1}^N$ is a source of complexity. Main source of complexity when $\sigma \lesssim r_N^*$. It is measured by the quadratic process $(P - P_N)(\left\langle \cdot, u\right\rangle^2)_{u \in (T-T)\cap r\mathcal{S}^{n-1}}$.

$\Rightarrow$ study of the isomorphic structure over $x - x_0 \in (T - T) \cap r\mathcal{S}^{n-1}$

$$P\mathcal{L}_x(a,y) = P\big\langle a, x - x_0\big\rangle^2 + 2\sigma P[g\big\langle a, x - x_0\big\rangle] = P\big\langle a, x - x_0\big\rangle^2 = \|x - x_0\|_2^2 = r^2.$$

Estimation of x_0 using observations $(a_i, \big\langle a_i, x_0\big\rangle + \sigma g_i)_{i=1}^N$: 2 sources of statistical complexity :

- the projection $P_{\mathbb{A}} : x_0 \in \mathbb{R}^d \mapsto \big(\big\langle a_i, x_0\big\rangle\big)_{i=1}^N$ is a source of complexity. Main source of complexity when $\sigma \lesssim r_N^*$. It is measured by the quadratic process $(P - P_N)(\big\langle \cdot, u\big\rangle^2)_{u \in (T-T)\cap r\mathcal{S}^{n-1}}$.
- the noise $y_i = \big\langle a_i, x_0\big\rangle + \sigma g_i$ is a source of statistical complexity.

$\Rightarrow$ study of the isomorphic structure over $x - x_0 \in (T - T) \cap r\mathcal{S}^{n-1}$

$$P\mathcal{L}_x(a,y) = P\big\langle a, x - x_0\big\rangle^2 + 2\sigma P[g\big\langle a, x - x_0\big\rangle] = P\big\langle a, x - x_0\big\rangle^2 = \|x - x_0\|_2^2 = r^2.$$

Estimation of x_0 using observations $(a_i, \big\langle a_i, x_0\big\rangle + \sigma g_i)_{i=1}^N$: 2 sources of statistical complexity :

- the projection $P_{\mathbb{A}} : x_0 \in \mathbb{R}^d \mapsto \big(\big\langle a_i, x_0\big\rangle\big)_{i=1}^N$ is a source of complexity. Main source of complexity when $\sigma \lesssim r_N^*$. It is measured by the quadratic process $(P - P_N)(\big\langle \cdot, u\big\rangle^2)_{u \in (T-T)\cap r\mathcal{S}^{n-1}}$.
- the noise $y_i = \big\langle a_i, x_0\big\rangle + \sigma g_i$ is a source of statistical complexity. Main source of complexity when $\sigma \gtrsim r_N^*$.

$\Rightarrow$ study of the isomorphic structure over $x - x_0 \in (T - T) \cap r\mathcal{S}^{n-1}$

$$P\mathcal{L}_x(a,y) = P\left\langle a, x - x_0\right\rangle^2 + 2\sigma P[g\left\langle a, x - x_0\right\rangle] = P\left\langle a, x - x_0\right\rangle^2 = \|x - x_0\|_2^2 = r^2.$$

Estimation of x_0 using observations $(a_i, \left\langle a_i, x_0\right\rangle + \sigma g_i)_{i=1}^N$: 2 sources of statistical complexity :

- the projection $P_{\mathbb{A}} : x_0 \in \mathbb{R}^d \mapsto \left(\left\langle a_i, x_0\right\rangle\right)_{i=1}^N$ is a source of complexity. Main source of complexity when $\sigma \lesssim r_N^*$. It is measured by the quadratic process $(P - P_N)(\left\langle \cdot, u\right\rangle^2)_{u \in (T-T)\cap r\mathcal{S}^{n-1}}$.
- the noise $y_i = \left\langle a_i, x_0\right\rangle + \sigma g_i$ is a source of statistical complexity. Main source of complexity when $\sigma \gtrsim r_N^*$. This complexity is measured by the multiplier process $(P - P_N)(g\left\langle a, u\right\rangle)_{u \in (T-T)\cap r\mathcal{S}^{n-1}}$.

2 empirical processes - 2 statistical complexities - 2 regimes

1. The quadratic process $((P - P_N)(\langle \cdot, u\rangle^2)_{u \in (T-T)\cap r\mathcal{S}^{n-1}}$.
   [Mendelson-Pajor-Tomczak] : w.h.p.

$$\sup_{v \in V} \left| \frac{1}{N} \sum_{i=1}^{N} \langle a_i, v\rangle^2 - \|v\|_2^2 \right| \lesssim \left(\mathrm{diam}(V, \ell_2^d) \frac{\ell(V)}{\sqrt{N}} + \frac{\ell^2(V)}{N} \right).$$

2 empirical processes - 2 statistical complexities - 2 regimes

1. The quadratic process $((P - P_N)(\langle \cdot, u\rangle^2)_{u \in (T-T)\cap r\mathcal{S}^{n-1}}$.
   [Mendelson-Pajor-Tomczak] : w.h.p.

$$\sup_{v\in V}\left|\frac{1}{N}\sum_{i=1}^{N}\langle a_i, v\rangle^2 - \|v\|_2^2\right| \lesssim \left(\mathrm{diam}(V,\ell_2^d)\frac{\ell(V)}{\sqrt{N}} + \frac{\ell^2(V)}{N}\right).$$

Measures the complexity coming from the projection via the fixed point

$$r_N^*(Q) = \inf\left(r > 0 : \ell\big((T-T)\cap r\mathcal{S}^{d-1}\big) \leq Qr\sqrt{N}\right)$$

2 empirical processes - 2 statistical complexities - 2 regimes

1. The quadratic process $((P-P_N)(\langle \cdot, u\rangle^2)_{u\in(T-T)\cap r\mathcal{S}^{n-1}}$.
   [Mendelson-Pajor-Tomczak] : w.h.p.

$$\sup_{v\in V}\left|\frac{1}{N}\sum_{i=1}^N \langle a_i, v\rangle^2 - \|v\|_2^2\right| \lesssim \left(\mathrm{diam}(V,\ell_2^d)\frac{\ell(V)}{\sqrt{N}} + \frac{\ell^2(V)}{N}\right).$$

   Measures the complexity coming from the projection via the fixed point

$$r_N^*(Q) = \inf\left(r>0 : \ell\big((T-T)\cap r\mathcal{S}^{d-1}\big) \le Qr\sqrt{N}\right)$$

2. The multiplier process $((P-P_N)(\sigma g\langle a, u\rangle))_{u\in(T-T)\cap r\mathcal{S}^{n-1}}$.
   [Mendelson] : w.h.p.

$$\sup_{v\in V}\left|\frac{1}{N}\sum_{i=1}^N \sigma g_i\langle a_i, v\rangle\right| \lesssim \sigma\frac{\ell(V)}{\sqrt{N}}.$$

2 empirical processes - 2 statistical complexities - 2 regimes

1. The quadratic process $((P-P_N)(\langle\cdot,u\rangle^2)_{u\in(T-T)\cap r\mathcal{S}^{n-1}}$.
   [Mendelson-Pajor-Tomczak] : w.h.p.

$$\sup_{v\in V}\left|\frac{1}{N}\sum_{i=1}^N \langle a_i,v\rangle^2-\|v\|_2^2\right|\lesssim\left(\mathrm{diam}(V,\ell_2^d)\frac{\ell(V)}{\sqrt{N}}+\frac{\ell^2(V)}{N}\right).$$

   Measures the complexity coming from the projection via the fixed point

$$r_N^*(Q)=\inf\left(r>0:\ell\big((T-T)\cap r\mathcal{S}^{d-1}\big)\le Qr\sqrt{N}\right)$$

2. The multiplier process $((P-P_N)(\sigma g\langle a,u\rangle))_{u\in(T-T)\cap r\mathcal{S}^{n-1}}$.
   [Mendelson] : w.h.p.

$$\sup_{v\in V}\left|\frac{1}{N}\sum_{i=1}^N \sigma g_i\langle a_i,v\rangle\right|\lesssim\sigma\frac{\ell(V)}{\sqrt{N}}.$$

   Measures the complexity coming from the noise via the fixed point :

$$s_N^*(Q)=\inf\left(r>0:\sigma\ell\big((T-T)\cap r\mathcal{S}^{d-1}\big)\le Qr^2\sqrt{N}\right)$$

Complexity of $T_{\geq r}$ and the bias/variance trade-off

$$\mathrm{comp}^2(T_{\geq r}) \lesssim \max\left(\sigma \frac{\ell(T \cap rB_2^d)}{\sqrt{N}}, r\frac{\ell(T \cap rB_2^d)}{\sqrt{N}} + \frac{\ell^2(T \cap rB_2^d)}{N}\right).$$

Complexity of $T_{\geq r}$ and the bias/variance trade-off

$$\mathrm{comp}^2(T_{\geq r}) \lesssim \max\left(\sigma \frac{\ell(T \cap rB_2^d)}{\sqrt{N}}, r\frac{\ell(T \cap rB_2^d)}{\sqrt{N}} + \frac{\ell^2(T \cap rB_2^d)}{N}\right).$$

$$rate_N^* = \inf\left(r \geq 0 : \mathrm{comp}(T_{\geq r}) \leq r\right) \sim$$

Complexity of $T_{\geq r}$ and the bias/variance trade-off

$$\mathrm{comp}^2(T_{\geq r}) \lesssim \max\left(\sigma \frac{\ell(T \cap rB_2^d)}{\sqrt{N}}, r\frac{\ell(T \cap rB_2^d)}{\sqrt{N}} + \frac{\ell^2(T \cap rB_2^d)}{N}\right).$$

$$rate_N^* = \inf\left(r \geq 0 : \mathrm{comp}(T_{\geq r}) \leq r\right) \sim \begin{cases} r_N^* & \text{if } \sigma \leq r_N^* \\ s_N^* & \text{otherwise} \end{cases}$$

Complexity of $T_{\geq r}$ and the bias/variance trade-off

$$\mathrm{comp}^2(T_{\geq r}) \lesssim \max\left(\sigma\frac{\ell(T\cap rB_2^d)}{\sqrt{N}}, r\frac{\ell(T\cap rB_2^d)}{\sqrt{N}} + \frac{\ell^2(T\cap rB_2^d)}{N}\right).$$

$$rate_N^* = \inf\left(r\geq 0 : \mathrm{comp}(T_{\geq r})\leq r\right) \sim \begin{cases} r_N^* & \text{if } \sigma\leq r_N^* \\ s_N^* & \text{otherwise}\end{cases}$$

When $T - T$ is star-shaped in 0, then with high probability :

$$\|\hat{x} - x_0\|_2 \leq rate_N^*$$

Complexity of $T_{\geq r}$ and the bias/variance trade-off

$$\mathrm{comp}^2(T_{\geq r}) \lesssim \max\left(\sigma \frac{\ell(T \cap rB_2^d)}{\sqrt{N}}, r\frac{\ell(T \cap rB_2^d)}{\sqrt{N}} + \frac{\ell^2(T \cap rB_2^d)}{N}\right).$$

$$rate_N^* = \inf\left(r \geq 0 : \mathrm{comp}(T_{\geq r}) \leq r\right) \sim \begin{cases} r_N^* & \text{if } \sigma \leq r_N^* \\ s_N^* & \text{otherwise} \end{cases}$$

When $T - T$ is star-shaped in 0, then with high probability :

$$\|\hat{x} - x_0\|_2 \leq rate_N^*$$

When $\sigma = 0$, the rate is

$$r_N^* = \inf\left(r > 0 : \ell\left((T - T) \cap r\mathcal{S}^{d-1}\right) \leq c_0 r\sqrt{N}\right)$$

which may be 0 : exact reconstruction when N is large enough.

Complexity of $T_{\geq r}$ and the bias/variance trade-off

$$\mathrm{comp}^2(T_{\geq r}) \lesssim \max\left(\sigma \frac{\ell(T \cap rB_2^d)}{\sqrt{N}}, r\frac{\ell(T \cap rB_2^d)}{\sqrt{N}} + \frac{\ell^2(T \cap rB_2^d)}{N}\right).$$

$$rate_N^* = \inf\left(r \geq 0 : \mathrm{comp}(T_{\geq r}) \leq r\right) \sim \begin{cases} r_N^* & \text{if } \sigma \leq r_N^* \\ s_N^* & \text{otherwise} \end{cases}$$

When $T - T$ is star-shaped in 0, then with high probability :

$$\|\hat{x} - x_0\|_2 \leq rate_N^*$$

When $\sigma = 0$, the rate is

$$r_N^* = \inf\left(r > 0 : \ell\big((T - T) \cap r\mathcal{S}^{d-1}\big) \leq c_0 r\sqrt{N}\right)$$

which may be 0 : exact reconstruction when N is large enough.
ex. : when T is the set of s-sparse vectors
$\ell\big((T - T) \cap r\mathcal{S}^{d-1}\big) \sim r\sqrt{s\log(ed/s)}$.

Complexity of $T_{\geq r}$ and the bias/variance trade-off

$$\mathrm{comp}^2(T_{\geq r}) \lesssim \max\left(\sigma \frac{\ell(T \cap rB_2^d)}{\sqrt{N}}, r\frac{\ell(T \cap rB_2^d)}{\sqrt{N}} + \frac{\ell^2(T \cap rB_2^d)}{N}\right).$$

$$rate_N^* = \inf\left(r \geq 0 : \mathrm{comp}(T_{\geq r}) \leq r\right) \sim \begin{cases} r_N^* & \text{if } \sigma \leq r_N^* \\ s_N^* & \text{otherwise} \end{cases}$$

When $T - T$ is star-shaped in 0, then with high probability :

$$\|\hat{x} - x_0\|_2 \leq rate_N^*$$

When $\sigma = 0$, the rate is

$$r_N^* = \inf\left(r > 0 : \ell\left((T - T) \cap r\mathcal{S}^{d-1}\right) \leq c_0 r\sqrt{N}\right)$$

which may be 0 : exact reconstruction when N is large enough.
ex. : when T is the set of s-sparse vectors
$\ell\left((T - T) \cap r\mathcal{S}^{d-1}\right) \sim r\sqrt{s\log(ed/s)}$. if $N \gtrsim s\log(ed/s)$ then $r_N^* = 0$.

Phase recovery

Empirical risk minimization in phase recovery

- **Data :** $y_i = \left\langle a_i, x_0 \right\rangle^2 + \sigma g_i, i = 1, \ldots, N$

Empirical risk minimization in phase recovery

- **Data :** $y_i = \left\langle a_i, x_0 \right\rangle^2 + \sigma g_i, i = 1, \ldots, N$
- **Model :** $x_0 \in T \subset \mathbb{R}^d$

Empirical risk minimization in phase recovery

- **Data :** $y_i = \left\langle a_i, x_0 \right\rangle^2 + \sigma g_i, i = 1, \ldots, N$
- **Model :** $x_0 \in T \subset \mathbb{R}^d$
- **Aim :** We want to construct $\hat{x}$ such that with high probability (w.h.p.) :
$$\min \left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2 \right) \leq rate.$$

Empirical risk minimization in phase recovery

- **Data :** $y_i = \left\langle a_i, x_0 \right\rangle^2 + \sigma g_i, i = 1, \ldots, N$
- **Model :** $x_0 \in T \subset \mathbb{R}^d$
- **Aim :** We want to construct $\hat{x}$ such that with high probability (w.h.p.) :

$$\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right) \leq rate.$$

- **Estimator :** Empirical risk minimization

$$\hat{x} \in \underset{x \in T}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \left\langle a_i, x \right\rangle^2\right)^2$$

Empirical risk minimization in phase recovery

- **Data :** $y_i = \langle a_i, x_0 \rangle^2 + \sigma g_i, i = 1, \dots, N$
- **Model :** $x_0 \in T \subset \mathbb{R}^d$
- **Aim :** We want to construct $\hat{x}$ such that with high probability (w.h.p.) :

$$\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right) \le rate.$$

- **Estimator :** Empirical risk minimization

$$\hat{x} \in \underset{x \in T}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \langle a_i, x \rangle^2\right)^2$$

rem. : Even when T is convex, this is not a convex optimization problem (cf. E. Candès et al. or A. d'Aspremont for linear programming algorithms in phase recovery).

Loss and excess loss functions

Loss function :

$$\ell_x(a,y) = \left(y - \left\langle a, x \right\rangle^2\right)^2$$

Loss and excess loss functions

Loss function :

$$\ell_x(a,y) = \left(y - \left\langle a, x\right\rangle^2\right)^2$$

Excess loss and risk function : $\mathcal{L}_x = \ell_x - \ell_{x_0}$,

Loss and excess loss functions

Loss function :

$$\ell_x(a, y) = \left(y - \left\langle a, x \right\rangle^2\right)^2$$

Excess loss and risk function : $\mathcal{L}_x = \ell_x - \ell_{x_0}$,

$$P\mathcal{L}_x = \mathbb{E}\left\langle a, x - x_0 \right\rangle^2 \left\langle a, x + x_0 \right\rangle^2.$$

Loss and excess loss functions

Loss function :

$$\ell_x(a,y) = \left(y - \left\langle a, x \right\rangle^2\right)^2$$

Excess loss and risk function : $\mathcal{L}_x = \ell_x - \ell_{x_0}$,

$$P\mathcal{L}_x = \mathbb{E}\left\langle a, x - x_0 \right\rangle^2 \left\langle a, x + x_0 \right\rangle^2.$$

Assumption : for all $u, v \in T$,

$$\mathbb{E}|\left\langle a, u \right\rangle \left\langle a, v \right\rangle| \geq \kappa_0 \|u\|_2 \|v\|_2$$

Loss and excess loss functions

Loss function :

$$\ell_x(a, y) = \left(y - \langle a, x\rangle^2\right)^2$$

Excess loss and risk function : $\mathcal{L}_x = \ell_x - \ell_{x_0}$,

$$P\mathcal{L}_x = \mathbb{E}\langle a, x - x_0\rangle^2 \langle a, x + x_0\rangle^2.$$

Assumption : for all $u, v \in T$,

$$\mathbb{E}|\langle a, u\rangle\langle a, v\rangle| \geq \kappa_0 \|u\|_2 \|v\|_2$$

First aim : construct $\hat{x}$ such that

$$\|\hat{x} - x_0\|_2 \|\hat{x} + x_0\|_2 \leq \mathrm{small}$$

A word on the assumption : $\mathbb{E}|\langle a, u\rangle\langle a, v\rangle| \geq \kappa_0 \|u\|_2 \|v\|_2$

Definition

A random vector a in $\mathbb{R}^d$ satisfies the small ball assumption when for all $x \in \mathbb{R}^d$ and $\epsilon > 0$,

$$\mathbb{P}\left[|\langle a, x\rangle| \leq \epsilon \|x\|_2\right] \leq c_0 \epsilon.$$

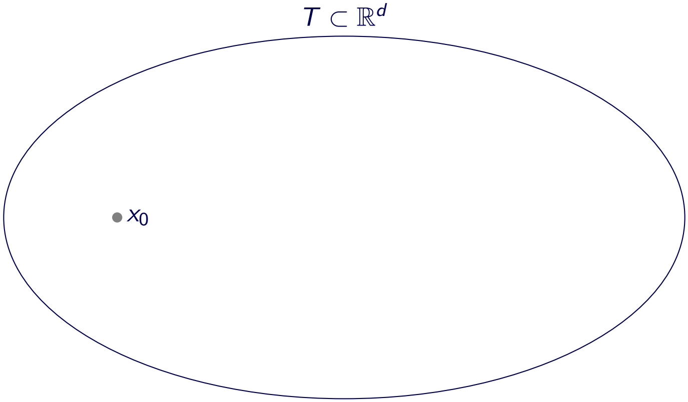
A word on the assumption : $\mathbb{E}|\langle a, u\rangle\langle a, v\rangle| \geq \kappa_0\|u\|_2\|v\|_2$

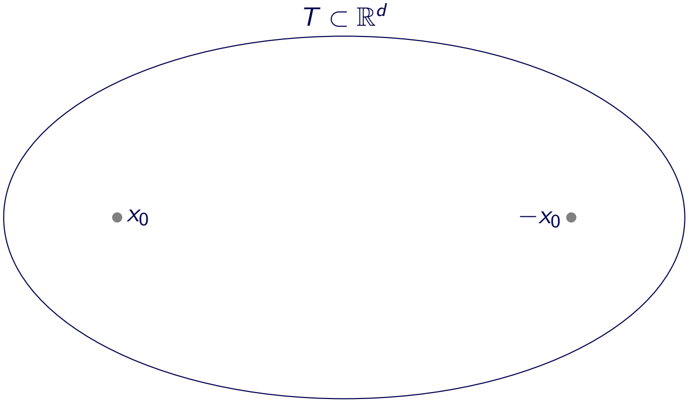
Definition

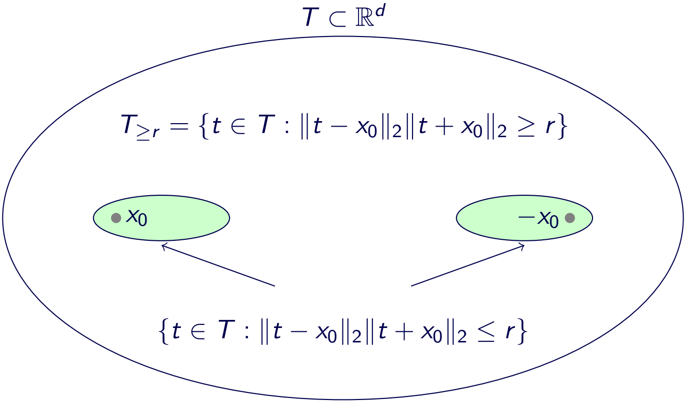
A random vector a in $\mathbb{R}^d$ satisfies the small ball assumption when for all $x \in \mathbb{R}^d$ and $\epsilon > 0$,

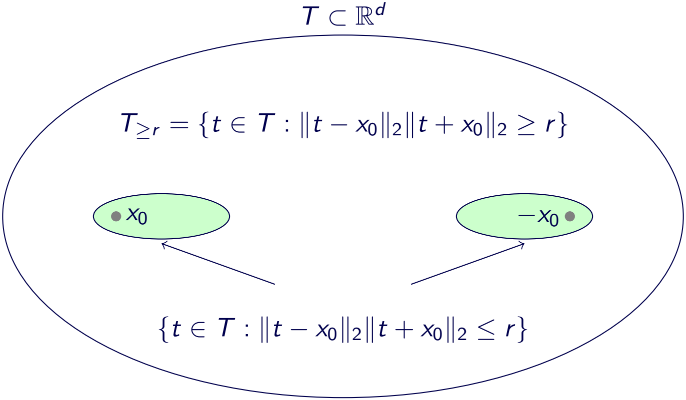
$$\mathbb{P}\big[|\langle a, x\rangle| \leq \epsilon\|x\|_2\big] \leq c_0\epsilon.$$

a satisfies the small ball assumption $\Rightarrow \mathbb{E}|\langle a, u\rangle\langle a, v\rangle| \geq \kappa_0\|u\|_2\|v\|_2$

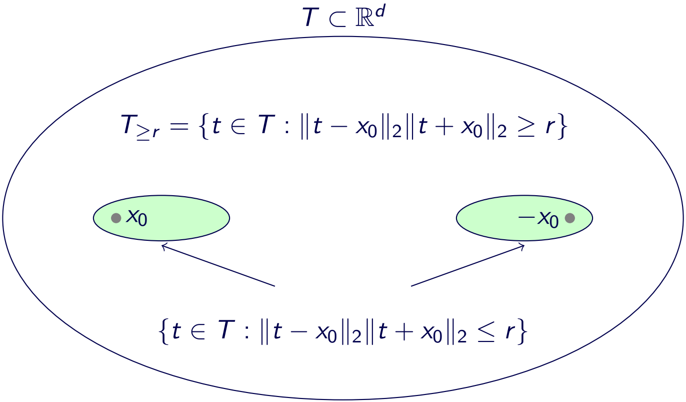

$T \subset \mathbb{R}^d$
x_0


$T \subset \mathbb{R}^d$
x_0
$-x_0$


$T \subset \mathbb{R}^d$
$T_{\geq r} = \{t \in T : \|t - x_0\|_2 \|t + x_0\|_2 \geq r\}$
x_0
$-x_0$
$\{t \in T : \|t - x_0\|_2 \|t + x_0\|_2 \leq r\}$



$$T_{\pm,r} = \left\{ \frac{t \pm x_0}{\|t \pm x_0\|_2} : t \in T, \|t - x_0\|_2 \|t + x_0\|_2 \geq r \right\}.$$



$$T_{\pm,r} = \left\{ \frac{t \pm x_0}{\|t \pm x_0\|_2} : t \in T, \|t - x_0\|_2 \|t + x_0\|_2 \geq r \right\}.$$

Measure of complexity :

$$E_r = \max\left(\ell(T_{-,r}), \ell(T_{+,r})\right)$$

The "quadratic" + "multiplier" decomposition

Decomposition of the empirical excess loss :

$$P_N\mathcal{L}_x = \frac{1}{N}\sum_{i=1}^{N}\left\langle a_i, x-x_0\right\rangle^2\left\langle a_i, x+x_0\right\rangle^2 - \frac{2\sigma}{N}\sum_{i=1}^{N} g_i\left\langle a_i, x-x_0\right\rangle\left\langle a_i, x+x_0\right\rangle$$

The "quadratic" + "multiplier" decomposition

Decomposition of the empirical excess loss :

$$P_N\mathcal{L}_x = \frac{1}{N}\sum_{i=1}^{N}\big\langle a_i, x - x_0\big\rangle^2\big\langle a_i, x + x_0\big\rangle^2 - \frac{2\sigma}{N}\sum_{i=1}^{N} g_i\big\langle a_i, x - x_0\big\rangle\big\langle a_i, x + x_0\big\rangle$$

1. "The quadratic term" :

$$\frac{1}{N}\sum_{i=1}^{N}\big\langle a_i, x - x_0\big\rangle^2\big\langle a_i, x + x_0\big\rangle^2$$

The “quadratic” + “multiplier” decomposition

Decomposition of the empirical excess loss :

$$P_N\mathcal{L}_x = \frac{1}{N}\sum_{i=1}^{N}\big\langle a_i, x-x_0\big\rangle^2\big\langle a_i, x+x_0\big\rangle^2 - \frac{2\sigma}{N}\sum_{i=1}^{N} g_i\big\langle a_i, x-x_0\big\rangle\big\langle a_i, x+x_0\big\rangle$$

1. “The quadratic term” :

$$\frac{1}{N}\sum_{i=1}^{N}\big\langle a_i, x-x_0\big\rangle^2\big\langle a_i, x+x_0\big\rangle^2$$

power 4 of sub-gaussian variables (badly concentrated $\sim \psi_{1/2}$)

The "quadratic" + "multiplier" decomposition

Decomposition of the empirical excess loss :

$$P_N\mathcal{L}_x = \frac{1}{N}\sum_{i=1}^{N}\left\langle a_i, x-x_0\right\rangle^2\left\langle a_i, x+x_0\right\rangle^2 - \frac{2\sigma}{N}\sum_{i=1}^{N} g_i\left\langle a_i, x-x_0\right\rangle\left\langle a_i, x+x_0\right\rangle$$

1. "The quadratic term" :

$$\frac{1}{N}\sum_{i=1}^{N}\left\langle a_i, x-x_0\right\rangle^2\left\langle a_i, x+x_0\right\rangle^2$$

power 4 of sub-gaussian variables (badly concentrated $\sim \psi_{1/2}$) - control via an "empirical small ball estimate".

The "quadratic" + "multiplier" decomposition

Decomposition of the empirical excess loss :

$$P_N \mathcal{L}_x = \frac{1}{N}\sum_{i=1}^N \big\langle a_i, x - x_0\big\rangle^2 \big\langle a_i, x + x_0\big\rangle^2 - \frac{2\sigma}{N}\sum_{i=1}^N g_i \big\langle a_i, x - x_0\big\rangle \big\langle a_i, x + x_0\big\rangle$$

1. "The quadratic term" :

$$\frac{1}{N}\sum_{i=1}^N \big\langle a_i, x - x_0\big\rangle^2 \big\langle a_i, x + x_0\big\rangle^2$$

   power 4 of sub-gaussian variables (badly concentrated $\sim \psi_{1/2}$) - control via an "empirical small ball estimate".

2. "The multiplier term" :

$$\frac{2\sigma}{N}\sum_{i=1}^N g_i \big\langle a_i, x - x_0\big\rangle \big\langle a_i, x + x_0\big\rangle$$

The "quadratic" + "multiplier" decomposition

Decomposition of the empirical excess loss :

$$P_N\mathcal{L}_x = \frac{1}{N}\sum_{i=1}^N \big\langle a_i, x-x_0\big\rangle^2 \big\langle a_i, x+x_0\big\rangle^2 - \frac{2\sigma}{N}\sum_{i=1}^N g_i \big\langle a_i, x-x_0\big\rangle \big\langle a_i, x+x_0\big\rangle$$

1. "The quadratic term" :

$$\frac{1}{N}\sum_{i=1}^N \big\langle a_i, x-x_0\big\rangle^2 \big\langle a_i, x+x_0\big\rangle^2$$

power 4 of sub-gaussian variables (badly concentrated $\sim \psi_{1/2}$) - control via an "empirical small ball estimate".

2. "The multiplier term" :

$$\frac{2\sigma}{N}\sum_{i=1}^N g_i \big\langle a_i, x-x_0\big\rangle \big\langle a_i, x+x_0\big\rangle$$

power 3 of a subgaussian variables ($\sim \psi_{2/3}$) - control via contraction principle : a $\sqrt{\log N}$ extra term.

Control of the "quadratic" term via empirical small ball estimate

Proposition

If $\sqrt{N} \gtrsim E_r$ then w.h.p. for any $t \in T$ such that $\|t - x_0\|_2 \|t + x_0\|_2 \geq r$, there exists $I_t \subset \{1, \dots, N\}$ such that $|I_t| \gtrsim N$

Control of the "quadratic" term via empirical small ball estimate

Proposition

If $\sqrt{N} \gtrsim E_r$ then w.h.p. for any $t \in T$ such that $\|t - x_0\|_2 \|t + x_0\|_2 \geq r$, there exists $I_t \subset \{1, \ldots, N\}$ such that $|I_t| \gtrsim N$ and $\forall i \in I_t$

$$|\langle t - x_0, a_i \rangle \langle t + x_0, a_i \rangle| \gtrsim \|t - x_0\|_2 \|t + x_0\|_2.$$

Control of the "quadratic" term via empirical small ball estimate

Proposition

If $\sqrt{N} \gtrsim E_r$ then w.h.p. for any $t \in T$ such that $\|t - x_0\|_2 \|t + x_0\|_2 \geq r$, there exists $I_t \subset \{1, \dots, N\}$ such that $|I_t| \gtrsim N$ and $\forall i \in I_t$

$$\left|\left\langle t - x_0, a_i \right\rangle \left\langle t + x_0, a_i \right\rangle\right| \gtrsim \|t - x_0\|_2 \|t + x_0\|_2.$$

If $\sqrt{N} \gtrsim E_r$ then w.h.p if $t \in T$ is such that $\|t - x_0\|_2 \|t + x_0\|_2 \geq r$:

$$\frac{1}{N} \sum_{i=1}^{N} \left\langle t - x_0, a_i \right\rangle^2 \left\langle t + x_0, a_i \right\rangle^2 \gtrsim \|t - x_0\|_2^2 \|t + x_0\|_2^2.$$

Control of the "quadratic" term via empirical small ball estimate

Proposition

If $\sqrt{N} \gtrsim E_r$ then w.h.p. for any $t \in T$ such that $\|t - x_0\|_2 \|t + x_0\|_2 \geq r$, there exists $I_t \subset \{1, \ldots, N\}$ such that $|I_t| \gtrsim N$ and $\forall i \in I_t$

$$\left|\left\langle t - x_0, a_i \right\rangle \left\langle t + x_0, a_i \right\rangle\right| \gtrsim \|t - x_0\|_2 \|t + x_0\|_2.$$

If $\sqrt{N} \gtrsim E_r$ then w.h.p if $t \in T$ is such that $\|t - x_0\|_2 \|t + x_0\|_2 \geq r$:

$$\frac{1}{N} \sum_{i=1}^{N} \left\langle t - x_0, a_i \right\rangle^2 \left\langle t + x_0, a_i \right\rangle^2 \gtrsim \|t - x_0\|_2^2 \|t + x_0\|_2^2.$$

Sharp control of the "quadratic term" as long as $\sqrt{N} \gtrsim E_r$: "fixed point equation" for r when the noise is small.

Control of the "multiplier" process

w.h.p. for any $t \in T$ such that $\|t - x_0\|_2 \|t + x_0\|_2 \geq r$,

$$\left| \frac{1}{N} \sum_{i=1}^{N} \sigma g_i \langle a_i, t - x_0 \rangle \langle a_i, t + x_0 \rangle \right| \lesssim \sigma \sqrt{\log N} \frac{E_r}{\sqrt{N}} \|t - x_0\|_2 \|t + x_0\|.$$

Control of the "multiplier" process

w.h.p. for any $t \in T$ such that $\|t - x_0\|_2 \|t + x_0\|_2 \geq r$,

$$\left| \frac{1}{N} \sum_{i=1}^{N} \sigma g_i \langle a_i, t - x_0 \rangle \langle a_i, t + x_0 \rangle \right| \lesssim \sigma \sqrt{\log N} \frac{E_r}{\sqrt{N}} \|t - x_0\|_2 \|t + x_0\|.$$

Following these two estimates : if $\sqrt{N} \gtrsim E_r$

Control of the "multiplier" process

w.h.p. for any $t \in T$ such that $\|t - x_0\|_2 \|t + x_0\|_2 \geq r$,

$$\left| \frac{1}{N} \sum_{i=1}^{N} \sigma g_i \langle a_i, t - x_0 \rangle \langle a_i, t + x_0 \rangle \right| \lesssim \sigma \sqrt{\log N} \frac{E_r}{\sqrt{N}} \|t - x_0\|_2 \|t + x_0\|.$$

Following these two estimates : if $\sqrt{N} \gtrsim E_r$ and if

$$\sigma \sqrt{\log N} \frac{E_r}{\sqrt{N}} \lesssim r$$

Control of the “multiplier” process

w.h.p. for any $t \in T$ such that $\|t - x_0\|_2 \|t + x_0\|_2 \geq r$,

$$\left| \frac{1}{N} \sum_{i=1}^{N} \sigma g_i \langle a_i, t - x_0 \rangle \langle a_i, t + x_0 \rangle \right| \lesssim \sigma \sqrt{\log N} \frac{E_r}{\sqrt{N}} \|t - x_0\|_2 \|t + x_0\|.$$

Following these two estimates : if $\sqrt{N} \gtrsim E_r$ and if

$$\sigma \sqrt{\log N} \frac{E_r}{\sqrt{N}} \lesssim r$$

then for all $t \in T_{\geq r}$, $P_N \mathcal{L}_t > 0$

Control of the "multiplier" process

w.h.p. for any $t \in T$ such that $\|t - x_0\|_2 \|t + x_0\|_2 \geq r$,

$$\left| \frac{1}{N} \sum_{i=1}^{N} \sigma g_i \langle a_i, t - x_0 \rangle \langle a_i, t + x_0 \rangle \right| \lesssim \sigma \sqrt{\log N} \frac{E_r}{\sqrt{N}} \|t - x_0\|_2 \|t + x_0\|.$$

Following these two estimates : if $\sqrt{N} \gtrsim E_r$ and if

$$\sigma \sqrt{\log N} \frac{E_r}{\sqrt{N}} \lesssim r$$

then for all $t \in T_{\geq r}$, $P_N \mathcal{L}_t > 0$ but $P_N \mathcal{L}_{\hat{x}} \leq 0$ therefore, $\hat{x} \notin T_{\geq r}$:

$$\|\hat{x} - x_0\|_2 \|\hat{x} + x_0\|_2 \leq r.$$

Two fixed points

1. Complexity coming from the noise measured via the "multiplier process's fixed point" :

Two fixed points

1. Complexity coming from the noise measured via the "multiplier process's fixed point" :

$$r_2^* = \inf\left(r > 0 : \sigma\sqrt{\log N} E_r \lesssim r\sqrt{N}\right)$$

Two fixed points

1. Complexity coming from the noise measured via the "multiplier process's fixed point" :

$$r_2^* = \inf\left(r > 0 : \sigma\sqrt{\log N} E_r \lesssim r\sqrt{N}\right)$$

2. Complexity coming from the projection measured via the "quadratic process's fixed point" :

Two fixed points

1. Complexity coming from the noise measured via the "multiplier process's fixed point" :

$$r_2^* = \inf\left(r > 0 : \sigma\sqrt{\log N} E_r \lesssim r\sqrt{N}\right)$$

2. Complexity coming from the projection measured via the "quadratic process's fixed point" :

$$r_0^* = \inf\left(r > 0 : E_r \lesssim \sqrt{N}\right).$$

Two fixed points

1. Complexity coming from the noise measured via the "multiplier process's fixed point" :

$$r_2^* = \inf\left(r > 0 : \sigma\sqrt{\log N} E_r \lesssim r\sqrt{N}\right)$$

2. Complexity coming from the projection measured via the "quadratic process's fixed point" :

$$r_0^* = \inf\left(r > 0 : E_r \lesssim \sqrt{N}\right).$$

W.h.p.

$$\|\hat{x} - x_0\|_2 \|\hat{x} + x_0\|_2 \lesssim \max(r_0^*, r_2^*).$$

Two fixed points

1. Complexity coming from the noise measured via the "multiplier process's fixed point" :

$$r_2^* = \inf\left(r > 0 : \sigma\sqrt{\log N} E_r \lesssim r\sqrt{N}\right)$$

2. Complexity coming from the projection measured via the "quadratic process's fixed point" :

$$r_0^* = \inf\left(r > 0 : E_r \lesssim \sqrt{N}\right).$$

W.h.p.

$$\|\hat{x} - x_0\|_2 \|\hat{x} + x_0\|_2 \lesssim \max(r_0^*, r_2^*).$$

Nevertheless, it is easier to understand a result like

$$\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right) \leq rate.$$

Relation between $\|\hat{x} - x_0\|_2 \|\hat{x} + x_0\|_2$ and $\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right)$

Relation between $\|\hat{x} - x_0\|_2 \|\hat{x} + x_0\|_2$ and $\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right)$

1. when $\|x_0\|_2$ is large ($\geq \min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right)$)

Relation between $\|\hat{x} - x_0\|_2 \|\hat{x} + x_0\|_2$ and $\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right)$

1. when $\|x_0\|_2$ is large $(\geq \min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right))$

$$\|\hat{x} - x_0\|_2 \|\hat{x} + x_0\|_2 \sim \|x_0\|_2 \min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right)$$

Relation between $\|\hat{x}-x_0\|_2\|\hat{x}+x_0\|_2$ and $\min\left(\|\hat{x}-x_0\|_2, \|\hat{x}+x_0\|_2\right)$

1. when $\|x_0\|_2$ is large ($\geq \min\left(\|\hat{x}-x_0\|_2, \|\hat{x}+x_0\|_2\right)$)

$$\|\hat{x}-x_0\|_2\|\hat{x}+x_0\|_2 \sim \|x_0\|_2 \min\left(\|\hat{x}-x_0\|_2, \|\hat{x}+x_0\|_2\right)$$

2. when $\|x_0\|_2$ is small ($\leq \min\left(\|\hat{x}-x_0\|_2, \|\hat{x}+x_0\|_2\right)$)

Relation between $\|\hat{x}-x_0\|_2\|\hat{x}+x_0\|_2$ and $\min\left(\|\hat{x}-x_0\|_2, \|\hat{x}+x_0\|_2\right)$

1. when $\|x_0\|_2$ is large ($\geq \min\left(\|\hat{x}-x_0\|_2, \|\hat{x}+x_0\|_2\right)$)

$$\|\hat{x}-x_0\|_2\|\hat{x}+x_0\|_2 \sim \|x_0\|_2 \min\left(\|\hat{x}-x_0\|_2, \|\hat{x}+x_0\|_2\right)$$

2. when $\|x_0\|_2$ is small ($\leq \min\left(\|\hat{x}-x_0\|_2, \|\hat{x}+x_0\|_2\right)$)

$$\|\hat{x}-x_0\|_2\|\hat{x}+x_0\|_2 \sim \min\left(\|\hat{x}-x_0\|_2^2, \|\hat{x}+x_0\|_2^2\right)$$

Relation between $\|\hat{x}-x_0\|_2\|\hat{x}+x_0\|_2$ and $\min\left(\|\hat{x}-x_0\|_2, \|\hat{x}+x_0\|_2\right)$

1. when $\|x_0\|_2$ is large $\left(\geq \min\left(\|\hat{x}-x_0\|_2, \|\hat{x}+x_0\|_2\right)\right)$

$$\|\hat{x}-x_0\|_2\|\hat{x}+x_0\|_2 \sim \|x_0\|_2 \min\left(\|\hat{x}-x_0\|_2, \|\hat{x}+x_0\|_2\right)$$

2. when $\|x_0\|_2$ is small $\left(\leq \min\left(\|\hat{x}-x_0\|_2, \|\hat{x}+x_0\|_2\right)\right)$

$$\|\hat{x}-x_0\|_2\|\hat{x}+x_0\|_2 \sim \min\left(\|\hat{x}-x_0\|_2^2, \|\hat{x}+x_0\|_2^2\right)$$

$\Rightarrow$ 2 regimes for (the localization and thus) the complexity term E_r depending on $\|x_0\|_2$.

$$r_N^* = \inf\left(r > 0 : \ell(2T \cap rB_2^d) \lesssim r\sqrt{N}\right)$$

$$s_N^* = \inf\left(s > 0 : \ell(2T \cap sB_2^d) \lesssim \frac{\|x_0\|_2}{\sigma\sqrt{\log N}} s^2\sqrt{N}\right)$$

$$v_N^* = \inf\left(s > 0 : \ell(2T \cap vB_2^d) \lesssim \frac{1}{\sigma\sqrt{\log N}} v^3\sqrt{N}\right)$$

w.h.p.

$$\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right) \leq rate$$

$$r_N^* = \inf\left(r > 0 : \ell(2T \cap rB_2^d) \lesssim r\sqrt{N}\right)$$

$$s_N^* = \inf\left(s > 0 : \ell(2T \cap sB_2^d) \lesssim \frac{\|x_0\|_2}{\sigma\sqrt{\log N}} s^2\sqrt{N}\right)$$

$$v_N^* = \inf\left(s > 0 : \ell(2T \cap vB_2^d) \lesssim \frac{1}{\sigma\sqrt{\log N}} v^3\sqrt{N}\right)$$

w.h.p.

$$\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right) \leq rate$$

where

$rate$	$\sigma\sqrt{\log N} \leq \|x_0\|_2 r_N^*$	$\sigma\sqrt{\log N} \geq \|x_0\|_2 r_N^*$
$\|x_0\|_2 \geq v_N^*$	r_N^*	s_N^*
$\|x_0\|_2 \leq v_N^*$	r_N^*	v_N^*

Sparse vectors

$$T = W_s = \{x \in \mathbb{R}^d : |\mathrm{supp}(x)| \le s\}$$

$$T = W_s = \{x \in \mathbb{R}^d : |\mathrm{supp}(x)| \leq s\}$$

1. Gaussian complexity of localized sets :

$$\ell(W_s \cap rB_2^n) \sim r\sqrt{s \log\left(ed/s\right)}.$$

$$T = W_s = \{x \in \mathbb{R}^d : |\mathrm{supp}(x)| \leq s\}$$

1. Gaussian complexity of localized sets :

$$\ell(W_s \cap rB_2^n) \sim r\sqrt{s \log\left(ed/s\right)}.$$

2. Sudakov complexity of localized sets

$$r \log^{1/2} N\left(W_s \cap 2rB_2^d, rB_2^d\right) \sim r\sqrt{s \log\left(ed/s\right)}.$$

$$T = W_s = \{x \in \mathbb{R}^d : |\mathrm{supp}(x)| \leq s\}$$

1. Gaussian complexity of localized sets :

$$\ell(W_s \cap rB_2^n) \sim r\sqrt{s \log\left(ed/s\right)}.$$

2. Sudakov complexity of localized sets

$$r \log^{1/2} N\left(W_s \cap 2rB_2^d, rB_2^d\right) \sim r\sqrt{s \log\left(ed/s\right)}.$$

Sudakov inequality is sharp : $r \log^{1/2} N\left(W_s \cap 2rB_2^d, rB_2^d\right) \sim \ell(W_s \cap rB_2^n)$

$$\Downarrow$$

ERM is minimax in linear regression and phase recovery (up to $\sqrt{\log N}$) in the noisy setup.

Computing the three fixed points

1. when $N \gtrsim s \log\left(ed/s\right)$ then

$$r_N^* = 0$$

Computing the three fixed points

1. when $N \gtrsim s \log\left(ed/s\right)$ then

$$r_N^* = 0$$

(otherwise, we don't have isomorphy in Linear Regression and small ball estimate in phase recovery).

Computing the three fixed points

1. when $N \gtrsim s \log\left(ed/s\right)$ then

$$r_N^* = 0$$

   (otherwise, we don't have isomorphy in Linear Regression and small ball estimate in phase recovery).

2. $$s_N^*(\eta) \sim \frac{1}{\eta}\sqrt{\frac{s\log(ed/s)}{N}}$$

   $\eta \sim \sigma^{-1}$ in Linear Regression

Computing the three fixed points

1. when $N \gtrsim s \log\left(ed/s\right)$ then
$$r_N^* = 0$$
(otherwise, we don't have isomorphy in Linear Regression and small ball estimate in phase recovery).

2. $$s_N^*(\eta) \sim \frac{1}{\eta}\sqrt{\frac{s\log(ed/s)}{N}}$$
$\eta \sim \sigma^{-1}$ in Linear Regression and $\eta \sim \|x_0\|_2/\sigma$ in Phase Recovery.

Computing the three fixed points

1. when $N \gtrsim s \log\left(ed/s\right)$ then
$$r_N^* = 0$$
(otherwise, we don't have isomorphy in Linear Regression and small ball estimate in phase recovery).

2. $$s_N^*(\eta) \sim \frac{1}{\eta}\sqrt{\frac{s\log(ed/s)}{N}}$$
$\eta \sim \sigma^{-1}$ in Linear Regression and $\eta \sim \|x_0\|_2/\sigma$ in Phase Recovery.

3. $$v_N^* \sim \left[\sigma\sqrt{\frac{s\log(ed/s)}{N}}\right]^{1/2}.$$

Rates of convergence over W_s in linear regression and phase recovery

When $N \gtrsim s \log\left(ed/s\right)$. In linear regression :

1. if $\sigma = 0$ then w.h.p. $\hat{x} = x_0$.

Rates of convergence over W_s in linear regression and phase recovery

When $N \gtrsim s \log\left(ed/s\right)$. In linear regression :

1. if $\sigma = 0$ then w.h.p. $\hat{x} = x_0$.
2. if $\sigma > 0$, then w.h.p.

$$\|\hat{x} - x_0\|_2 \lesssim s_N^*(\sigma^{-1}) \sim \sigma \sqrt{\frac{s \log\left(ed/s\right)}{N}}.$$

Rates of convergence over W_s in linear regression and phase recovery

When $N \gtrsim s \log\left(ed/s\right)$. In linear regression :

1. if $\sigma = 0$ then w.h.p. $\hat{x} = x_0$.
2. if $\sigma > 0$, then w.h.p.

$$\|\hat{x} - x_0\|_2 \lesssim s_N^*(\sigma^{-1}) \sim \sigma\sqrt{\frac{s\log\left(ed/s\right)}{N}}.$$

In phase recovery :

1. if $\sigma = 0$ then w.h.p. $\hat{x} = x_0$ or $\hat{x} = -x_0$.

Rates of convergence over W_s in linear regression and phase recovery

When $N \gtrsim s \log\left(ed/s\right)$. In linear regression :

1. if $\sigma = 0$ then w.h.p. $\hat{x} = x_0$.
2. if $\sigma > 0$, then w.h.p.

$$\|\hat{x} - x_0\|_2 \lesssim s_N^*(\sigma^{-1}) \sim \sigma \sqrt{\frac{s \log\left(ed/s\right)}{N}}.$$

In phase recovery :

1. if $\sigma = 0$ then w.h.p. $\hat{x} = x_0$ or $\hat{x} = -x_0$.
2. if $\sigma > 0$, then
   - if $\|x_0\|_2 \geq v_N^*$, then w.h.p.

Rates of convergence over W_s in linear regression and phase recovery

When $N \gtrsim s\log\left(ed/s\right)$. In linear regression :

1. if $\sigma = 0$ then w.h.p. $\hat{x} = x_0$.
2. if $\sigma > 0$, then w.h.p.

$$\|\hat{x} - x_0\|_2 \lesssim s_N^*(\sigma^{-1}) \sim \sigma\sqrt{\frac{s\log\left(ed/s\right)}{N}}.$$

In phase recovery :

1. if $\sigma = 0$ then w.h.p. $\hat{x} = x_0$ or $\hat{x} = -x_0$.
2. if $\sigma > 0$, then
   - if $\|x_0\|_2 \geq v_N^*$, then w.h.p.

$$\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right) \leq s_N^*\left(\frac{\|x_0\|_2}{\sigma}\right) \sim \frac{\sigma}{\|x_0\|_2}\sqrt{\frac{s\log\left(ed/s\right)}{N}}$$

Rates of convergence over W_s in linear regression and phase recovery

When $N \gtrsim s\log\left(ed/s\right)$. In linear regression :

1. if $\sigma = 0$ then w.h.p. $\hat{x} = x_0$.
2. if $\sigma > 0$, then w.h.p.

$$\|\hat{x} - x_0\|_2 \lesssim s_N^*(\sigma^{-1}) \sim \sigma\sqrt{\frac{s\log\left(ed/s\right)}{N}}.$$

In phase recovery :

1. if $\sigma = 0$ then w.h.p. $\hat{x} = x_0$ or $\hat{x} = -x_0$.
2. if $\sigma > 0$, then
   - if $\|x_0\|_2 \geq v_N^*$, then w.h.p.

$$\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right) \leq s_N^*\left(\frac{\|x_0\|_2}{\sigma}\right) \sim \frac{\sigma}{\|x_0\|_2}\sqrt{\frac{s\log\left(ed/s\right)}{N}}$$

   - if $\|x_0\|_2 \leq v_N^*$, then w.h.p.

Rates of convergence over W_s in linear regression and phase recovery

When $N \gtrsim s \log\left(ed/s\right)$. In linear regression :

1. if $\sigma = 0$ then w.h.p. $\hat{x} = x_0$.
2. if $\sigma > 0$, then w.h.p.

$$\|\hat{x} - x_0\|_2 \lesssim s_N^*(\sigma^{-1}) \sim \sigma \sqrt{\frac{s \log\left(ed/s\right)}{N}}.$$

In phase recovery :

1. if $\sigma = 0$ then w.h.p. $\hat{x} = x_0$ or $\hat{x} = -x_0$.
2. if $\sigma > 0$, then
   - if $\|x_0\|_2 \geq v_N^*$, then w.h.p.

$$\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right) \leq s_N^*\left(\frac{\|x_0\|_2}{\sigma}\right) \sim \frac{\sigma}{\|x_0\|_2} \sqrt{\frac{s \log\left(ed/s\right)}{N}}$$

   - if $\|x_0\|_2 \leq v_N^*$, then w.h.p.

$$\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right) \leq v_N^*\left(\sigma^{-1}\right) \sim \left[\sigma \sqrt{\frac{s \log\left(ed/s\right)}{N}}\right]^{1/2}$$

The unit B_1^d-ball

$$T = B_1^d = \{x \in \mathbb{R}^d : \|x\|_1 \le 1\}$$

$$T = B_1^d = \{x \in \mathbb{R}^d : \|x\|_1 \leq 1\}$$

1. Gaussian complexity of localized sets :

$$\ell(B_1^d \cap rB_2^d) \sim \begin{cases} \sqrt{\log\left(edr^2\right)} & \text{if } d^2 r \geq 1 \\ r\sqrt{d} & \text{otherwise} \end{cases}$$

$$T = B_1^d = \{x \in \mathbb{R}^d : \|x\|_1 \le 1\}$$

1. Gaussian complexity of localized sets :

$$\ell(B_1^d \cap rB_2^d) \sim \begin{cases} \sqrt{\log\left(edr^2\right)} & \text{if } d^2 r \ge 1 \\ r\sqrt{d} & \text{otherwise} \end{cases}$$

2. Sudakov complexity of localized sets is sharp : for any $r > 0$,

$$r \log^{1/2} N\left(B_1^d \cap 2rB_2^d, rB_2^d\right) \sim \ell(B_1^d \cap rB_2^d).$$

$$T = B_1^d = \{x \in \mathbb{R}^d : \|x\|_1 \leq 1\}$$

1. Gaussian complexity of localized sets :

$$\ell(B_1^d \cap rB_2^d) \sim \begin{cases} \sqrt{\log\left(edr^2\right)} & \text{if } d^2 r \geq 1 \\ r\sqrt{d} & \text{otherwise} \end{cases}$$

2. Sudakov complexity of localized sets is sharp : for any $r > 0$,

$$r \log^{1/2} N\big(B_1^d \cap 2rB_2^d, rB_2^d\big) \sim \ell(B_1^d \cap rB_2^d).$$

$$\Downarrow$$

ERM is minimax in linear regression and phase recovery (up to $\sqrt{\log N}$) in the noisy case (and also in the noise free case).

Computing the three fixed points

$$r_N^*(Q) \begin{cases} \sim \left(\frac{1}{Q^2 N}\log\left(\frac{n}{Q^2 N}\right)\right)^{1/2} & \text{if } n \geq C_0 Q^2 N \\ \lesssim \frac{1}{N} & \text{if } C_1 Q^2 N \leq n \leq C_0 Q^2 N \\ = 0 & \text{if } n \leq C_1 Q^2 N. \end{cases}$$

Computing the three fixed points

$$r_N^*(Q) \begin{cases} \sim \left(\frac{1}{Q^2 N}\log\left(\frac{n}{Q^2 N}\right)\right)^{1/2} & \text{if } n \geq C_0 Q^2 N \\ \lesssim \frac{1}{N} & \text{if } C_1 Q^2 N \leq n \leq C_0 Q^2 N \\ = 0 & \text{if } n \leq C_1 Q^2 N. \end{cases}$$

$$s_N^*(\eta) \sim \begin{cases} \left(\frac{1}{\eta^2 N}\log\left(\frac{n^2}{\eta^2 N}\right)\right)^{1/4} & \text{if } n \geq \eta\sqrt{N} \\ \sqrt{\frac{n}{\eta^2 N}} & \text{if } n \leq \eta\sqrt{N} \end{cases}$$

Computing the three fixed points

$$
r_N^*(Q) \begin{cases} \sim \left(\frac{1}{Q^2 N} \log\left(\frac{n}{Q^2 N} \right) \right)^{1/2} & \text{if } n \geq C_0 Q^2 N \\ \lesssim \frac{1}{N} & \text{if } C_1 Q^2 N \leq n \leq C_0 Q^2 N \\ = 0 & \text{if } n \leq C_1 Q^2 N. \end{cases}
$$

$$
s_N^*(\eta) \sim \begin{cases} \left(\frac{1}{\eta^2 N} \log\left(\frac{n^2}{\eta^2 N} \right) \right)^{1/4} & \text{if } n \geq \eta\sqrt{N} \\ \sqrt{\frac{n}{\eta^2 N}} & \text{if } n \leq \eta\sqrt{N} \end{cases}
$$

$$
v_N^*(\zeta) \sim \begin{cases} \left(\frac{1}{\zeta^2 N} \log\left(\frac{n^3}{\zeta^2 N} \right) \right)^{1/6} & \text{if } n \geq \zeta^{2/3} N^{1/3} \\ \left(\frac{n}{\zeta^2 N} \right)^{1/4} & \text{if } n \leq \zeta^{2/3} N^{1/3}. \end{cases}
$$

Rates of convergence over B_1^d when $\sigma = 0$

1. when $d \gtrsim N$ then
   - in Linear regression : w.h.p.

$$\|\hat{x} - x_0\|_2 \lesssim \sqrt{\frac{1}{N} \log\left(\frac{ed}{N}\right)}$$

Rates of convergence over B_1^d when $\sigma = 0$

1. when $d \gtrsim N$ then
   - in Linear regression : w.h.p.

$$\|\hat{x} - x_0\|_2 \lesssim \sqrt{\frac{1}{N}\log\left(\frac{ed}{N}\right)}$$

   - in Phase recovery : w.h.p.

$$\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right) \lesssim \sqrt{\frac{1}{N}\log\left(\frac{ed}{N}\right)}$$

Rates of convergence over B_1^d when $\sigma = 0$

1. when $d \gtrsim N$ then
   - in Linear regression : w.h.p.
     $$\|\hat{x} - x_0\|_2 \lesssim \sqrt{\frac{1}{N}\log\left(\frac{ed}{N}\right)}$$
   - in Phase recovery : w.h.p.
     $$\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right) \lesssim \sqrt{\frac{1}{N}\log\left(\frac{ed}{N}\right)}$$
2. when $d \lesssim N$ then
   - in Linear regression : w.h.p. $\hat{x} = x_0$

Rates of convergence over B_1^d when $\sigma = 0$

1. when $d \gtrsim N$ then
   - in Linear regression : w.h.p.

$$\|\hat{x} - x_0\|_2 \lesssim \sqrt{\frac{1}{N}\log\left(\frac{ed}{N}\right)}$$

   - in Phase recovery : w.h.p.

$$\min\left(\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\right) \lesssim \sqrt{\frac{1}{N}\log\left(\frac{ed}{N}\right)}$$

2. when $d \lesssim N$ then
   - in Linear regression : w.h.p. $\hat{x} = x_0$
   - in Phase recovery : w.h.p. $\hat{x} = x_0$ or $\hat{x} = -x_0$.

Thanks for your attention