Statistical Inverse Problems –
an Example Based Tutorial

Axel Munk
Institute for Mathematical Statistics
Georg August University Göttingen

munk@math.uni-goettingen.de

www.stochastik.math.uni-goettingen.de/munk
www.stochastik.math.uni-goettingen.de/for916

This talk is (almost) not about:

- Algorithms/computation/optimisation
- Sparsity/Compression

Outline:

1. Introduction/Examples
2. SIP: Some Theory, Kernel Density estimation
3. Linear Spectral Regularisation methods
4. Statistical Inverse Problems and Statistical Multiscale Analysis

Inverse problems:

Two problems are inverses of each other, if the formulation of each involves all or part of the solution of the other. Historically, one direction of the problem is well-studied since long time ago (usually called the "direct" problem), the other only recently studied and less well understood ("inverse problem")

(J.B. Keller '76, Am. Math. Monthly)

In other words: Here is the answer...

... but what was the question?

Or: Climbing a rock often is easier than getting down

Google: ~ 3Mio hits

Statistical Regularisation

Statistical Inverse Problems (SIPs)

Google: ~1.5Mio hits (still...)

Areas of application (a selection):

- Physics: astronomy, material science, ...
- Medical physics: NMR, MRT, tomography,
  biophiyscs, molecular microscopy
- Engeneering: chemical engineering, control theory
- Medical statistics: clinical trials, survival analysis
- Computer science: learning theory,
  network tomography
- Financial mathematics: numerical option pricing,
  financial statistics

Statistical Regularisation

What is a SIP?

1. Johnstone/Silverman'91
   „Data in one domain but interest lies in another domain where measurements cannot be taken"

2. Evans/Stark'02

   $$X \sim P_\theta, \qquad \theta \in \Theta$$ abstract parameter space

   Learn from X about θ, i.e.

   invert the „forward" operator (1-1)

   $$K : \theta \longrightarrow P_\theta$$

   But this is essentially all what Statistics is about...

   Important: In a SIP the inversion

   $$K^{-1} : P_\theta \longrightarrow \theta$$

   is „instable".

Statistical Regularisation

What is a SIP?

3. Hadamard (1923) (he thought on boundary value problems PDE‘s)

„A problem is well posed if

1. There exists a solution (**existence**)
2. There is at most one solution (**uniqueness**)
3. The solution depends continuously on the data (**stability**)

Most SIPs are ill posed (3.)

i.e. small pertubations of the data may affect the solution drastically

Statistics: pertubations are modelled **randomly**

Numerical Analysis: (traditionally)

Pertubations are modelled **deterministically**

Statistical Regularisation

What is a SIP?

´ 4. My own definition (very prelim., strange, not good at all)

Let $(X_1, \cdots, X_n) \sim P_\theta,\ \theta \in \Theta$.

A SIP is a statistical decision problem (estimation, prediction,...) which does not allow $n^{1/2}$-consistent decisions (e.g. in terms of mininax or Bayes-risk), i.e.

$$\liminf_{n\to\infty} n^{1/2} R_n(\hat{\delta}) = \infty$$

I fully agree of you don‘t like it...

One aim of this lecture series is to convince you that it has some appeal and comes close to Hadamard‘s notion of instability.

What is a SIP?

All definitions contain some important ingredient of a SIP

1. „Indirect observations“
2. „Inversion problem“ $K^{-1} : P_\theta \longrightarrow \theta$
3. „Instability“
4. „Instability“:states 3. in a quantitative way

In any case: **It‘s a helpful point of view** (microstructure)

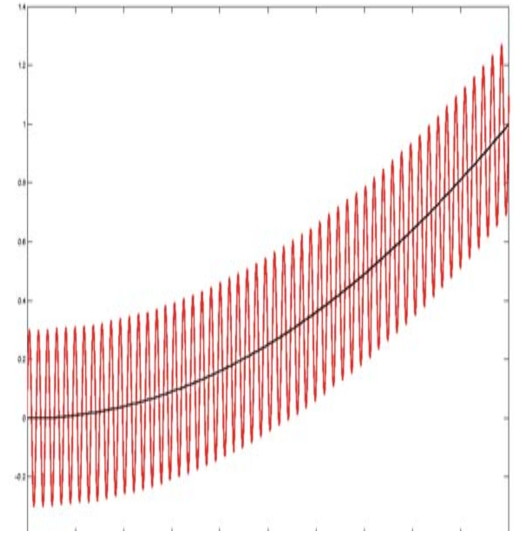
Remarks: - the wheel has been reinvented several times
(EM/R.L.-algorithm, ridge-regression/Tikhnov ...)
- caution: different fields require different treatments
and use different terminology
Examples:
- estimate (stat.) = model (atsrophysics)
- E[ ] = < >

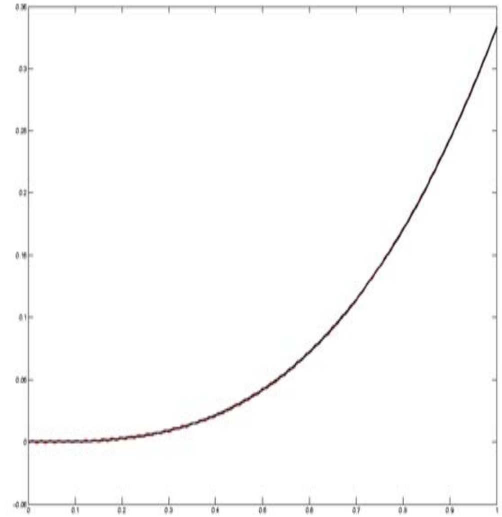
Example 1: (high dimensional regression)

Blackboard-Tutorial

Example 2: Num diff. / densitiy estimation

$$\delta = 0.3, \quad n = 100$$





(Num. differentiation / integration)

Direct problem: Integration

$$F(x) = (Kf)(x) = \int_0^x f(t)dt, \qquad f \in C^{(0)}[0,1]$$

Deterministic error - model

$$f_{n,\delta}(t) = f(t) + \delta \sin\left(\frac{nt}{\delta}\right), \quad \delta \text{ noise level}$$

This implies

$$\begin{aligned} ||f_{n,\delta} - f||_\infty &= \delta \\ ||F_{n,\delta} - F||_\infty &= \left|\left|\frac{\delta^2}{n}\int_0^{nx/\delta} \sin(t)dt\right|\right|_\infty = O(\frac{\delta^2}{n}) \end{aligned}$$

Hence: Any numerical integration rule which approximates Riemann integration reasonably will work

$$\text{error} = 0.9 \times 10^{-4}$$

Statistical Regularisation

Example 2: Num diff. / densitiy estimation

Inverse Problem: Differentiation (same error model)

$$F_{n,\delta}(x) \;=\; F(x) + \delta \sin\left(\frac{nx}{\delta}\right)$$

$$||F_{n,\delta} - F||_\infty \;=\; \delta$$

but

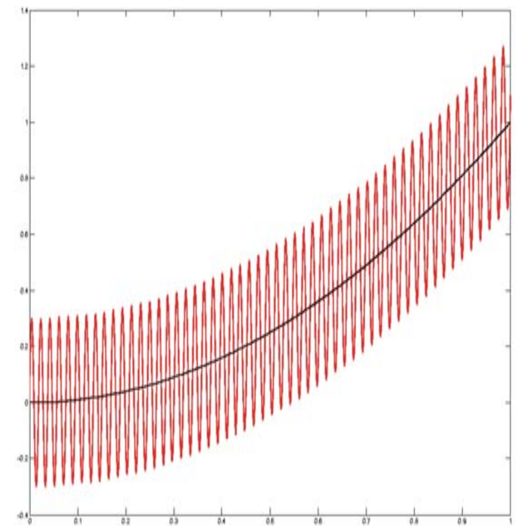
$$F'_{n,\delta}(x) \;=\; F'(x) + n \cos\left(\frac{nx}{\delta}\right)$$

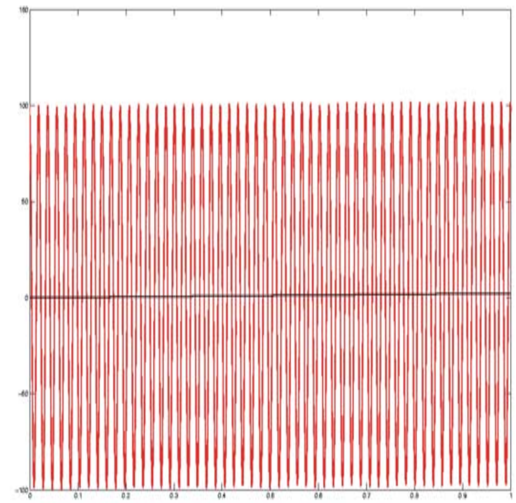
hence

$$||F'_{n,\delta} - F'||_\infty \;=\; n, \qquad \delta \to 0$$

Hence: Even if you could differentiate numerically as good as ‚formal' the error is huge

Error $= 0.3$





Error $= 100$

Some theory: (Tikhonov, Arsenin'77, Engl, Hanke, Neubauer'96,...)

Deterministic error model

$$Y(x) := F_\delta(x) = (Kf)(x) + \delta\epsilon(x), \quad ||\epsilon|| \leq 1$$

$$(Kf)(x) = \int_0^x f$$

Deterministic minimax error:

$$err(\delta) := \inf_{\hat{f}} \sup_{f \in W_2^k} \sup_{||\epsilon|| \leq 1} ||f - \hat{f}||$$

– If $f \in W_2^k$, then the best possible rate of reconstruction of f' is $O(\delta^{k/(k+1)})$, $\delta \to 0$.

– The rate $O(\delta)$ is only achieved for well posed problems.

– Ill posed problems can be characterized by best possible rates which are slower than $O(\delta)$.

Statistical Regularisation

(Estimation of a density/distribution funct.)

$$F(x) = (Kf)(x) = \int_0^x f(t)dt$$

Data model

$$X_1(f), \cdots, X_n(f) \sim F, \qquad \text{i.i.d.}$$

Direct problem: Estimate F, e.g. by F_n, empirical c.d.f.

$$F \quad = F_n + (F - F_n)$$

Error model

$$||F_n - F||_\infty = O_p(\frac{1}{\sqrt{n}}), \qquad 1/\sqrt{n} \sim \delta$$

Statistical Regularisation

Inverse problem: Estimate $F' = f$.

$err(n) := \inf_{\hat{f}} \sup_{f \in W_2^k} E||f - \hat{f}||_2$

Minimax rates for $f \in W_2^k$ are $O(n^{-k/(2k+1)})$

Gaussian white noise model:

$$Y = Kf + \delta\epsilon$$

ϵ is a WN process, s.t. $W' = \epsilon$ for a Wiener process (Nussbaum/Pereverzev'99, $\cdots$)

In a Gaussian W.N. model: $O(\delta^{k/(1+k+1/2))})$

Recall: determinstic $O(\delta^{k/(1+k))})$ (1/2-diff. WP)

Num. Analysis: If the rate δ is achieved the problem is called **well posed**.

Statistics: If the rate $O(n^{-1/2})$ is achieved the problem might be called **parametric**

$$\textbf{Nonparametrics} \subset \textbf{SIPs}$$

Statistical Regularisation

Example 3. Boosting (Bühlmann, Yu'04, JASA, Yao et al.'07, Constr. Approx.)

Regression model: $Y = Kf + \epsilon$.

$$\boxed{K : {W_2}^k \to L^2 \text{ embedding operator}}$$

How to improve (**boost**) an initial estimate $\hat{f}$ (**weak learner**)?
Idea: Use the remaining information in the residuals

$$r = Y - K\hat{f}$$

and update $\hat{f}$ into this direction

$$\begin{aligned} \hat{f}_{new} &= \hat{f} + aK^*r \\ &= \hat{f} - aK^*(K\hat{f} - Y) \\ &= (I - aK^*K)\hat{f} + aK^*Y \end{aligned}$$

Iterate: $\hat{f}_m = (I - aK^*K)\hat{f}_{m-1} + aK^*Y$

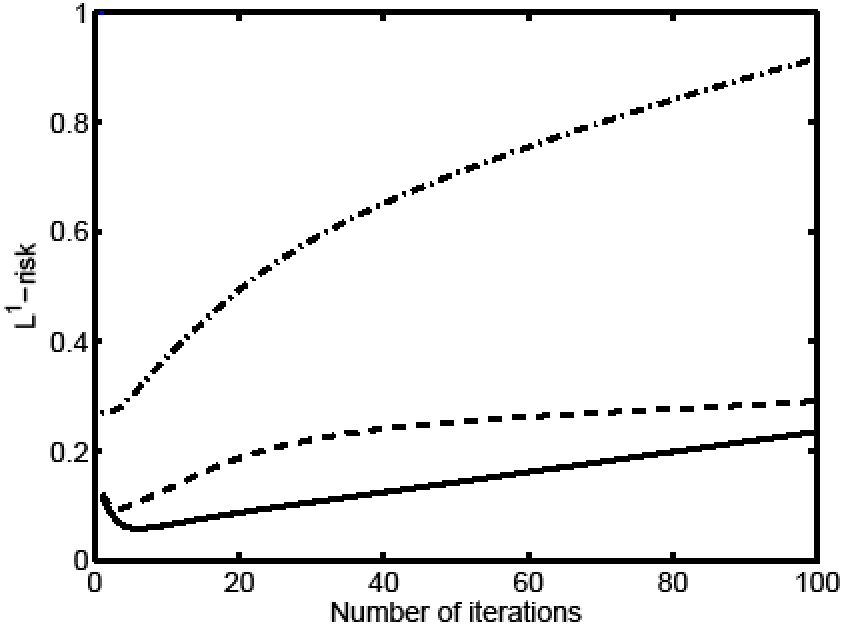
Statistical Regularisation

Example 3. Boosting

Striking feature: The prediction error can be signifcantly improved by boosting the weak learner.

However, simulations show that proper stopping is required. When to stop properly?

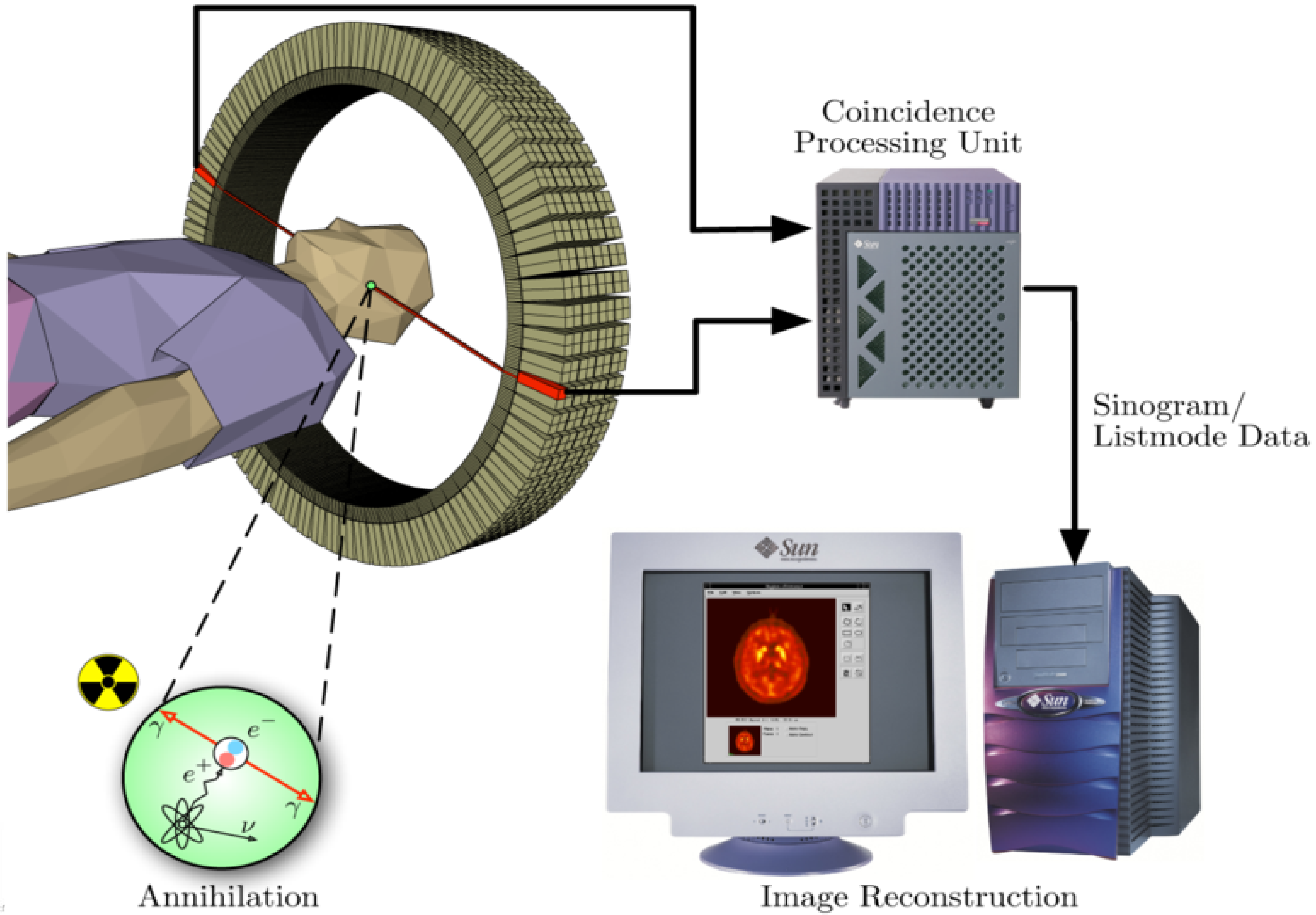
Risk for different weak learners:



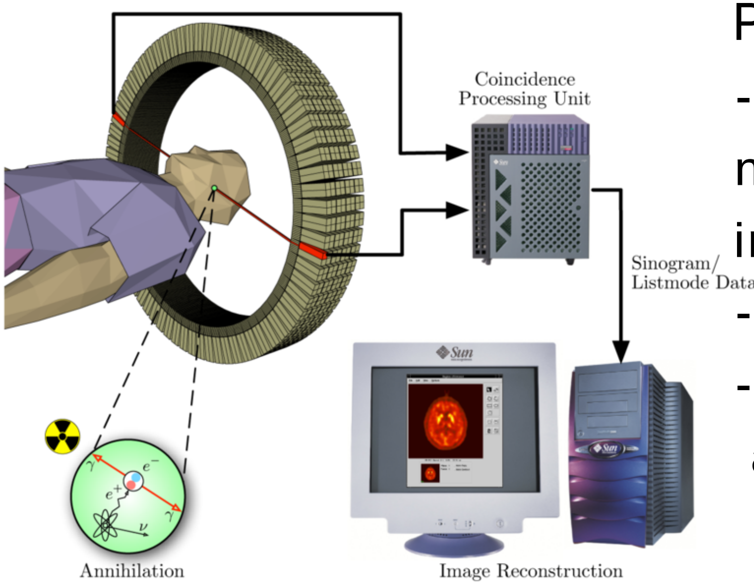
We will analyze this with SIP techniques and we will see that it corrsponds to Landweber iteration (1951, Am J. Math.) (see also Bilay'59, Arch. Rat. Mech. Anal., Friedman'56 Usp. Mat. Nauk.)

Statistical Regularisation

4. Positron Emission Tomography (PET)

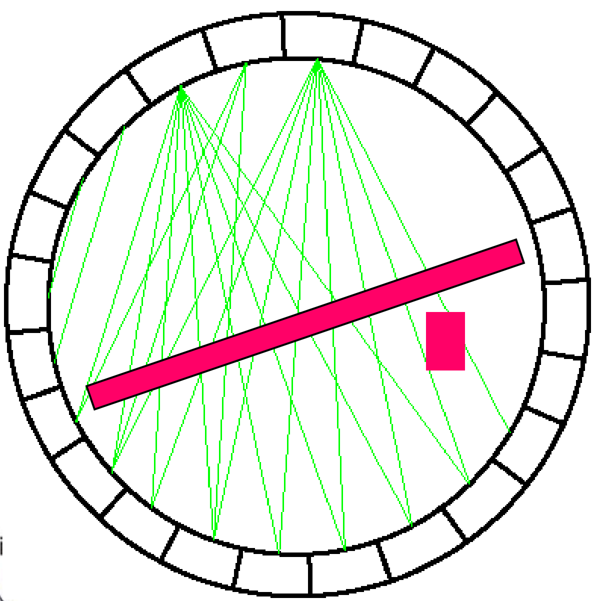


4. Positron Emission Tomography (PET)



Physics of PET (in a nutshell)

- Emission of a positron anihilates with nearby electron: two x-ray photons fly in opposite direction
- This will be counted in opposite detectors
- Directions are approx. uniform and gives an intensity distribution λ^*



$$p(b,d) = P(\text{detected in } d | \text{ emitted in } b)$$

$$\begin{aligned} N^* &= (N_1^*, \cdots, N_D^*) \\ N^* &\sim \mathbf{Poiss}([P\lambda]), \\ E[N^*] &= \lambda^* = P\lambda \end{aligned}$$

4. Positron Emission Tomography (PET)

$$\begin{aligned} N^* &= (N_1^*, \cdots, N_D^*) \\ N^* &\sim \textbf{Poiss}([P\lambda]), \\ E[N^*] &= \lambda^* = P\lambda \end{aligned}$$

The EM algorithm
(Dempster et al.'77, JRSSB
Vardi et al.'85, JASA)

Let $N(b,d) = \#$ emissions from box b detected in d
(incomplete data)
Observe $N_d^* = N(\cdots, d)$ total emissions in d (complete data)

E-step: Estimate emissions in b

$$\hat{N}(b) = E[n(b,\cdot)|\lambda_{old}, N^*]$$

M-step: (MLE for λ) $\lambda_{new} = \hat{N}(b)$

Iterate

Statistical Regularisation

4. Positron Emission Tomography (PET)

Assume $N(b,d)$ independent Poisson

$$[N(b,d) | N_d^*] \sim \text{Bin}\left[N_d^*, \frac{\lambda_{old}(b,d)}{\lambda_{old}(d)}\right]$$

EM-algorithm:

$$\lambda_{new}(b) = \lambda_{old}(b) \sum_d \frac{N^*(d)p(b,d)}{\sum_{b'} \lambda_{old}(b')p(b',d)}$$

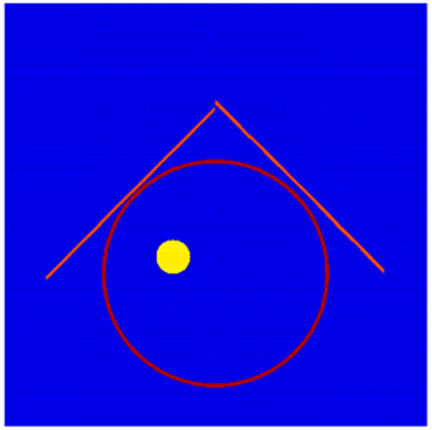
$$\lambda_{k+1} = \lambda_k \left[P^t \left[\frac{N^*}{P\lambda_k}\right]\right]$$

Statistical Regularisation

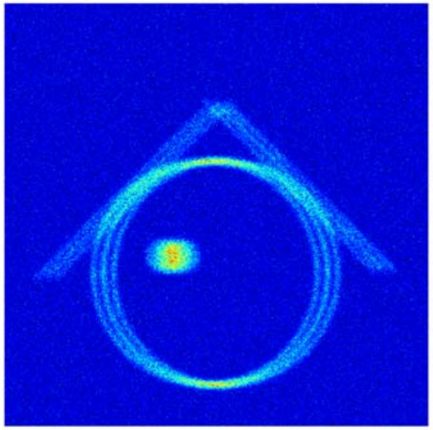
4. Positron Emission Tomography (PET)

The EM algorithm: self regularisation
and stopping

True object



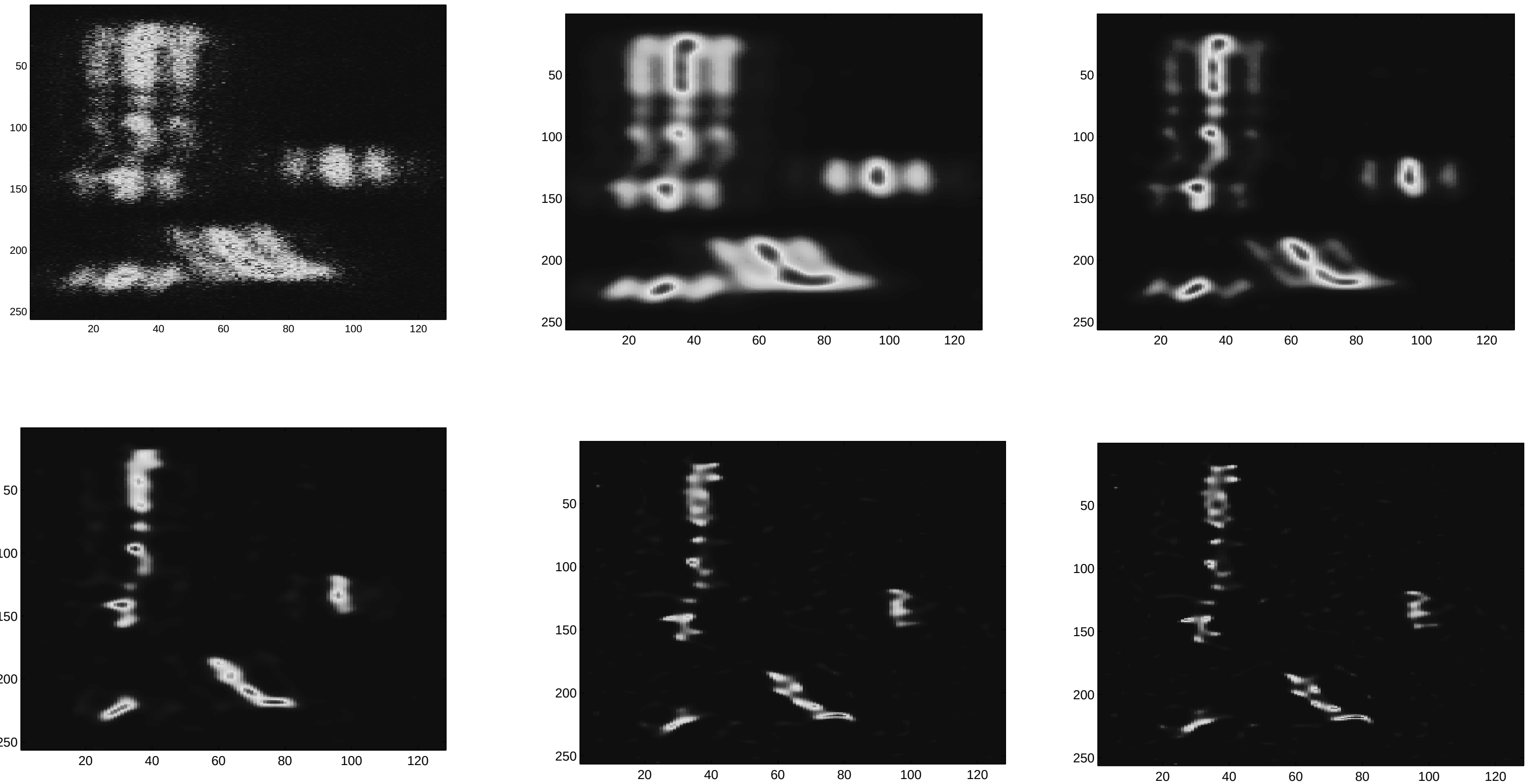
Raw data



St Regularisation

4. Positron Emission Tomography (PET)

The EM algorithm: When to stop?

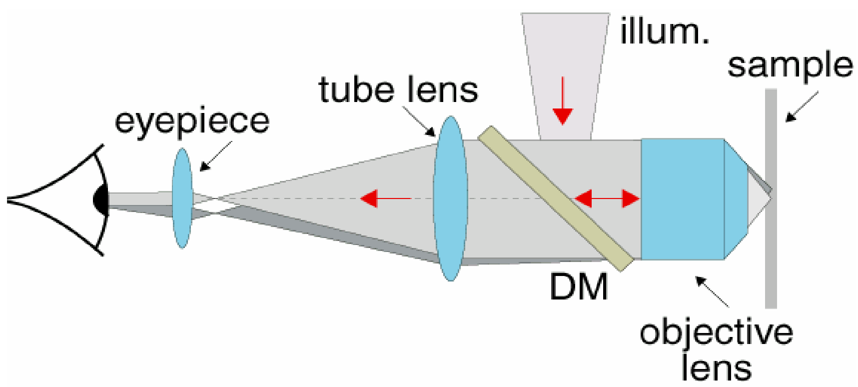


L.t.r.: 4Pi:(trans Golgi protein), obsered data, #iteration=3,9,59,599,999

Pricop, M.‘09

Statistical Regularisation

5. Nanoscopy: Molecuar Microscopy



Confocal microscopy

$$\Delta x \geq \frac{\lambda}{2n\sin(\alpha)}$$

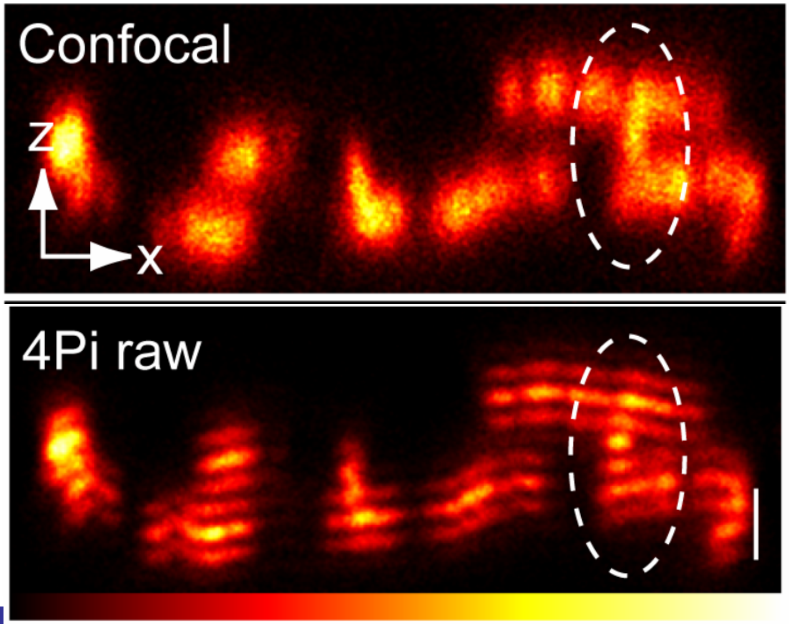
Abbe‘s limit

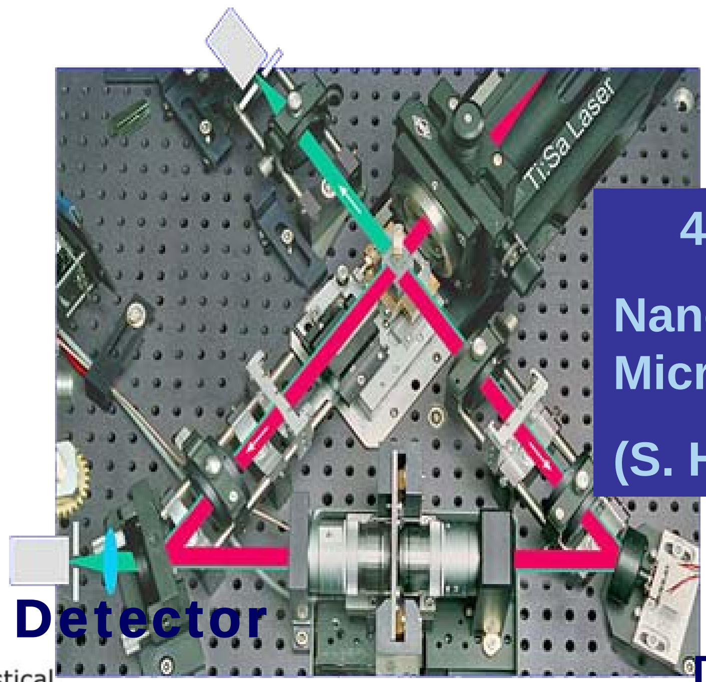
$$\begin{aligned} N^* &= (N_1^*, \cdots, N_D^*) \\ N^* &\sim \mathbf{Poiss}([P\lambda]), \\ E[N^*] &= \lambda^* = P\lambda \end{aligned}$$

Statistical Regularisation

Nanoscopy: Breaking the Abbe limit

4 Pi microscopy:
Cy3 labelled cis-Golgi protein GM130 in Vero cell

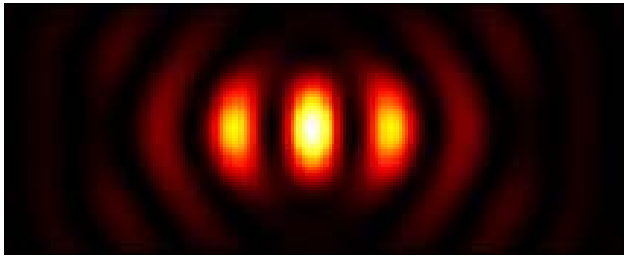




4Pi

Nanoscale Microscopy

(S. Hell)



Detector

Detector

PSF 4 Pi

Statistical Regularisation

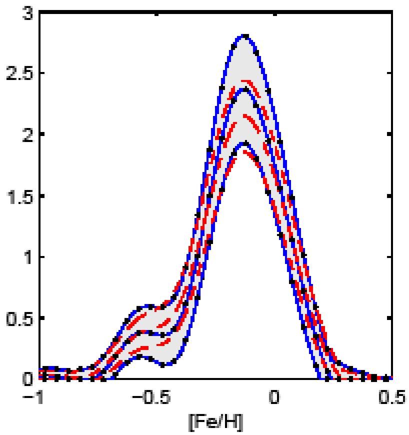
6. Density Deconvolution

Estimation of metal poor G dwarfs in the solar neighborhood
Obs: ~1900 stellar metallicities from the Genova-Copenhagen survey (Nordström et al.‘04).

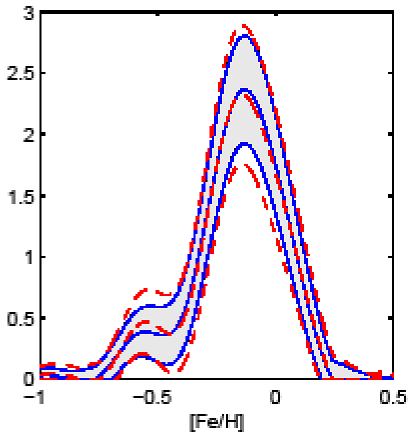
$$X = [Fe/H] + W$$

Kernel density deconvolution with 90% confidence band (Bickel/Rosenblatt’73)

Spiral Galaxy NGC 1232 - VLT UT 1 + FORS1
ESO PR Photo 37e/98 (23 September 1998)
©European Southern Observatory



W normal, (std. 0.1)



W Laplace

Bissantz et al.’07, JRSSB