Statistical Inverse Problems: Theory II.1 (Deconvolution)

Axel Munk

Institut für Mathematische Stochastik, Göttingen

www.stochastik.math.uni-goettingen.de/munk
munk@math.uni-goettingen.de

January 22, 2010

Apologizes

"What a day! The computer broke down and we all had to think!"

CSIRO
Mathematical and Information Sciences

II. 1. Kernel Deconvolution

II.1. Kernel deconvolution

Model: observe $Y_1, \ldots, Y_n \overset{iid}{\sim} H$; density h

$$Y_i = X_i + \epsilon_i \quad i = 1, \ldots, n \quad \text{corrupted by noise}$$

$X_i \overset{iid}{\sim} f$ (density) to be estimated, independent of
$\epsilon_i \overset{iid}{\sim} g$ (error) known.

Hence

$$h(\cdot) = (g * f)(\cdot) = \int g(\cdot - y) f(y) dy \quad (convolution) \tag{1}$$

Goal: Recover f
"Inversion" of (1) is called **deconvolution**.

II.1. Kernel deconvolution

Model: observe $Y_1, \ldots, Y_n \overset{iid}{\sim} H$; density h

$$Y_i = X_i + \epsilon_i \quad i = 1, \ldots, n \quad \text{corrupted by noise}$$

$X_i \overset{iid}{\sim} f$ (density) to be estimated, independent of

$\epsilon_i \overset{iid}{\sim} g$ (error) known.

Hence

$$h(\cdot) = (g * f)(\cdot) = \int g(\cdot - y) f(y) dy \quad (convolution) \tag{1}$$

Goal: Recover f

"Inversion" of (1) is called **deconvolution**.

To see the difficulty of the problem,

assume f and g are discrete,

then (1) becomes with $g(x_i - y_j) = g_{ij}$

$$\begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = (g_{ij})_{\substack{j=1,\dots,m \\ i=1,\dots,n}} \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix} \tag{2}$$

Hence, even if we would observe g and h,

"inversion" of (2) is instable.

$$\Psi_h(t) := \int exp\{itx\}\ h(x)\ dx$$
$$= E\left(exp\{itY_1\}\right) \qquad \text{(characteristic function)}$$

One has

$$\Psi_h(t) = E\left(exp\{it\left(X_1 + \epsilon_1\right)\}\right)$$
$$\overset{\text{indep.}}{=} E\left(exp\{itX_1\}\right)\ E\left(exp\{it\epsilon_1\}\right)$$
$$= \Psi_f(t) \cdot \Psi_g(t)$$

$$\Psi_h(t) := \int exp\{itx\}\ h(x)\ dx$$
$$= E\left(exp\{itY_1\}\right) \qquad \text{(characteristic function)}$$

One has

$$\Psi_h(t) = E\left(exp\{it\left(X_1+\epsilon_1\right)\}\right)$$
$$\overset{\text{indep.}}{=} E\left(exp\{itX_1\}\right)\ E\left(exp\{it\epsilon_1\}\right)$$
$$= \Psi_f(t)\cdot\Psi_g(t)$$

Hence (formal division):

$$\Psi_f = \frac{\Psi_h}{\Psi_g} \tag{3}$$

Fourier-transform

$$\Psi_h(t) := \int exp\{itx\}\ h(x)\ dx$$
$$= E\left(exp\{itY_1\}\right) \qquad \text{(characteristic function)}$$

One has

$$\begin{aligned} \Psi_h(t) &= E\left(exp\{it\left(X_1+\epsilon_1\right)\}\right) \\ &\overset{\text{indep.}}{=} E\left(exp\{itX_1\}\right)\ E\left(exp\{it\epsilon_1\}\right) \\ &= \Psi_f(t)\cdot\Psi_g(t) \end{aligned}$$

Hence (formal division):

$$\Psi_f = \frac{\Psi_h}{\Psi_g} \tag{3}$$

f bounded, continuous, then $\Psi_f \in L^1$

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} exp\{-itx\}\ \Psi_f(t)dt \qquad (4)$$

$$= \frac{1}{2\pi}\Psi_{\Psi_f}(-x)$$

Estimation (first attempt)

We estimate Ψ_h by

$$\widehat{\Psi}_h(t) = \frac{1}{n}\sum_{j=1}^{n} exp\{itY_j\}$$

Plug in (4):

$$\widehat{f}(x) = \frac{1}{2\pi}\int exp\{-itx\}\,\frac{\widehat{\Psi}_h(t)}{\Psi_g(t)}dt \qquad \text{"does not exist"}$$

Estimation (first attempt)

We estimate Ψ_h by

$$\widehat{\Psi}_h(t) = \frac{1}{n}\sum_{j=1}^{n} exp\{itY_j\}$$

Plug in (4):

$$\widehat{f}(x) = \frac{1}{2\pi}\int exp\{-itx\}\,\frac{\widehat{\Psi}_h(t)}{\Psi_g(t)}dt \qquad \text{"does not exist"}$$

Problem I: $\widehat{\Psi}_h(t)$ not integrable over $\mathbb{R}$: $\int_{\mathbb{R}} \underbrace{exp\{itY\}}_{|\cdot|=1} dt = \infty$

Estimation (first attempt)

We estimate Ψ_h by

$$\widehat{\Psi}_h(t) = \frac{1}{n}\sum_{j=1}^{n} exp\{itY_j\}$$

Plug in (4):

$$\widehat{f}(x) = \frac{1}{2\pi}\int exp\{-itx\}\,\frac{\widehat{\Psi}_h(t)}{\Psi_g(t)}dt \qquad \text{"does not exist"}$$

Problem I: $\widehat{\Psi}_h(t)$ not integrable over $\mathbb{R}$: $\int_{\mathbb{R}} \underbrace{exp\{itY\}}_{|\cdot|=1} dt = \infty$

Quite funny: "on average" things are OK:

$$E\left(\widehat{\Psi}_h(t)\right) = E\left(\exp\{itY_1\}\right)$$
$$= \int \exp\{ity\} \underbrace{h(y)}_{\text{"damping function"}} dy$$

Problem II: In (3) and $\hat{f}$, division by Ψ_g is very "instable", if $\Psi_g \sim 0$.

This can be drastic (exponential decay)

$$\text{for} \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}x^2\} \qquad \text{(normal)}$$

$$\text{Claim:} \qquad \Psi_\Phi(t) = \exp\{-\frac{1}{2}t^2\}.$$

Because of

$$\Psi'_\Phi(t) = \frac{1}{\sqrt{2\pi}} \int \underbrace{(e^{itx} i)}_{f} \underbrace{\left(x \exp\{-\frac{1}{2}x^2\}\right)}_{g'} dx$$

$$\overset{\text{partial int.}}{=} -\sqrt{2\pi}\ t\ \Phi(t) = -t\ \Psi_\Phi(t)$$

This can be drastic (exponential decay)

$$\text{for} \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}x^2\} \qquad \text{(normal)}$$

$$\text{Claim:} \quad \Psi_\Phi(t) = \exp\{-\frac{1}{2}t^2\}.$$

Because of

$$\Psi'_\Phi(t) = \frac{1}{\sqrt{2\pi}} \int \underbrace{(e^{itx} i)}_{f} \underbrace{\left(x \exp\{-\frac{1}{2}x^2\}\right)}_{g'} dx$$

$$\overset{\text{partial int.}}{=} -\sqrt{2\pi}\ t\ \Phi(t) = -t\ \Psi_\Phi(t)$$

Kernel Deconvolution (second attempt)

Estimate h by a kernel estimator first

$$\widehat{h}(x) = \frac{1}{nb} \sum_{j=1}^{n} k\left(\frac{x - Y_j}{b}\right)$$

(local averaging)

$b > 0$ bandwidth,

$k \in L^1 \cap L^2$ kernel; $\quad \int k = 1$

$$
\begin{aligned}
\Psi_{\widehat{h}}(t) &= \frac{1}{nb}\sum_{j=1}^{n}\int \exp\{itx\}\cdot k\left(\frac{x-Y_j}{b}\right)\ dx \\
&\overset{(*)}{=} \frac{1}{n}\sum_{j=1}^{n}\int \exp\left(it\left(zb+Y_j\right)\right)\cdot k(z)\ dz \\
&= \frac{1}{n}\sum_{j=1}^{n}\exp\{itY_j\}\int \exp\{itbz\}\cdot k(z)\ dz \\
&= \widehat{\Psi}_h(t) \qquad \underbrace{\Psi_k(tb)}_{\text{e.g. } \sim\exp\{-\frac{1}{2}(tb)^2\}} \qquad (5)
\end{aligned}
$$

"regulates" the spectrum of Ψ_h (cf. (3))

$(*)$ substitute $z=\frac{x-Y_i}{b}$

Now, let us apply Fourier-inversion (4) to (5):

$$\widehat{f}(x) = \frac{1}{2\pi} \int \exp(-itx) \frac{\Psi_k(tb)}{\Psi_g(t)} \; \frac{1}{n} \sum_{j=1}^{n} \exp\{itY_j\} dt$$

(Deconvolution Kernel Estimate DKE)

(first introduced by Carroll, Hall '88 JASA)

Remarks:

- Note that f is only well defined, if $\Psi_g \neq 0$
- If $\Psi_g = 0$ then things become rather involved
  Boxcar deconvolution (Donoho et al.'04, JRSSB)

$$\Psi_g = \frac{\sin t}{t}$$

- $\widehat{f}$ is complex-valued, practice: $Re(\widehat{f})$

Remarks:

- Note that f is only well defined, if $\Psi_g \neq 0$
- If $\Psi_g = 0$ then things become rather involved
  Boxcar deconvolution (Donoho et al.'04, JRSSB)

$$\Psi_g = \frac{\sin t}{t}$$

- $\widehat{f}$ is complex-valued, practice: $Re(\widehat{f})$

Spectral regularisation of kernel smoothing

- Regularisation in the spectral domain by kernel smoothing

$$\Psi_{\widehat{f}}(t) = \widehat{\Psi}_h(t) \cdot \underbrace{\frac{\Psi_k(tb)}{\Psi_g(t)}}_{\alpha_b(t)}$$

$b = 0 \quad$ (no regularisation)

$b \to \infty, \ \alpha_b \to 0 \quad$ small spectrum of g will be regularised

In order to understand the statistical properties of $\widehat{f}$

regularity assumptions on g and f have to be imposed.

For simplicity (generalisations to Sobolev-spaces possible)

Conditions on f:

- $||f||_\infty \leq c$

  $f \in C^{(\beta)}(\mathbb{R})$ $\quad$ (β-times cont. diff.); $\beta \in \mathbb{N}$

- $|f^{(\lfloor\beta\rfloor)}(x) - f^{(\lfloor\beta\rfloor)}(y)| \leq c|x-y|^{\beta-\lfloor\beta\rfloor}$ $\quad$ $\beta > 0$

  (Hölder-cont. derivative)

Conditions on g:

- Polynomial smoothness (PS)

$$\Psi_g(t) \sim (1 + |t|)^{-\alpha} \qquad \alpha > 0$$

- exponential smoothness (ES)

$$\Psi_g(t) \sim \exp(-d|t|^{\gamma}) \qquad \gamma > 0$$

a) Cauchy-deconvolution

$g(x) = \frac{1}{\pi}\left(1 + x^2\right)^{-1}$ $\qquad \Psi_g(t) = \exp\left(-|t|\right)$

(ES, $\gamma = 1, d = 1$)

b) vice versa: Laplace-deconvolution

$g(x) = \frac{1}{2}\exp\{-|x|\}$ $\qquad \Psi_g(t) = (1 + t^2)^{-1}$

(PS, $\alpha = 2$)

c) normal (most prominent)

$\Psi_g(t) = \exp\{-\frac{1}{2}t^2\}$ $\qquad$ (EF, $\gamma = 2,\ d = \frac{1}{2}$)

Note: Smoothness of g translates into tail behaviour of Ψ_g

Examples

a) Cauchy-deconvolution

$g(x) = \frac{1}{\pi}\left(1+x^2\right)^{-1}$ $\qquad$ $\Psi_g(t) = \exp\left(-|t|\right)$

(ES, $\gamma = 1, d = 1$)

b) vice versa: Laplace-deconvolution

$g(x) = \frac{1}{2}\exp\{-|x|\}$ $\qquad$ $\Psi_g(t) = (1+t^2)^{-1}$

(PS, $\alpha = 2$)

c) normal (most prominent)

$\Psi_g(t) = \exp\{-\frac{1}{2}t^2\}$ $\qquad$ (EF, $\gamma = 2,\ d = \frac{1}{2}$)

Note: Smoothness of g translates into tail behaviour of Ψ_g

Condition on k: (β-order kernel)

$$\int k(z)\ z^j\ dz = 0; \qquad j = 1, \ldots, \lfloor \beta \rfloor$$

MSE-calculations (pointwise)

$$\begin{aligned}\text{MSE}\left(\widehat{f}(x)\right) &= E\left|\widehat{f}(x) - f(x)\right|^2 \\ &= Var\left(\widehat{f}(x)\right) + \{\underbrace{E\left(\widehat{f} - f\right)(x)}_{(\text{Bias})^2}\}^2\end{aligned}$$

Exactly as for kernel-estimation:

$$
\begin{aligned}
E\left(\widehat{f}(x)\right) &= \frac{1}{2\pi}\int exp(-itx)\frac{\Psi_k(tb)}{\Psi_g(t)}E\left(exp\{itY_1\}\right)dt \\
&= \frac{1}{2\pi}\int exp(-itx)\Psi_k(tb)\Psi_f(t)dt \\
&= \frac{1}{2\pi}\int exp(-itx)\Psi_{(k_b * f)}(t)dt
\end{aligned}
$$

Fourier-inversion (cond. to be checked! $k_b * f$ is continuous)

$$= (k_b * f)(x)$$

Hence g does not enter the picture!

$\overset{\text{(Taylor)}}{\Rightarrow}$ $\text{Bias}^2 = O(b^{2\beta})$ as $\beta \searrow 0$

Bounding the Bias

Exactly as for kernel-estimation:

$$
\begin{aligned}
E\left(\widehat{f}(x)\right) &= \frac{1}{2\pi}\int exp(-itx)\frac{\Psi_k(tb)}{\Psi_g(t)}E\left(exp\{itY_1\}\right)dt \\
&= \frac{1}{2\pi}\int exp(-itx)\Psi_k(tb)\Psi_f(t)dt \\
&= \frac{1}{2\pi}\int exp(-itx)\Psi_{(k_b*f)}(t)dt
\end{aligned}
$$

Fourier-inversion (cond. to be checked! $k_b * f$ is continuous)

$$= \left(k_b * f\right)(x)$$

Hence g does not enter the picture!

$\overset{\text{(Taylor)}}{\Rightarrow}$ $\text{Bias}^2 = O(b^{2\beta})$ as $\beta \searrow 0$

Bounding the variance

Let

$$\frac{\Psi_k(t\ b)}{\Psi_g(t)} = \Psi(t)$$

$$\begin{aligned}
Var\widehat{f}(x) &= \frac{1}{(2\pi)^2}\frac{1}{n^2}\sum_{j=1}^{n} Var\int exp\{-it(x-Y_i)\}\ \Psi(t)dt \\
&\leq c\frac{1}{n}E\left|\int exp\{-it(x-Y_1)\}\ \Psi(t)dt\right|^2 \\
&= c\ n^{-1}\int\left|\int exp\{-itz\}\ \Psi(t)dt\right|^2 (f*g)(x-z)dz \\
&\leq \underbrace{||f*g||_\infty}_{bounded} c\ n^{-1}\int\left|\int exp\{-itz\}\Psi(t)dt\right|^2 dz \\
&\overset{\text{Parseval}}{=} n^{-1}c'||\Psi||_2^2 \qquad (6)
\end{aligned}$$

The last term (6): let k a kernel with support $[-1,1]$ for Ψ_k

$$\int \left|\frac{\Psi_k(tb)}{\Psi_g(t)}\right|^2 dt \leq \left(\frac{1}{\min_{|t|\leq\frac{1}{b}} |\Psi_g(t)|}\right)^2 \frac{1}{b} \underbrace{\int |\Psi_k(t)|^2\, dt}_{<\infty}$$

Hence (6) becomes

$$C\frac{1}{b\ n}\left(\min_{|t|\leq\frac{1}{b}} |\Psi_g(t)|\right)^{-2}$$

The last term (6): let k a kernel with support $[-1,1]$ for Ψ_k

$$\int \left|\frac{\Psi_k(tb)}{\Psi_g(t)}\right|^2 dt \leq \left(\frac{1}{\min_{|t|\leq \frac{1}{b}} |\Psi_g(t)|}\right)^2 \frac{1}{b} \underbrace{\int |\Psi_k(t)|^2\, dt}_{<\infty}$$

Hence (6) becomes

$$C\frac{1}{b\ n}\left(\min_{|t|\leq \frac{1}{b}} |\Psi_g(t)|\right)^{-2}$$

(PS) $O\left(\frac{b^{-1-2\alpha}}{n}\right), \quad n \to \infty, b \to \infty$

The last term (6): let k a kernel with support $[-1,1]$ for Ψ_k

$$\int \left|\frac{\Psi_k(tb)}{\Psi_g(t)}\right|^2 dt \leq \left(\frac{1}{\min_{|t|\leq\frac{1}{b}}|\Psi_g(t)|}\right)^2 \frac{1}{b} \underbrace{\int |\Psi_k(t)|^2 \, dt}_{<\infty}$$

Hence (6) becomes

$$C\frac{1}{b\ n}\left(\min_{|t|\leq\frac{1}{b}}|\Psi_g(t)|\right)^{-2}$$

(PS) $O\left(\frac{b^{-1-2\alpha}}{n}\right)$, $\quad n\to\infty, b\to\infty$

(ES) $O\left(\frac{1}{bn}exp\{2db^{-\gamma}\}\right)$ This grows extremely fast as $b\searrow 0$

The last term (6): let k a kernel with support $[-1,1]$ for Ψ_k

$$\int \left|\frac{\Psi_k(tb)}{\Psi_g(t)}\right|^2 dt \leq \left(\frac{1}{\min_{|t|\leq\frac{1}{b}} |\Psi_g(t)|}\right)^2 \frac{1}{b} \underbrace{\int |\Psi_k(t)|^2 \, dt}_{<\infty}$$

Hence (6) becomes

$$C\frac{1}{b\ n}\left(\min_{|t|\leq\frac{1}{b}} |\Psi_g(t)|\right)^{-2}$$

(PS) $O\left(\frac{b^{-1-2\alpha}}{n}\right), \quad n \to \infty, b \to \infty$

(ES) $O\left(\frac{1}{bn} exp\{2db^{-\gamma}\}\right)$ This grows extremely fast as $b \searrow 0$

MSE bound

Balancing this with $\text{bias}^2 = O(b^{2\beta})$ gives

$$
\begin{array}{llll}
 & & \text{MSE} & \\
\text{(PS)} & b \sim n^{-1/(2\beta+2\alpha+1)} & O\left(n^{-2\beta/(2\beta+2\alpha+1)}\right) & \\
 & & & \textbf{(R)} \\
\text{(ES)} & b \sim (\log n)^{-1/\gamma} & O\left((\log n)^{-2\beta/\gamma}\right) &
\end{array}
$$

- Careful inspection shows that these rates are uniform in a neighbourhood of x.

Balancing this with $\text{bias}^2 = O(b^{2\beta})$ gives

$$
\begin{array}{lll}
 & & \text{MSE} \\
\text{(PS)} & b \sim n^{-1/(2\beta+2\alpha+1)} & O\left(n^{-2\beta/(2\beta+2\alpha+1)}\right) \\
\text{(ES)} & b \sim (\log n)^{-1/\gamma} & O\left((\log n)^{-2\beta/\gamma}\right)
\end{array}
\qquad \textbf{(R)}
$$

- Careful inspection shows that these rates are uniform in a neighbourhood of x.

- In fact, one can show (Fan'91, JASA)

$$\inf_{\widehat{f}} \sup_{f \in \mathcal{F}_C} E||\widehat{f} - f||^2 \geq c \cdot \text{ rates in } (\mathbf{R})$$

The DKE is rate minimax-optimal

- In fact, one can show (Fan'91, JASA)

$$\inf_{\widehat{f}} \sup_{f \in \mathcal{F}_C} E||\widehat{f} - f||^2 \geq c \cdot \text{ rates in } (\mathbf{R})$$

The DKE is rate minimax-optimal

The degree of ill posedness in deconvolution

Examples:

			without convolution
a)	Cauchy ($\gamma = 1$)	$(\log n)^{-2\beta}$	$n^{-2\beta/(2\beta+1)}$
b)	Laplace ($\alpha = 2$)	$n^{-2\beta/(2\beta+5)}$	$n^{-2\beta/(2\beta+1)}$
c)	Normal ($\gamma = 2$)	$(\log n)^{-\beta}$	$n^{-2\beta/(2\beta+1)}$

The gaussian case looks horrible and Gaussian deconvolution seems completely hopeless (as a referee once remarked smartly)!

The degree of ill posedness in deconvolution

Examples:

			without convolution
a)	Cauchy ($\gamma = 1$)	$(\log n)^{-2\beta}$	$n^{-2\beta/(2\beta+1)}$
b)	Laplace ($\alpha = 2$)	$n^{-2\beta/(2\beta+5)}$	$n^{-2\beta/(2\beta+1)}$
c)	Normal ($\gamma = 2$)	$(\log n)^{-\beta}$	$n^{-2\beta/(2\beta+1)}$

The gaussian case looks horrible and Gaussian deconvolution seems completely hopeless (as a referee once remarked smartly)!
Unfortunalety, in practice we do it all the time...

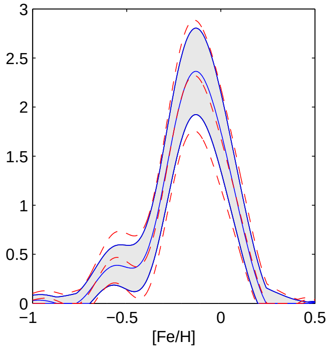
The degree of ill posedness in deconvolution

Examples:

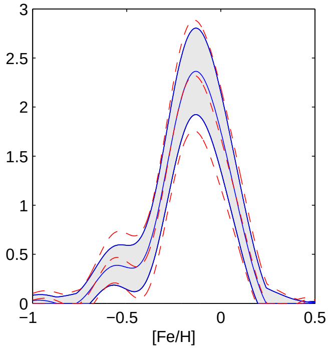
			without convolution
a)	Cauchy ($\gamma = 1$)	$(\log n)^{-2\beta}$	$n^{-2\beta/(2\beta+1)}$
b)	Laplace ($\alpha = 2$)	$n^{-2\beta/(2\beta+5)}$	$n^{-2\beta/(2\beta+1)}$
c)	Normal ($\gamma = 2$)	$(\log n)^{-\beta}$	$n^{-2\beta/(2\beta+1)}$

The gaussian case looks horrible and Gaussian deconvolution seems completely hopeless (as a referee once remarked smartly)! Unfortunalety, in practice we do it all the time...

Laplace vs. Gauss Deconvolution



Laplace (blue) together with 90% confidence band (bootstrap), Gaussian (red), Bissantz et al.'07 JRSSB. Scaled such that variances are equal.



Laplace (blue) together with 90% confidence band (bootstrap), Gaussian (red), Bissantz et al.'07 JRSSB. Scaled such that variances are equal.

Summary Part II.1

- Deconvolution is a SIP and requires regularisation
- Simple kernel estimators are (rate) minimax

Summary Part II.1

- Deconvolution is a SIP and requires regularisation
- Simple kernel estimators are (rate) minimax
- Fourier analysis helps to analyze and to understand their behaviour:
  They regularize in the spectral domain

Summary Part II.1

- Deconvolution is a SIP and requires regularisation
- Simple kernel estimators are (rate) minimax
- Fourier analysis helps to analyze and to understand their behaviour:
  They regularize in the spectral domain
- Minimax: The pessimist's point of view:
  Worst case scenario can be very misleading

Summary Part II.1

- Deconvolution is a SIP and requires regularisation
- Simple kernel estimators are (rate) minimax
- Fourier analysis helps to analyze and to understand their behaviour:
  They regularize in the spectral domain
- Minimax: The pessimist's point of view:
  Worst case scenario can be very misleading
- Nevertheless, it tells something

Summary Part II.1

- Deconvolution is a SIP and requires regularisation
- Simple kernel estimators are (rate) minimax
- Fourier analysis helps to analyze and to understand their behaviour:
  They regularize in the spectral domain
- Minimax: The pessimist's point of view:
  Worst case scenario can be very misleading
- Nevertheless, it tells something