Still blind deconvolution

Elisabeth Gassiat

Université Paris-Saclay (LM-Orsay) and CNRS

MASCOT NUM 2023

Outline

Once upon a time

Once upon a time

The Annals of Statistics
1996, Vol. 24, No. 5, 1964–1981

BLIND DECONVOLUTION OF DISCRETE LINEAR SYSTEMS

BY F. GAMBOA AND E. GASSIAT

The Annals of Statistics
1997, Vol. 25, No. 1, 328–350

BAYESIAN METHODS AND MAXIMUM ENTROPY FOR ILL-POSED INVERSE PROBLEMS

BY F. GAMBOA[1] AND E. GASSIAT[2]

Once upon a time

The Annals of Statistics
1996, Vol. 24, No. 5, 1964–1981

BLIND DECONVOLUTION OF DISCRETE LINEAR SYSTEMS

BY F. GAMBOA AND E. GASSIAT

The Annals of Statistics
1997, Vol. 25, No. 1, 328–350

BAYESIAN METHODS AND MAXIMUM ENTROPY FOR ILL-POSED INVERSE PROBLEMS

BY F. GAMBOA[1] AND E. GASSIAT[2]

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 45, NO. 7, NOVEMBER 1999

MEM Pixel Correlated Solutions for Generalized Moment and Interpolation Problems

Imre Csiszár, *Fellow, IEEE*, Fabrice Gamboa, and Elisabeth Gassiat

Erdos number = 2 ! ! !

Deconvolution

The observation $\mathbf{Y}$ is given by

$$\mathbf{Y} = \mathbf{X} + \varepsilon \, ,$$

$\mathbf{X}$ is the signal and ε is the noise, $\mathbf{X}$ and ε are independent random variables.

Goal : on the basis of observations $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, learn the distribution of $\mathbf{X}$.

Deconvolution

The observation $\mathbf{Y}$ is given by

$$\mathbf{Y} = \mathbf{X} + \varepsilon\,,$$

$\mathbf{X}$ is the signal and ε is the noise, $\mathbf{X}$ and ε are independent random variables.

Goal : on the basis of observations $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, learn the distribution of $\mathbf{X}$.

Characteristic functions verify

$$\mathbb{E}\left(e^{i\langle t,\mathbf{Y}\rangle}\right) = \mathbb{E}\left(e^{i\langle t,\mathbf{X}\rangle}\right)\mathbb{E}\left(e^{i\langle t,\varepsilon\rangle}\right)\,.$$

$\rightarrow$ estimators based on knowledge of the characteristic function of the noise.

Topic of my talk

Deconvolution with totally unknown noise is possible for multidimensional signals (under very weak assumptions) !

Application to

- density estimation
- support estimation
- low-dimensional distribution estimation

When the characteristic function of the noise is known

or with independent data of the noise

Estimators based on knowledge of the characteristic function of the noise (Fourier/Kernel method ; bandwith selection ; non vanishing characteristic function of the noise)

$$\widehat{\mathbb{E}\left(e^{i\langle t,\mathbf{X}\rangle}\right)} = \frac{\widehat{\mathbb{E}\left(e^{i\langle t,\mathbf{Y}\rangle}\right)}}{\mathbb{E}\left(e^{i\langle t,\varepsilon\rangle}\right)}$$

$\rightarrow$ Hardness depends (mostly) on the way the characteristic function of the noise decreases at infinity.

Review of known results : density estimation

When the characteristic function of the noise is known

$$|\mathbb{E}\left(e^{i\langle t,\varepsilon\rangle}\right)| \geq A(\|t\|^2+1)^{-\gamma/2}\exp(-B\|t\|^{\alpha})$$

If $\mathbf{X}$ has density f, let $R_n(\mathcal{F}) = \inf_{\widehat{f}}\sup_{f\in\mathcal{F}}\|\widehat{f}-f\|_2^2$ and let $\mathcal{F}$ be a Sobolev class with regularity β. Observations are in $\mathbb{R}^d$.

- Exponentially smooth errors : $B>0$.
  Then $R_n(\mathcal{F})$ is of order $(\log n)^{-2\beta/\alpha}$.

- Ordinary smooth errors : $B=0$.
  Then $R_n(\mathcal{F})$ is of order $n^{-2\beta/(2\beta+2\gamma+d)}$.

Some (in the huge) literature : Carroll and Hall (1988), Fan (1991) ; Tang (1994) ; Matias (2002) ; Meister (2009) ; Comte and Lacour (2011).

Review of known results : distribution estimation

When the characteristic function of the noise is known

Wasserstein metrics : estimation of the distribution without assuming a density. If $\mathbf{X}$ has law μ, let

$$R_n(\mathcal{G}) = \inf_{\widehat{\mu}} \sup_{\mu \in \mathcal{G}} W_p(\widehat{\mu}, \mu)^p$$

• Exponentially smooth errors : general lower bound in any dimension (Dedecker and Michel JMA 2013). Here $\mathcal{G}$ requires bounded $2p + a$ moments (for some $a > 1$). When the noise is Gaussian (that is $\alpha = 2$), then $R_n(\mathcal{G})$ is of order $(\log n)^{-p/2}$.

• Ordinary smooth errors : in dimension 1 (Dedecker, Fisher, Michel EJS 2015) and moment-like conditions.

Then $R_n(\mathcal{G})$ is less than $\begin{cases} n^{-p/(2p+2\gamma-1)} & \text{if } \gamma > 1/2 \\ \sqrt{(\log n)/n} & \text{if } \gamma = 1/2 \\ 1/\sqrt{n} & \text{if } \gamma < 1/2 \end{cases}$

Review of known results : support estimation

When the characteristic function of the noise is known

Gaussian noise and (truncated) Hausdorff loss.

Genovese et al (AOS 2012). Assumption on the regularity of the support (lower bounded reach). Estimation using the level set of a smoothed density, requires the knowledge of the intrinsic dimension.

Maximum risk upper bounded by $C\left(\frac{1}{\log n}\right)^{\frac{1}{2}-\delta}$.

Lower bound of the minimax risk : $\frac{c}{\log n}$.

Brunel et al. (Bernoulli 2021). The support is a full dimensional convex body. Estimation using endpoints estimation.

Maximum risk upper bounded by $C\left(\frac{\log\log n}{\sqrt{\log n}}\right)$.

Lower bound of the minimax risk with reach more than $\tau \in (0,1)$:

$c\left(\frac{1}{\log n}\right)^{\frac{2}{\tau}}$.

Robustness to the assumptions on the noise

• Misspecification : Meister (2004) proves that the mean integrated squared error of such estimators can grow to infinity when the noise distribution is misspecified.

Robustness to the assumptions on the noise

• Misspecification : Meister (2004) proves that the mean integrated squared error of such estimators can grow to infinity when the noise distribution is misspecified.

• Knowledge of the characteristic function of the noise : reduction to knowledge on a compact interval (Meister 2007). Dimension 1.
$\mathcal{F}$: class of compactly supported densities with Sobolev regularity β.
Characteristic function of the noise known on $[-\nu, \nu]$ and lower bounded by μ.

$$\text{Then } R_n(\mathcal{F}) \text{ is of order } \left(\frac{\log n}{\log\log n}\right)^{-2\beta}.$$

Robustness to the assumptions on the noise

• Misspecification : Meister (2004) proves that the mean integrated squared error of such estimators can grow to infinity when the noise distribution is misspecified.

• Knowledge of the characteristic function of the noise : reduction to knowledge on a compact interval (Meister 2007). Dimension 1. $\mathcal{F}$: class of compactly supported densities with Sobolev regularity β. Characteristic function of the noise known on $[-\nu, \nu]$ and lower bounded by μ.

$$\text{Then } R_n(\mathcal{F}) \text{ is of order } \left(\frac{\log n}{\log\log n}\right)^{-2\beta}.$$

$\rightarrow$ Is it possible to get rid of the knowledge of the noise ?

Joint works with

L. Lehéricy

S. Le Corff

J. Capitao-Miniconi

First basic question : identifiability

Is the distribution of $\mathbf{X}$ uniquely determined by the distribution of $\mathbf{Y}$? That is :

Can $\mathbf{X} + \varepsilon$ have the same distribution of $\mathbf{X}' + \varepsilon'$ with $\mathbf{X}'$ having a different distribution than $\mathbf{X}$?

What assumptions to get identifiability (up to translation) ?

First basic question : identifiability

Is the distribution of $\mathbf{X}$ uniquely determined by the distribution of $\mathbf{Y}$? That is :

Can $\mathbf{X} + \varepsilon$ have the same distribution of $\mathbf{X}' + \varepsilon'$ with $\mathbf{X}'$ having a different distribution than $\mathbf{X}$?

What assumptions to get identifiability (up to translation) ?

Good news : no assumption on the noise and weak structure assumptions on the signal allow identifiability

- Multidimensional observations : $\mathbf{Y}$, $\mathbf{X}$, ε are in $\mathbb{R}^D$, $D \geq 2$
- No distributional assumption on the noise, except that it has independent components
- The distribution of the signal has light tails
- Some dependency assumption on the components of the signal

Identifiability theorem

- With $d_1 \geq 1$, $d_2 \geq 1$ $(d_1 + d_2 = D)$:

$$\mathbf{Y} = \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} + \begin{pmatrix} \varepsilon^{(1)} \\ \varepsilon^{(2)} \end{pmatrix} = \mathbf{X} + \varepsilon\,.$$

Identifiability theorem

- With $d_1 \geq 1$, $d_2 \geq 1$ $(d_1 + d_2 = D)$:

$$\mathbf{Y} = \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} + \begin{pmatrix} \varepsilon^{(1)} \\ \varepsilon^{(2)} \end{pmatrix} = \mathbf{X} + \varepsilon \,.$$

- $\varepsilon^{(1)}$ is independent of $\varepsilon^{(2)}$.

Notations : $\mathbb{P}_{G,Q}$ is the distribution of $\mathbf{Y}$ when $\mathbf{X}$ has distribution G and for $i \in \{1,2\}$, $\varepsilon^{(i)}$ has distribution $Q^{(i)}$, with $Q = Q^{(1)} \otimes Q^{(2)}$.

$$\forall (z_1, z_2) \in \mathbb{C}^{d_1} \times \mathbb{C}^{d_2},\ \Phi_G(z_1, z_2) = \int \exp\left(i z_1^\top x_1 + i z_2^\top x_2 \right) G(\mathrm{d}x_1, \mathrm{d}x_2).$$

- Assumption H(ρ) : There exist $A > 0, B > 0$ such that for all $\lambda \in \mathbb{R}^D$, $\Phi_G(i\lambda) \leq A \exp\left(B \|\lambda\|^\rho\right)$.

When H(ρ) holds, Φ_G is a multivariate analytic function

.

Identifiability theorem

- Assumption HD :

For any $z_1 \in \mathbb{C}^{d_1}$, $z \mapsto \Phi_G(z_1, z)$ is not identically zero and for any $z_2 \in \mathbb{C}^{d_2}$, $z \mapsto \Phi_G(z, z_2)$ is not identically zero.

Identifiability theorem

• Assumption HD :
For any $z_1 \in \mathbb{C}^{d_1}$, $z \mapsto \Phi_G(z_1, z)$ is not identically zero and
for any $z_2 \in \mathbb{C}^{d_2}$, $z \mapsto \Phi_G(z, z_2)$ is not identically zero.

When $X^{(1)}$ and $X^{(2)}$ are not deterministic, if H(ρ), $\rho < 2$, and HD hold, $X^{(1)}$ and $X^{(2)}$ can not be independent random variables

Identifiability theorem

• Assumption HD :
For any $z_1 \in \mathbb{C}^{d_1}$, $z \mapsto \Phi_G(z_1, z)$ is not identically zero and for any $z_2 \in \mathbb{C}^{d_2}$, $z \mapsto \Phi_G(z, z_2)$ is not identically zero.

When $X^{(1)}$ and $X^{(2)}$ are not deterministic, if H(ρ), $\rho < 2$, and HD hold, $X^{(1)}$ and $X^{(2)}$ can not be independent random variables

Let $U^{(1)}$ and $U^{(2)}$ be two independent random variables in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$ respectively that satisfy H(ρ) for some $\rho \geq 1$, then $U = \begin{pmatrix} U^{(1)} \\ U^{(2)} \end{pmatrix}$ satisfies HD if and only if $U^{(1)}$ and $U^{(2)}$ are Gaussian or Dirac random variables.
From $\Phi_U(z_1, z_2) = \Phi_{U^{(1)}}(z_1)\Phi_{U^{(2)}}(z_2)$ and Hadamard+Marckinciewitz.

Identifiability theorem

Theorem (EG, S. Le Corff, L. Lehéricy (AOS 2022))

Assume that G and $\widetilde{G}$ are probability distributions on $\mathbb{R}^D$ which satisfy assumption $H(\rho)$ for some $\rho < 2$ and which satisfy HD. Then, $\mathbb{P}_{G,Q} = \mathbb{P}_{\widetilde{G},\widetilde{Q}}$ implies that $G = \widetilde{G}$ and $Q = \widetilde{Q}$ up to translation.

When does (HD) hold ?

- Repeated measurements

$$\mathbf{Y} = \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} = \begin{pmatrix} X^{(1)} \\ X^{(1)} \end{pmatrix} + \begin{pmatrix} \varepsilon^{(1)} \\ \varepsilon^{(2)} \end{pmatrix} = \mathbf{X} + \varepsilon \,.$$

Assumption HD always holds.

$$\Phi_X(z_1, z_2) = \Phi_{X^{(1)}}(z_1 + z_2).$$

The particular case of repeated observations : Delaigle et al. (2008) ; Meister (2010).

When does (HD) hold ?

- Errors in variable regression models

$$\mathbf{Y} = \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} = \begin{pmatrix} X^{(1)} \\ g(X^{(1)}) \end{pmatrix} + \begin{pmatrix} \varepsilon^{(1)} \\ \varepsilon^{(2)} \end{pmatrix} = \mathbf{X} + \varepsilon\,,$$

where $g : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$.

Asume $H(\rho)$. Then Assumption HD holds as soon as g is one-to-one on a subset of the support of X_1 with positive probability

$$\Phi_X(z_1, z) = E\left[E\left(\exp\left(iz_1^\top X^{(1)}\right)|g(X^{(1)})\right)\exp\left(iz^\top X^{(1)}\right)\right].$$

When does (HD) hold ?

IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 45, NO. 12, DECEMBER 1997

Source Separation when the Input Sources Are Discrete or Have Constant Modulus

Fabrice Gamboa and Elisabeth Gassiat

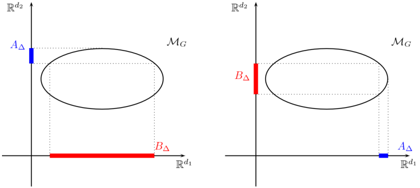
- Noisy non linear ICA (H. Halva, S. Le Corff, L. Lehéricy, J. So, Y. Zhu, EG, A. Hyvarinen (NeurIPS 2021))
  There exists an unknown integer $q \geq 1$, an unknown function $h : \mathbb{R}^q \to \mathbb{R}^D$, and a sequence of random vectors $\mathbf{S}_i \in \mathbb{R}^q$ with independent coordinates such that

$$\mathbf{Y}_i = h(\mathbf{S}_i) + \varepsilon_i \, .$$

When does HD hold ? Geometrical condition

(H1) For any $\Delta > 0$, there exists $A_\Delta \subset \mathbb{R}^{d_2}$ and $B_\Delta \subset \mathbb{R}^{d_1}$ such that $P(X^{(2)} \in A_\Delta) > 0$, $\lim_{\Delta \to 0} Diam(B_\Delta) = 0$ and $P(X^{(1)} \in B_\Delta \,|\, X^{(2)} \in A_\Delta) = 1$.

(H2) For any $\Delta > 0$, there exists $A_\Delta \subset \mathbb{R}^{d_1}$ and $B_\Delta \subset \mathbb{R}^{d_2}$ such that $P(X^{(1)} \in A_\Delta) > 0$, $\lim_{\Delta \to 0} Diam(B_\Delta) = 0$ and $P(X^{(2)} \in B_\Delta \,|\, X^{(1)} \in A_\Delta) = 1$.

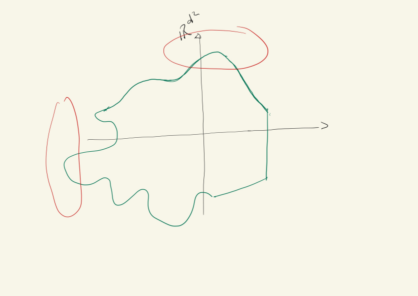


Theorem (J. Capitao-Miniconi, EG, L. Lehéricy)

Assume that the distribution of X satisfies H(ρ), (H1) and (H2). Then X satisfies H(ρ) and HD.

When does HD hold ? Examples of supports

- Any closed Euclidian sphere
- Any strictly convex compact set in $\mathbb{R}^D$
- The boundary of any strictly convex compact set
- Sets having strictly convex extremities in two 'components' or boundaries of such sets



Estimation of the characteristic function of the signal

From the Identifiability proof

Assume Φ_G and ϕ satisfy H(ρ) and (HD). Then for any $\nu > 0$,

$$\int_{[-\nu,\nu]^D} |\phi(t_1,t_2)\Phi_G(t_1,0)\Phi_G(0,t_2) - \Phi_G(t_1,t_2)\phi(t_1,0)\phi(0,t_2)|^2 \,\mathrm{d}t_1\mathrm{d}t_2$$

$$= 0 \iff \phi = \Phi_G .$$

Notice that with

$$M(\phi) = \int_{[-\nu,\nu]^D} |\phi(t_1,t_2)\Phi_Y(t_1,0)\Phi_G(0,t_2) - \Phi_Y(t_1,t_2)\phi(t_1,0)\phi(0,t_2)|^2 \mathrm{d}t_1\mathrm{d}t_2 \,,$$

$$M(\phi) = \int |\phi(t_1,t_2)\Phi_G(t_1,0)\Phi_G(0,t_2) - \Phi_G(t_1,t_2)\phi(t_1,0)\phi(0,t_2)|^2 |\phi_{\varepsilon^{(1)}}(t_1)\phi_{\varepsilon^{(2)}}(t_2)|^2 \mathrm{d}t_1\mathrm{d}t_2 .$$

Estimation of the characteristic function of the signal

We get an estimator $\widehat{\phi}_n$ of Φ_G by minimizing M_n over a set of functions satisfying H(ρ) and HD, with

$$M_n(\phi) = \int_{[-\nu_{\text{est}},\nu_{\text{est}}]^{d_1+d_2}} |\phi(t_1,t_2)\widetilde{\phi}_n(t_1,0)\widetilde{\phi}_n(0,t_2) - \widetilde{\phi}_n(t_1,t_2)\phi(t_1,0)\phi(0,t_2)|^2 \mathrm{d}t_1 \mathrm{d}t_2\,,$$

where $\widetilde{\phi}_n(t_1,t_2) = \frac{1}{n}\sum_{\ell=1}^n \mathrm{e}^{it_1^\top Y_\ell^{(1)} + it_2^\top Y_\ell^{(2)}}$.

Estimation of the characteristic function of the signal

We get an estimator $\widehat{\phi}_n$ of Φ_G by minimizing M_n over a set of functions satisfying H(ρ) and HD, with

$$M_n(\phi) = \int_{[-\nu_{\mathrm{est}},\nu_{\mathrm{est}}]^{d_1+d_2}} |\phi(t_1,t_2)\widetilde{\phi}_n(t_1,0)\widetilde{\phi}_n(0,t_2) - \widetilde{\phi}_n(t_1,t_2)\phi(t_1,0)\phi(0,t_2)|^2 \mathrm{d}t_1\mathrm{d}t_2\,,$$

where $\widetilde{\phi}_n(t_1,t_2) = \frac{1}{n}\sum_{\ell=1}^n \mathrm{e}^{it_1^\top Y_\ell^{(1)} + it_2^\top Y_\ell^{(2)}}$.

Proposition (EG, S. Le Corff, L. Lehéricy (AOS 2022))

(Simplified). For any $\delta > 0$, for any $n \geq n_0$, with high (controlled) probability,

$$\int_{[-\nu,\nu]^D} |\widehat{\phi}_n(t) - \Phi_G(t)|^2 \mathrm{d}t \leq C\left(\frac{1}{n^{1+\delta}}\right).$$

Estimation : basic ideas

Estimation of the characteristic function at almost parametric rate in $L^2([-\nu,\nu]^D)$ for any D.

$\rightarrow$ What can be written as a function of $(\Phi_G(z))_{|z|\leq M}$ can be estimated at almost parametric rate in $L^2([-\nu,\nu]^D)$ for any D. Example : radius of a sphere (EG, J. Capitao-Miniconi, EJS to appear).

$\rightarrow$ using that Φ_G is multivariate analytic

- Density estimation : Fourier inversion of a the truncation of a polynomial expansion of ϕ_n ;
- Support estimation : upper level set of the estimator of the density of a smoothed version of G ;
- Distribution estimation : restrict the estimator of the density of a smoothed version of G to the estimated support.

Density estimation

Model : X has a density

We truncate the polynomial expansion of $\widehat{\phi}_n$ at some degree $m_{\rho,n}$ (to be chosen) and

$$\widehat{f}_\rho(x) = \frac{1}{(2\pi)^d} \int_{[-\omega_{\rho,n},\omega_{\rho,n}]^{d_1+d_2}} \exp(-it^\top x) \left(T_{m_{\rho,n}} \widehat{\phi}_n \right)(t) \mathrm{d}t \,,$$

for some $\omega_{\rho,n} > 0$ to be chosen.

Minimax rate : density estimation

Fix

$$m_{\rho,n} = \left\lfloor \frac{\rho}{8} \left(\frac{\log n}{\log \log (n/4)} \right) \right\rfloor \text{ and } \omega_{\rho,n} = c_\rho m_{\rho,n}^{1/\rho}/S$$

for some constant $c_\rho \leq \nu_{\text{est}} \wedge \frac{2}{\rho} \exp(-(3d+5)/2)$.

Theorem (EG, S. Le Corff, L. Lehéricy)

The rate of convergence of $\mathbb{E}\|\widehat{f}_\rho - f\|_2^2$ *is* $\left(\frac{\log n}{\log \log n}\right)^{-2\beta/\rho}$ *uniformly for* f *in a Sobolev ball of regularity* β.

Minimax rate : density estimation

Fix

$$m_{\rho,n} = \left\lfloor \frac{\rho}{8}\left(\frac{\log n}{\log\log(n/4)}\right)\right\rfloor \text{ and } \omega_{\rho,n} = c_\rho m_{\rho,n}^{1/\rho}/S$$

for some constant $c_\rho \leq \nu_{\text{est}} \wedge \frac{2}{\rho}\exp(-(3d+5)/2)$.

Theorem (EG, S. Le Corff, L. Lehéricy)

The rate of convergence of $\mathbb{E}\|\widehat{f}_\rho - f\|_2^2$ is $\left(\frac{\log n}{\log\log n}\right)^{-2\beta/\rho}$ uniformly for f in a Sobolev ball of regularity β.

- Automatically adaptive in β.
- Adaptivity in ρ by using a Lepski's method $\to \widehat{\rho}$.
- Matching lower bound.
- (Almost) same rate as when the characteristic function of the noise is known on an interval

Estimation of the support

The i.i.d. X_i have distribution G with support $\mathcal{M}$ (of possibly low unknown dimension).
Let ψ_A a (well chosen) kernel, h a window, and $\overline{g}$ the density of the smoothed distribution :

$$\overline{g} = G \star \Psi_{A,h}.$$

Estimate $\overline{g}$ with

$$\forall y \in \mathbb{R}^D,\ \widehat{g}_{n,\rho}(y) = \left(\frac{1}{2\pi}\right)^D \int e^{-it^\intercal y} \mathcal{F}[\psi_A](ht)\ T_{m_\rho} \widehat{\Phi}_{n,\rho}(t) dt.$$

Finally, define an estimator of the support of the signal as the upper level set

$$\widehat{\mathcal{M}}_\rho = \left\{ y \in \mathbb{R}^D \mid \widehat{g}_{n,\rho}(y) > \lambda_{n,\rho} \right\}$$

for some $\lambda_{n,\rho}$.

Estimation of the support

The loss function : truncated Hausdorff.

Since we allow the support to be a non-compact set, we fix $\mathcal{K}$ a compact subset of $\mathbb{R}^D$ and for any S_1, S_2 subsets of $\mathbb{R}^D$, with d_H the Hausdorff distance,

$$H_{\mathcal{K}}(S_1, S_2) = d_H(S_1 \cap \mathcal{K}, S_2 \cap \mathcal{K}).$$

(a, d)-standard distributions.

For any positive constants a, d and r_0, we define $St_{\mathcal{K}}(a, d, r_0)$ as the set of positive measures G such that :

$$\forall x \in \mathcal{K},\ \forall r \leq r_0,\ G(B(x, r)) \geq ar^d.$$

Estimation of the support : upper bound

Theorem (J. Capitao-Miniconi, EG, L. Lehéricy)

Let $\rho \in [1,2)$, $a > 0$, $d \leq D$, $r_0 > 0$. It is possible to choose m_ρ, h and $\lambda_{n,\rho}$ depending whether $d < D$ or $d = D$ such that

$$\sup_{\substack{G \in St_{\mathcal{K}}(a,d,r_0) \cap \mathcal{L}(\rho,\mathcal{H}) \\ Q \in \mathcal{Q}^{(D)}(\nu, c(\nu), E)}} \mathbb{E}_{(G*Q)^{\otimes n}}[H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}}_\rho)] \leq C \frac{\log(\log(n))^{\frac{1}{\rho} + \frac{A+1}{A}}}{\log(n)^{\frac{1}{\rho}}}.$$

- The rate deteriorates for signals having distributions with heavier tails.
- Not needed to know the intrinsic dimension of the support.
- Adaptation to unknown ρ by model selection.
- Almost matching lower bound.

Estimation of the support : lower bound

Theorem (J. Capitao-Miniconi, EG, L. Lehéricy)

For any $\rho \in (1,2)$, $S > 0$, $a > 0$, $d \geq 1$, $r_0 > 0$, *there exists a closed set* $\mathcal{H}^\star$ *such that, for all* $n \geq n_0$,

$$\inf_{\widehat{\mathcal{M}}} \sup_{\substack{G \in St_{\mathcal{K}}(a,d,r_0) \cap \mathcal{L}(\rho, \mathcal{H}^\star) \\ Q \in \mathcal{Q}^{(D)}(\nu, c(\nu), E)}} \mathbb{E}_{(G*Q)^{\otimes n}}[H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}})] \geq \frac{C}{\log(n)^{\frac{1}{\rho}}},$$

and for any $\delta \in (0,1)$, *for* $n \geq n_\delta$,

$$\inf_{\widehat{\mathcal{M}}} \sup_{\substack{G \in St_{\mathcal{K}}(a,d,r_0) \cap \mathcal{L}(1, \mathcal{H}^\star) \\ Q \in \mathcal{Q}^{(D)}(\nu, c(\nu), E)}} \mathbb{E}_{(G*Q)^{\otimes n}}[H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}})] \geq \frac{C}{\log(n)^{1+\delta}}.$$

The proof uses the two-points method : the idea is to find two distributions having supports as far as possible in the $H_{\mathcal{K}}$-loss, and a noise such that the joint distributions of the observations have total variation distance upper bounded by some $\eta < 1$.

Estimation of the distribution
with compact support

G has compact support $\mathcal{M}_G$: $\rho = 1$.

Fix some $\eta > 0$ and define $\widehat{\mathcal{M}}^\eta$ the η-enlargment of $\widehat{\mathcal{M}}$.

Define $\widehat{P}_{n,\eta}$, for any $\mathcal{O}$ borelian set of $\mathbb{R}^D$

$$\widehat{P}_{n,\eta}(\mathcal{O}) = \frac{1}{\int_{(\widehat{\mathcal{M}})_{\eta^\star}} \widehat{g}^+_{n,1}(y)dy} \int_{\mathcal{O}\cap(\widehat{\mathcal{M}})_{\eta^\star}} \widehat{g}^+_{n,1}(y)dy,$$

where $\widehat{g}^+_{n,1} = \max\{0, \widehat{g}_{n,1}\}$.

Estimation of the distribution

with compact support

Theorem (J. Capitao-Miniconi, EG, L. Lehéricy)

For any $S > 0$, $a > 0$, $d \geq 1$, $r_0 > 0$, $n \geq n_0$,

$$\sup_{\substack{G \in St_{\mathcal{K}}(a,d,r_0) \cap \mathcal{L}(1,\mathcal{H}^\star) \\ Q \in \mathcal{Q}^{(D)}(\nu, c(\nu), E)}} \mathbb{E}_{(G*Q)^{\otimes n}}[W_2(G, \widehat{P}_{n,\eta})] \leq \frac{C \log \log n}{\log n}.$$

Moreover there exists a closed set $\mathcal{H}^\star$ *such that for any* $\delta \in (0,1)$, *for* $n \geq n_\delta$,

$$\inf_{\widehat{P}} \sup_{\substack{G \in St_{\mathcal{K}}(a,d,r_0) \cap \mathcal{L}(1,\mathcal{H}^\star) \\ Q \in \mathcal{Q}^{(D)}(\nu, c(\nu), E)}} \mathbb{E}_{(G*Q)^{\otimes n}}[W_2(G, \widehat{P})] \geq \frac{c}{(\log n)^{1+\delta}}.$$

Take-home message

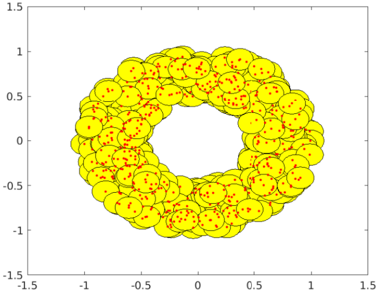
- Deconvolution is possible for multivariate signals without any knowledge of the noise distribution, under very weak assumptions.
- Estimation of a density, minimax risk.
- Simple geometric assumptions on the support of the signal allow to apply the theory.
- Estimation of the support in truncated Hausdorff risk : upper bounds, almost matching lower bounds ; rate in a power of $\log n$ (power depending on the tail of the signal distribution).
- Estimation of the distribution and minimax rates for the Wasserstein's risk ;
- (almost parametric rate for the estimation of the radius of a sphere).

Happy birthday brother !

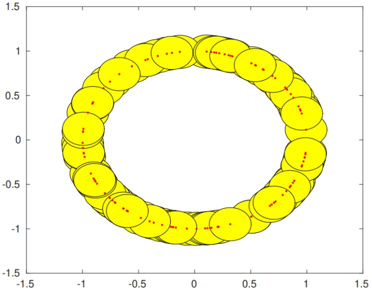
Manifold learning : some ideas (no noise)

Observations : points.
Aim : recover the (low dimensional) (nonlinear) support.

Balls centered on the observations.



Points on $B(0,1) \cap \mathcal{B}(0,0.5)^c$



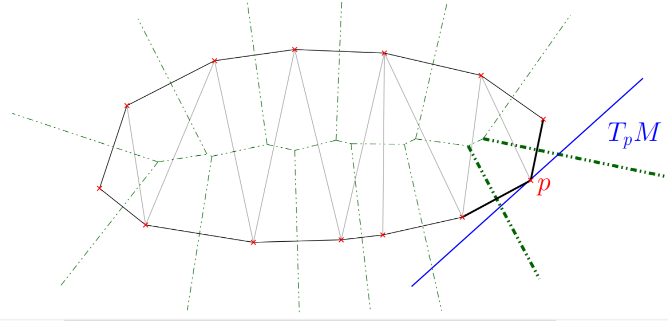
Points on $S(0,1)$

Manifold learning : some ideas (no noise)

Observations : points.

Aim : recover the (low dimensional) (nonlinear) support.

Tangential Delaunay Complex. Local PCA.

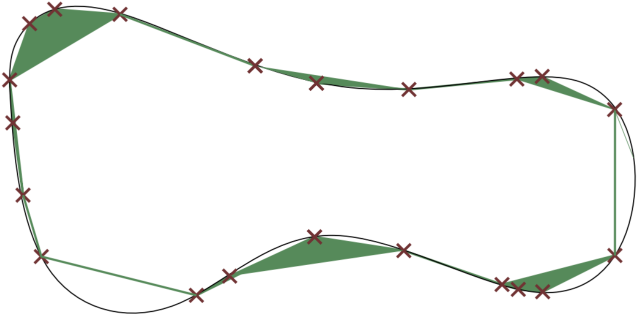


Manifold learning : some ideas (no noise)

Observations : points.

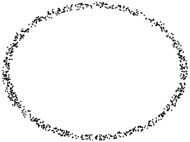
Aim : recover the (low dimensional) (nonlinear) support.

t-convex hull..

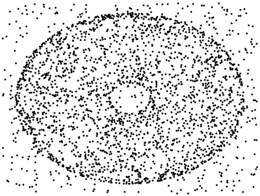


Noisy data. What happens with noise ? Geometric ideas

Small noise : stability of Tangential Delaunay Complex. Local PCA



Clutter noise : points $\sim (1-\pi)G + \pi U$



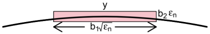


FIG. 2. *Given a manifold M and a point $y \in M$, $S_M(y)$ is a slab, centered at y, with size $O(\sqrt{\varepsilon_n})$ in the d directions corresponding to the tangent space $T_y M$ and size $O(\varepsilon_n)$ in the $D-d$ normal directions.*

Uniform noise around the shape

Additive noise : stochastic ideas

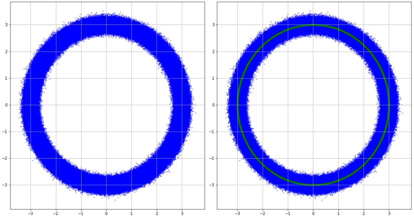
Data $Y_1, \ldots, Y_n$ are i.i.d. (independent and identically distributed)

The model is with additive independent noise :

$$\mathbf{Y} = \mathbf{X} + \varepsilon \, ,$$

$\mathbf{X}$ is the non-noisy variable of interest and ε is the noise,
$\mathbf{X}$ and ε are independent random variables.

Examples :



Additive noise : examples