# Free deconvolution- A tribute for Elisabeth work on deconvolution

Fabrice Gamboa

Work in progress with: Reda Chhaibi, Slim Kammoun and Mauricio Velasco

Elisabeth Gassiat - a path in modern statistics

June 2023

ANITI Université Fédérale Toulouse Midi-Pyrénées

INSTITUT de MATHEMATIQUES de TOULOUSE

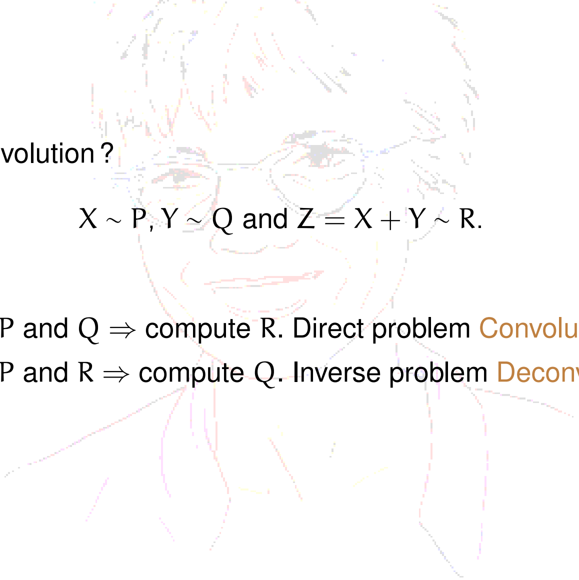
# Agenda

1. One of Elisabeth's favourite themes : Deconvolution
2. Sparsity and discrete mixtures
3. Free multiplicative deconvolution
4. RMT for Large Covariance Matrices
   - Brute force SVD is wrong in large dimension for spectral recovery
   - Free deconvolution and the good transform
   - Our method

# Overview

1. One of Elisabeth's favourite themes : Deconvolution
2. Sparsity and discrete mixtures
3. Free multiplicative deconvolution
4. RMT for Large Covariance Matrices

# One of Elisabeth's favourite themes : Blind deconvolution



What is deconvolution ?

$$X \sim P, Y \sim Q \text{ and } Z = X + Y \sim R.$$

- Knowing $P$ and $Q \Rightarrow$ compute $R$. Direct problem Convolution
- Knowing $P$ and $R \Rightarrow$ compute $Q$. Inverse problem Deconvolution

# Once upon the time, blind deconvolution

## Déconvolution aveugle

par Elisabeth Gassiat

Thèse de doctorat en Mathématiques

Sous la direction de Didier Dacunha-Castelle.

Soutenue en 1989
à Paris 11 , en partenariat avec Université de Paris-Sud. Faculté des sciences d'Orsay (Essonne) (autre partenaire) .

Description en français | **Description en anglais**

### Titre traduit

Blind deconvolution

### Résumé

Considering a signal X which is a process of random variables identically independently distributed, and the signal Y obtained by filtering X through a linear system s, we study the estimation of s from the observation of y in the following semi-parametric situation the law of X is unknown and non Gaussian, and s has an inverse of convolution with finite length. We need no assumption on the phase of the system, i. E. On the causality or non causality of s. We propose an estimation by maximum objective. The estimates are consistent and asymptotically Gaussian this result is still available what-ever the dimension of the index space of the series is. We study the asymptotic efficiency of the estimate and, in the causal case, we compare it to the usual minimum square estimates. The output y being an autoregressive field, we propose a consis- tent method of identification of the order of the model. We study different types of robustness robustness to underparametrization, robustness to additive noise on the observations. We also inves tigate the case where the law of X has infinite moments, and we show that, for "standardized cumulants" as objectives, and under assumptions which are in particular verified for laws in the attraction demains of stable laws, the obtained estimates are still consistent, and the speed of convergence is, in the causal case, better than for laws with finite variance.

# Elisabeth Ph. D main results

*Ann. Inst. Henri Poincaré,*
Vol. 26, n° 1, 1990, p. 181-205. *Probabilités et Statistiques*

**Estimation semi-paramétrique d'un modèle autorégressif stationnaire multiindice non nécessairement causal**

par

**Élisabeth GASSIAT**
Université de Paris-Sud, U.A. n° 743, C.N.R.S., Statistique appliquée, Mathématiques, bat. n° 425, 91405 Orsay Cedex, France

RÉSUMÉ. — Nous étudions le problème d'estimation du paramètre $\theta$ d'un modèle de processus autorégressif stationnaire multiindice dans le cas où l'on ne fait sur $\theta$ aucune hypothèse. La loi du bruit est supposée inconnue (déconvolution aveugle). En particulier on ne suppose pas que le bruit blanc associé soit l'innovation (modèle non nécessairement causal). Nous proposons une méthode d'estimation qui résoud complètement le problème si la loi du bruit blanc n'est pas gaussienne (dans ce cas le problème est sans solution). Nous étudions la consistance et la loi asymptotique des estimateurs ainsi que leur robustesse. Nous proposons une estimation de l'ordre du modèle.

JOURNAL OF MULTIVARIATE ANALYSIS **32**, 161–170 (1990)

Semi-parametric Estimation of a Stationary, Non-necessary Causal $AR(P)$ Process with Infinite Variance

ELISABETH GASSIAT
*Université de Paris-Sud, 91405 Orsay Cédex, France*

*Communicated by the Editors*

We study the estimation problem of the parameter of a stationary $AR(p)$ process with infinite variance when there is no assumption on the causality of the model. We propose consistent estimates. In the causal case, we obtain a speed of convergence. 

# Elisabeth Ph. D main results

Frame in a nutshell :

- $(Z_n)_{n \in \mathbb{Z}^d}$ observed A.R. built on an i.i.d. non Gaussian sequence $(X_n)_{n \in \mathbb{Z}^d}$,
- $b^* = (b_j^*)_{j \in S}$ the unknown weight vector ($|S| = m < \infty$),
- $Y_n(b^*) := \sum_{j \in S} b_j^* Z_{n-j}$,

$$Z_n = Y_n(b^*) + X_n, \; n \in \mathbb{Z}^d$$

Observing only $Z_1, \ldots Z_n$ recover both $b^*$ and $X_1, \ldots X_n$
$\Rightarrow$Blind deconvolution (much more greedy than the deconvolution ! !)

# Elisabeth Ph. D main results

Results in a nutshell :

- Build $\widehat{b}$ as an $M$-estimator of $b^*$,
- The good $M$ functions are the ones decreasing by convolution (typically normalized cumulants),
- Under very weak assumptions (even no variance for $X$ ! !) convergence and asymptotic rates for $\widehat{b}$

# One step beyond : optimality

*The Annals of Statistics*
1993, Vol. 21, No. 4, 2022–2042

**ADAPTIVE ESTIMATION IN NONCAUSAL STATIONARY AR PROCESSES**

BY E. GASSIAT

*Université Paris-Sud*

We consider the estimation problem of the parameter $b$ of a stationary AR($p$) process without any of the usual causality assumptions. The aim of the paper is to derive asymptotic minimax bounds for estimators of $b$. When the distribution of the noise is known, we show LAN properties of the model and derive locally asymptotically minimax (LAM) estimators. The most important results are about the case of unknown distribution: The main result shows that, if one uses the usual parametrization, these bounds depend heavily on the causality or the noncausality of the process, so that adaptive efficient estimation is impossible in the noncausal situation: The scaling factor is shown to give the hardest one-dimensional subproblem, and an unusual scaling is exhibited that could lead to adaptive efficient estimation of the rescaled parameter even in the noncausal case.

# One step beyond : optimality

Results in a nutshell :

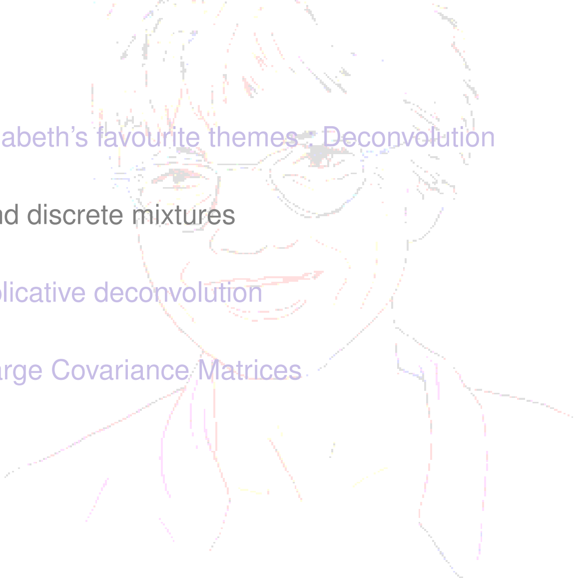
- For the estimation of $b^*$ under assumptions on the density of $X$ the model is LAN

$$\log\left(\frac{dL_{b^*+\frac{h}{\sqrt{n}}}}{dL_{b^*}}\right) = \langle h, \Delta_n\rangle - \frac{1}{2}h^T I_{b^*} h + o_p(1),$$

  with $\Delta_n \overset{\mathcal{L}}{\Rightarrow} \mathcal{N}_m(0, I_{b^*})$

- First consequence : lower bound for the asymptotic variance of a regular estimator ($\widehat{b}$ is regular).
- $\widehat{b}$ may be modified using one step Newton method to built asymptotic optimal $\widehat{\widehat{b}}$

# Overview



1. One of Elisabeth's favourite themes : Deconvolution
2. Sparsity and discrete mixtures
3. Free multiplicative deconvolution
4. RMT for Large Covariance Matrices

# Uniqueness and stability in the generalized moment problem

C. R. Acad. Sci. Paris, t. 310, Série I, p. 41-44, 1990 41

Analyse mathématique/*Mathematical Analysis*

## Problèmes des moments et concentration de mesure

Elisabeth GASSIAT

***Résumé*** – Nous nous intéressons au problème des moments au voisinage des points singuliers. Nous mettons en évidence des hypothèses portant sur les fonctions de moments sous lesquelles : nous étudions tout d'abord l'information apportée par les moments sur la concentration de la mesure cherchée, et nous montrons que le diamètre de Prokhorov de l'ensemble des mesures qui sont solution du problème des moments tend vers 0 au voisinage des points singuliers. Ces hypothèses sont vérifiées lorsque les moments sont définis par des T-systèmes. Nous indiquons des applications en traitement du signal.

## The moment problem and the concentration of measures

***Abstract*** – *We study the moment problem for constraints in the vicinity of singular points. We indicate assumptions under which: we first study the information on the concentration of the measure given by its moments and we then show that the Prokhorov diameter of the convex set of the measures solution of the moment problem tends to 0, when the constraint tends to singular points. These assumptions hold for moments defined by T-systems. We give some ideas of applications in signal processing.*

# Uniqueness and stability in the Hausdorff moment problem

- $\nu^*(dx) := \sum_{j=1}^{l} p_j^* \delta_{x_j^*}(dx) \in \mathcal{P}(]0,1[)$ is completely determined by its $2l$ first algebraic moments

$$m_j(\nu^*) = \int_{]0,1[} x^j \nu^*(dx), \ j = 1 \dots 2l.$$

- Hankel Matrix : $H_{p+1}(\nu) = (m_{i+j-2}(\nu))_{1 \leqslant i,j \leqslant p+1}$, $\nu \in \mathcal{P}(]0,1[)$,
  - $H_{p+1}(\nu) \geqslant 0$
  - $\det H_{p+1}(\nu) = 0 \Leftrightarrow \nu$ is supported by less than $p+1$ points.
  - So that $\det H_{p+1}$ can be used as $M$-function for deconvolution if the noise is discretely supported!!

# Blind discrete deconvolution

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

**BLIND DECONVOLUTION OF DISCRETE LINEAR SYSTEMS**

BY F. GAMBOA AND E. GASSIAT

*Laboratoire de Statistiques Orsay and Université Paris-Nord, and Université d'Evry Val d'Essonne*

We study the blind deconvolution problem in the case where the input noise has a finite discrete support and the transfer linear system is not necessarily minimum phase. We propose a new family of estimators built using algebraic considerations. The estimates are consistent under very wide assumptions: The input signal need not be independently distributed; the cardinality of the finite support may be estimated simultaneously. We consider in particular AR systems: In this case, we prove that the estimator of the parameters is perfect a.s. with a finite number of observations.

# Extension to source separation and mixtures

3062 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

***Abstract*— In this paper, we present a new method for the source separation problem when some prior information on the input sources is available. More specifically, we study the situation where the distributions of the input signals are discrete or are concentrated on a circle. The method is based on easy properties of Hankel forms and on the divisibility of Gaussian distributions. In both situations, we prove that the estimator converges in absence of noise or if we know the first moments of the noise up to its scale. Moreover, in the absence of noise, the estimate converges with a finite number of observations.**

***Index Terms*— Estimation, Hankel forms, identification, independent sources, source separation.**

The problem of interest is the restitution of the signal $\mathrm{s}(t)$, $t = 1, \cdots, N$ on the basis of the observations $\mathrm{x}(t)$, $t = 1, \cdots, N$. This is the source separation problem. In absence of additive noise, it can be solved by estimating a decoupling matrix $\mathbf{B}$ to the observations, which may be seen as an estimator of some inverse matrix (up to scale and permutation) of $\mathbf{A}$. Different separation procedures have been proposed in the past few years, for instance, using higher order statistics; see [2], [5], [6], [10], [15], [17], [22], [24], [25], [29], and [31]. Maximum likelihood estimates are considered in [1] and [11], whereas adaptive solutions are studied in [9], [18],

*Bernoulli* **3**(3), 1997, 279–299

## The estimation of the order of a mixture model

DIDIER DACUNHA-CASTELLE[1] and ELISABETH GASSIAT[2*]

[1]*Equipe de Modélisation Stochastique et Statistique, Unité de Recherche associée au CNRS 743, Université Paris-Sud, 91405, Orsay Cédex, France*
[2]*Equipe d'Analyse et de Probabilités, Université d'Evry, 91025 Evry Cédex, France*

We propose a new method to estimate the number of different populations when a large sample of a mixture of these populations is observed. It is possible to define the number of different populations as the number of points in the support of the mixing distribution. For discrete distributions having a finite support, the number of support points can be characterized by Hankel matrices of the first algebraic moments, or Toeplitz matrices of the trigonometric moments. Namely, for one-dimensional distributions, the cardinality of the support may be proved to be the least integer such that the Hankel matrix (or the Toeplitz matrix) degenerates. Our estimator is based on this property. We first prove the convergence of the estimator, and then its exponential convergence under wide assumptions. The number of populations is not a priori bounded. Our method applies to a large number of models such as translation mixtures with known or unknown variance, scale mixtures, exponential families and various multivariate models. The method has an obvious computational advantage since it avoids any computation of estimates of the mixing parameters. Finally we give some numerical examples to illustrate the effectiveness of the method in the most popular cases.

*Keywords:* Hankel matrix; mixture models; order estimation; penalization

# Recent contribution using this sparsity paradigm

*The Annals of Statistics*
2021, Vol. 49, No. 3, 1779–1809
https://doi.org/10.1214/20-AOS2022


## SUPERMIX: SPARSE REGULARIZATION FOR MIXTURES

BY Y. DE CASTRO[1], S. GADAT[2], C. MARTEAU[3] AND C. MAUGIS-RABUSSEAU[4]

[1]*Université Lyon, École Centrale de Lyon, Institut Camille Jordan*, yohann.de-castro@ec-lyon.fr
[2]*Toulouse School of Economics, Université Toulouse 1 Capitole*, sebastien.gadat@tse-fr.eu
[3]*Université Lyon, Université Claude Bernard Lyon 1, Institut Camille Jordan*, marteau@math.univ-lyon1.fr
[4]*Institut de Mathématiques de Toulouse, Université de Toulouse*, maugis@insa-toulouse.fr

This paper investigates the statistical estimation of a discrete mixing measure $\mu^0$ involved in a kernel mixture model. Using some recent advances in $\ell_1$-regularization over the space of measures, we introduce a "data fitting and regularization" convex program for estimating $\mu^0$ in a grid-less manner from a sample of mixture law, this method is referred to as Beurling-LASSO.

Our contribution is two-fold: we derive a lower bound on the bandwidth of our data fitting term depending only on the support of $\mu^0$ and its so-called "minimum separation" to ensure quantitative support localization error bounds; and under a so-called "nondegenerate source condition" we derive a nonasymptotic support stability property. This latter shows that for a sufficiently large sample size $n$, our estimator has exactly as many weighted Dirac masses as the target $\mu^0$, converging in amplitude and localization towards the true ones. Finally, we also introduce some tractable algorithms for solving this convex program based on "Sliding Frank–Wolfe" or "Conic Particle Gradient Descent".

Statistical performances of this estimator are investigated designing a so-called "dual certificate", which is appropriate to our setting. Some classical situations as, for example, mixtures of super-smooth distributions (see, e.g., Gaussian distributions) or ordinary-smooth distributions (see, e.g., Laplace distributions), are discussed at the end of the paper.

# Recent contribution of Elisabeth on deconvolution

*The Annals of Statistics*
2022, Vol. 50, No. 1, 303–323
https://doi.org/10.1214/21-AOS2106

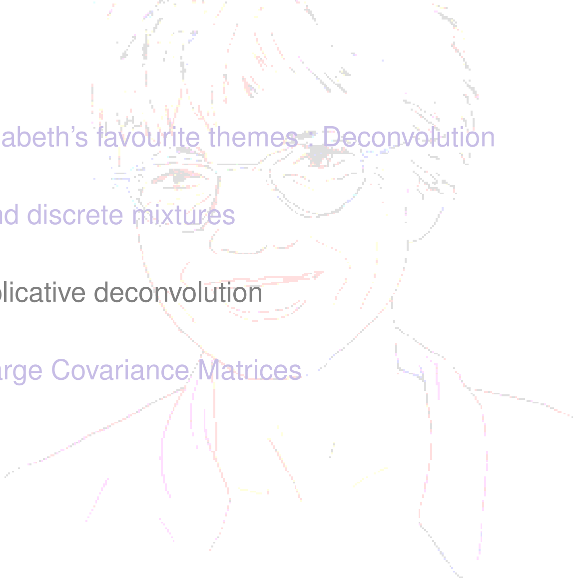

## DECONVOLUTION WITH UNKNOWN NOISE DISTRIBUTION IS POSSIBLE FOR MULTIVARIATE SIGNALS

BY ÉLISABETH GASSIAT[1], SYLVAIN LE CORFF[2] AND LUC LEHÉRICY[3]

[1]*Laboratoire de mathématiques d'Orsay, Université Paris-Saclay, CNRS, elisabeth.gassiat@universite-paris-saclay.fr*
[2]*Samovar, Télécom SudParis, département CITI, TIPIC, Institut Polytechnique de Paris, sylvain.le_corff@telecom-sudparis.eu*
[3]*Laboratoire J. A. Dieudonné, Université Côte d'Azur, CNRS, luc.lehericy@univ-cotedazur.fr*

This paper considers the deconvolution problem in the case where the target signal is multidimensional and no information is known about the noise distribution. More precisely, no assumption is made on the noise distribution and no samples are available to estimate it: the deconvolution problem is solved based only on observations of the corrupted signal. We establish the identifiability of the model up to translation when the signal has a Laplace transform with an exponential growth $\rho$ smaller than 2 and when it can be decomposed into two dependent components. Then we propose an estimator of the probability density function of the signal, which is consistent for any unknown noise distribution with finite variance. We also prove rates of convergence and, as the estimator depends on $\rho$ which is usually unknown, we propose a model selection procedure to obtain an adaptive estimator with the same rate of convergence as the estimator with a known tail parameter. This rate of convergence is known to be minimax when $\rho = 1$.

# Overview



# Classical convolutions

- X and Y real random variables,
- $Z = X + Y$ classical additive convolution
- XY classical multiplicative convolution

If

$$\mathbb{P}(X \in dx) := f(x)dx \quad \text{and} \quad \mathbb{P}(Y \in dy) := g(y)dy$$

then

$$\mathbb{P}(Z \in dz) = f *_a g(x)dx = \left(\int f(z-y)g(y)dy\right) dz\,, \text{ and } \varphi_Z = \varphi_X \times \varphi_Y$$

$$\mathbb{P}(XY \in dz) = f *_m g(x)dx = \left(\int f(z/y)g(y)dy\right) dz\,.$$

# Classical convolution with matrices

$A^{(n)} \in M_n(\mathbb{C})$

$$\mu_{A^{(n)}} := \frac{1}{n} \sum_{\lambda \in \mathrm{Spec}(A^{(n)})} \delta_\lambda \quad \text{(spectral measure of } A^{(n)}\text{)}$$

Two i.i.d. independent samples of positive random variables with laws $\mu_1$ and $\mu_2$ $X_1, X_2, X_3, \dots, X_n$ and $Y_1, Y_2, Y_3, \dots, Y_n$.
$A^{(n)} := \mathrm{diag}(X_1, \dots, X_n)$ and $B^{(n)} := \mathrm{diag}(Y_1, \dots, Y_n)$.

## Lemma (Obvious but illustrative)

*We have*

$$\lim_{n\to\infty} \mu_{A^{(n)}} = \mu_1 \,, \qquad \lim_{n\to\infty} \mu_{B^{(n)}} = \mu_2 \,,$$

*and also*

$$\lim_{n\to\infty} \mu_{A^{(n)}B^{(n)}} = \lim_{n\to\infty} \mu_{(B^{(n)})^{\frac{1}{2}} A^{(n)} (B^{(n)})^{\frac{1}{2}}} = \mu_1 *_m \mu_2 \,.$$

# Free multiplicative convolution

Two iid independent samples of positive random variables with laws $\mu_1$ and $\mu_2$ :

$$X_1, X_2, X_3, \ldots, X_n, \ldots \text{ and } Y_1, Y_2, Y_3, \ldots, Y_n, \ldots$$

$A^{(n)} := \mathrm{diag}(X_1, \ldots, X_n)$ and $B^{(n)} := U_n \,\mathrm{diag}(Y_1, \ldots, Y_n) U_n^*$,
$U_n$ Haar on the orthogonal group (independent of X and Y).

## Theorem

*Of course, we still have*

$$\lim_{n\to\infty} \mu_{A^{(n)}} = \mu_1 \, , \qquad \lim_{n\to\infty} \mu_{B^{(n)}} = \mu_2 \, ,$$

*but this time :*

$$\lim_{n\to\infty} \mu_{(B^{(n)})^{\frac{1}{2}} A^{(n)} (B^{(n)})^{\frac{1}{2}}} = \mu_1 \boxtimes \mu_2 \, .$$

*Here $\mu_1 \boxtimes \mu_2$ is a a deterministic measure which depends only on $\mu_1$ and $\mu_2$ called their free multiplicative convolution.*

# Overview

# Two pioneering papers

*The Annals of Statistics*
2012, Vol. 40, No. 2, 1024–1060
DOI: 10.1214/12-AOS989


## NONLINEAR SHRINKAGE ESTIMATION OF LARGE-DIMENSIONAL COVARIANCE MATRICES

BY OLIVIER LEDOIT AND MICHAEL WOLF[1]

*University of Zurich*

Many statistical applications require an estimate of a covariance matrix and/or its inverse. When the matrix dimension is large compared to the sample size, which happens frequently, the sample covariance matrix is known to perform poorly and may suffer from ill-conditioning. There already exists an extensive literature concerning improved estimators in such situations. In the absence of further knowledge about the structure of the true covariance matrix, the most successful approach so far, arguably, has been shrinkage estimation. Shrinking the sample covariance matrix to a multiple of the identity, by taking a weighted average of the two, turns out to be equivalent to linearly shrinking the sample eigenvalues to their grand mean, while retaining the sample eigenvectors. Our paper extends this approach by considering nonlinear transformations of the sample eigenvalues. We show how to construct an estimator that is asymptotically equivalent to an oracle estimator suggested in previous work. As demonstrated in extensive Monte Carlo simulations, the resulting *bona fide* estimator can result in sizeable improvements over the sample covariance matrix and also over linear shrinkage.

*The Annals of Statistics*
2008, Vol. 36, No. 6, 2757–2790
DOI: 10.1214/07-AOS581


## SPECTRUM ESTIMATION FOR LARGE DIMENSIONAL COVARIANCE MATRICES USING RANDOM MATRIX THEORY[1]

BY NOUREDDINE EL KAROUI

*University of California, Berkeley*

Estimating the eigenvalues of a population covariance matrix from a sample covariance matrix is a problem of fundamental importance in multivariate statistics; the eigenvalues of covariance matrices play a key role in many widely used techniques, in particular in principal component analysis (PCA). In many modern data analysis problems, statisticians are faced with large datasets where the sample size, $n$, is of the same order of magnitude as the number of variables $p$. Random matrix theory predicts that in this context, the eigenvalues of the sample covariance matrix are not good estimators of the eigenvalues of the population covariance.

We propose to use a fundamental result in random matrix theory, the Marčenko–Pastur equation, to better estimate the eigenvalues of large dimensional covariance matrices. The Marčenko–Pastur equation holds in very wide generality and under weak assumptions. The estimator we obtain can be thought of as "shrinking" in a nonlinear fashion the eigenvalues of the sample covariance matrix to estimate the population eigenvalues. Inspired by ideas of random matrix theory, we also suggest a change of point of view when thinking about estimation of high-dimensional vectors: we do not try to estimate directly the vectors but rather a probability measure that describes them. We think this is a theoretically more fruitful way to think about these problems.

Our estimator gives fast and good or very good results in extended simulations. Our algorithmic approach is based on convex optimization. We also show that the proposed estimator is consistent.

# Our work in progress

arXiv > math > arXiv:2305.05646

**Mathematics > Probability**

[Submitted on 9 May 2023]

## Estimation of large covariance matrices via free deconvolution: computational and statistical aspects

Reda Chhaibi, Fabrice Gamboa, Slim Kammoun, Mauricio Velasco

The estimation of large covariance matrices has a high dimensional bias. Correcting for this bias can be reformulated via the tool of Free Probability Theory as a free deconvolution. The goal of this work is a computational and statistical resolution of this problem. Our approach is based on complex-analytic methods methods to invert $S$-transforms. In particular, one needs transforms live and an efficient computational scheme.

Comments: v1: Preliminary version
Subjects: **Probability (math.PR)**; Statistics Theory (math.ST); Computation (stat.CO)
Cite as: arXiv:2305.05646 **[math.PR]**
(or arXiv:2305.05646v1 **[math.PR]** for this version)
https://doi.org/10.48550/arXiv.2305.05646

### Submission history

From: Reda Chhaibi [view email]
**[v1]** Tue, 9 May 2023 17:45:04 UTC (29 KB)

# Brute force SVD is wrong in large dimension

Consider a sequence of *centered* observables in $\mathbb{R}^d$, $X_1, X_2, \dots$. The empirical covariance

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^* = \frac{1}{n}\mathbb{X}\mathbb{X}^*,$$

where $\mathbb{X} = (X_1, \dots, X_n) \in M_{d,n}(\mathbb{R})$ is the matrix with columns given by the $X_i$'s.

Take the $X_i$ having iid components as well.

- If $d$ fixed and $n \to \infty$,

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{X}\mathbb{X}^* = I_d\ ,$$

  recovering the true covariance.
- If both $d = d_n, n \to \infty$, this is not true.

## Theorem (Marchenko-Pastur 1967)

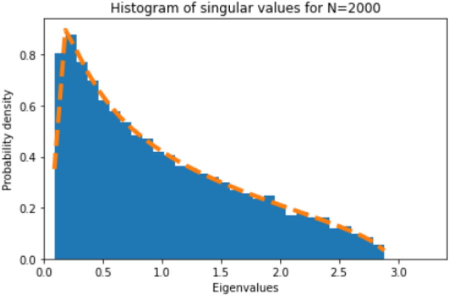
*Assume that* $d_n/n \to c < 1$. *Almost surely, as* $n \to \infty$, *we have the weak convergence of probability measures* $\lim_n \mu_{\frac{1}{n}\mathbb{X}\mathbb{X}^*} = MP_c \neq \delta_1$ *where*

$$MP_c(dx) := \mathbb{1}_{\{x \in [l,r]\}} \frac{\sqrt{(x-l)(r-x)}}{2\pi x} dx$$

*is the Marchenko-Pastur distribution,* $l = (1-\sqrt{c})^2$, $r = (1+\sqrt{c})^2$.

```
c = 1.0/2 # MP scale parameter
num_bins = 30
r = (1+np.sqrt(c))**2 #Right end
l = (1-np.sqrt(c))**2 #Left end

G = np.random.normal( size=( int(c*N),N) )
W = G.dot( G.transpose() )
W = W/N
diag, U = np.linalg.eig(W)
```



# The inverse problem

Assume

$$\mathbb{Z} = \mathbb{X}\mathbb{Y} \text{ with } \mathbb{X} = \Sigma^{\frac{1}{2}},$$

with $\Sigma \in M_d(\mathbb{R})$ true covariance, and $\mathbb{Y} \in M_{d,n}(\mathbb{R})$ with iid entries with distribution $\mathcal{N}(0, 1)$.

**Question**

*When $d/n \approx c > 0$, how to estimate the spectrum of $\Sigma$ ?*

More precisely, we will focuss on the case where the limit spectral measure is

$$\nu^* = \sum_{i=1}^{k} p_i^* \delta_{\sigma_i^*}$$

# Model

Who is $\nu^*$ ?

$$\lim_{n\to+\infty} \frac{1}{d_n} \mathrm{Tr}(\Sigma^r) = \int x^r \nu(dx), \ (r \in \mathbb{N}).$$

Very roughly speaking, $d_n p_i^*$ is the multiplicity of the eigenvalue $\sigma_i^*$
$\Rightarrow$ Model not well defined

- Parametric model $\theta^* = (\sigma_1^*, \dots, \sigma_k^*, p_1^*, \dots, p_{k-1}^*)^T$,
- As the white Gaussian is invariant under isometry take $\Sigma$ diagonal,
- Diagonal elements $\sigma_1^*, \dots, \sigma_k^*$ with random multiplicity $N_{1,n}, \dots, N_{k,n} \quad (N_{1,n} + \dots + N_{k,n} = d_n)$,
- $\lim_{n\to\infty} \frac{N_{i,n}}{d_n} = p_i^*$ and $n(\frac{N_{i,n}}{d_n} - p_i^*) = O_p(1)$.

# La La LAN

Under some assumptions on the random multiplicities :

## Proposition

*The model is LAN for the rate* $n^{-1}$

$$\log\left(\frac{dL_{\theta^*+\frac{h}{n}}}{dL_{\theta^*}}\right) = \langle h, \Delta_n\rangle - \frac{1}{2}h^T I_{\theta^*} h + o_p(1),$$

*with* $\Delta_n \overset{\mathcal{L}}{\Rightarrow} \mathcal{N}_{2k-1}(0, I_{\theta^*})$

Consequently, possible asymptotic efficient estimation could be possible (Newton Method).
But

- $I_{\theta^*}$ is not yet completely computable !!
- Our estimate (see next slides) have to be regular (not shown yet) !!

# Definitions and notations

- The Cauchy-Stieltjes transform of $\mu$, a probability measure on $\mathbb{R}_+$ is :

$$\begin{array}{rccc} G_\mu : & \mathbb{C}_+ & \to & \mathbb{C}_- \\ & z & \mapsto & \int_{\mathbb{R}_+} \frac{\mu(d\nu)}{z-\nu} \end{array},$$

  where

$$\mathbb{C}_\pm := \{z \in \mathbb{C} \mid \pm \Im z > 0\} .$$

- The moment generating function is $M_\mu(z) = zG_\mu(z) - 1$
- For $\mu \neq \delta_0$, $M_\mu$ is invertible in the neighborhood of $\infty$ and the inverse is denoted by $M_\mu^{\langle -1 \rangle}$. The S-transform of $\mu$ is defined as

$$S_\mu(m) = \frac{1+m}{mM_\mu^{\langle -1 \rangle}(m)},$$

  and is analytic in a neighborhood of $m = 0$. (WARNING : Inverse is multi-valued).
- For a diagonalizable matrix $A^{(N)} \in M_N(\mathbb{R})$, we write $S_{A^{(N)}} := S_{\mu_{A^{(N)}}}$, $G_{A^{(N)}} := G_{\mu_{A^{(N)}}}$, $M_{A^{(N)}} := M_{\mu_{A^{(N)}}}$.

# $S$-transforms and free convolution

A more complete statement on free multiplicative convolution :

## Theorem (Voiculescu, 1987)

*Consider two sequences of positive matrices, each element in* $M_N(\mathbb{R})$

$$\left(A^{(N)}\,;\,N \geqslant 1\right),\quad \left(B^{(N)}\,;\,N \geqslant 1\right),$$

*such that :*

$$\lim_{N\to\infty} \mu_{A^{(N)}} = \mu_A\,,\quad \lim_{N\to\infty} \mu_{B^{(N)}} = \mu_B\,.$$

*Under the (technical) assumption of asymptotic freeness for* $A^{(N)}$ *and* $B^{(N)}$*, there exists a deterministic probability measure* $\mu_A \boxtimes \mu_B$ *such that :*

$$\lim_{N\to\infty} \mu_{(A^{(N)})^{\frac{1}{2}} B^{(N)} (A^{(N)})^{\frac{1}{2}}} = \mu_A \boxtimes \mu_B\,.$$

*The operation* $\boxtimes$ *is the multiplicative free convolution. Moreover*

$$S_{\mu_A \boxtimes \mu_B}(m) = S_{\mu_A}(m) S_{\mu_B}(m)\,. \qquad (1)$$

Recall $\mathbb{X} = \Sigma^{\frac{1}{2}}\mathbb{Y}$ so that :

$$\frac{1}{n}\mathbb{X}\mathbb{X}^* = \Sigma^{\frac{1}{2}}\left(\frac{1}{n}\mathbb{Y}\mathbb{Y}^*\right)\Sigma^{\frac{1}{2}} .$$

**Corollary**

*Assume that $\mathbb{Y}$ has iid coefficients, and $\Sigma$ has a spectral measure converging to $\nu$. Under our assumptions, $\mu_{\frac{1}{n}\mathbb{X}\mathbb{X}^*}$ converges weakly to a measure $\mu$ satisfying :*

$$\mu = \nu \boxtimes MP_c .$$

$\rightsquigarrow$ We want to estimate $\nu$ by free deconvolution i.e. such that

$$S_\nu(m) = S_\mu(m)/S_{MP_c}(m) .$$

In particular we want to construct an empirical version $\hat{\nu}_n$ such that

$$S_{\hat{\nu}_n}(m) \approx S_{\frac{1}{n}\mathbb{X}\mathbb{X}^*}(m)/S_{MP_c}(m)$$

Ingredients for the estimator $\hat{\nu}_n$ :

**Proposition (Inversion)**

*It is possible to recover moments of $\hat{\nu}_n$ via (numerically constructible and small) contour integrals implicating $G_{\frac{1}{n}\mathbb{X}\mathbb{X}^*}$.*

**Proposition (Reconstruction)**

*From the observation of (noisy) moments $m_1, m_2, \ldots, m_k$, one can reconstruct an atomic probability measure $\hat{\nu}_n$ having these moments.*

# JOYEUX ANNIVERSAIRE ELISABETH