Total Variation and Tomographic Imaging from Projections

Per Christian Hansen & Jakob Heide Jørgensen

Technical University of Denmark

Joint work with

- Dr. Emil Sidky & Prof. Xiaochuan Pan Univ. of Chicago
- Tobias Lindstrøm Jensen Aalborg University, Denmark

DTU Informatics

Department of Informatics and Mathematical Modeling

Our "Tool": Total Variation

Total Variation is a well-known mathematical and computational tool for image reconstruction.

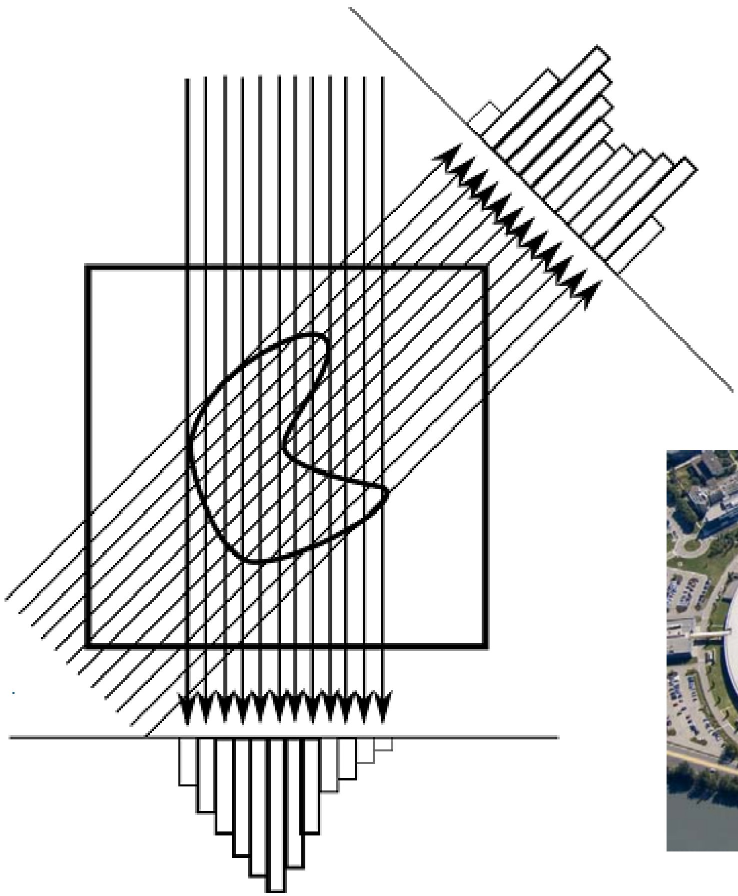
Example: image in-painting.

We will discuss its application in tomographic imaging.

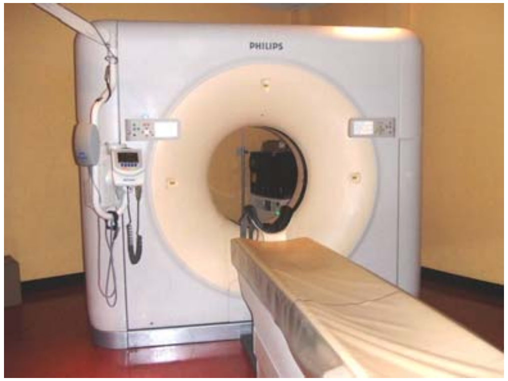
Our Problem: Tomography

Image reconstruction

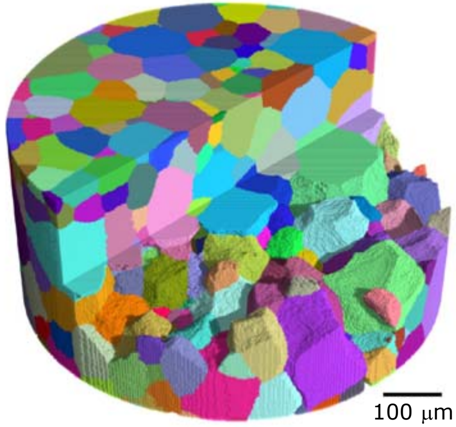
from projections



Medical scanning



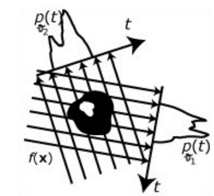
Mapping of metal grains



The Origin of Tomography

Johan Radon, *Über die Bestimmung von Funktionen durch ihre Integralwerte Längs gewisser Mannings-faltigkeiten*, Berichte Sächsische Akadamie der Wissenschaften, Leipzig, Math.-Phys. Kl., 69, pp. 262-277, 1917.

Main result: An object can be perfectly reconstructed from a full set of projections.



NOBELFÖRSAMLINGEN KAROLINSKA INSTITUTET
THE NOBEL ASSEMBLY AT THE KAROLINSKA INSTITUTE

11 October 1979

The Nobel Assembly of Karolinska Institutet has decided today to award the Nobel Prize in Physiology or Medicine for 1979 jointly to

Allan M Cormack and Godfrey Newbold Hounsfield

for the "development of computer assisted tomography".

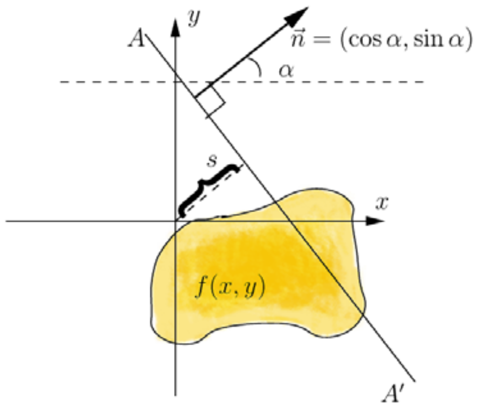
Reconstruction via “Analytic” Transform

The medical CT reconstruction method of choice is often the Inverse Radon Transform, implemented as the **Filtered Back Projection** or **Feldkamp-Davis-Kreiss** methods (2D & 3D).

Implementation: FFTs + low-pass filtering + interpolation.

Advantages:

- Fast – relies on the FFT
- Low memory requirements



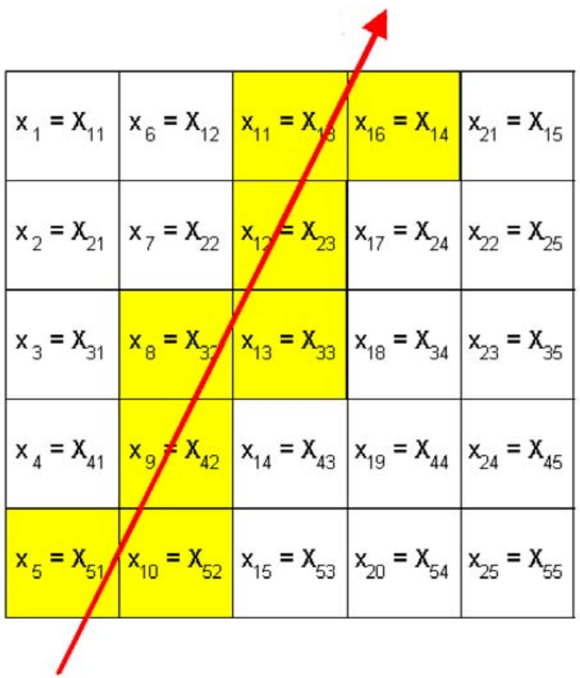
Drawbacks:

- Needs lots of data for accurate images
- Difficult to incorporate requirements on the reconstruction

Reconstruction via an Algebraic Model

Damping of *i*-th X-ray through domain:

$$b_i = \int_{\mathrm{ray}_i} \chi(\mathbf{s})\, d\ell, \qquad \chi(\mathbf{s}) = \text{attenuation coef.}$$



Discretization leads to a large, sparse, ill-conditioned system:

$$A\, x = b \qquad b = \bar{b} + e$$

Geometry

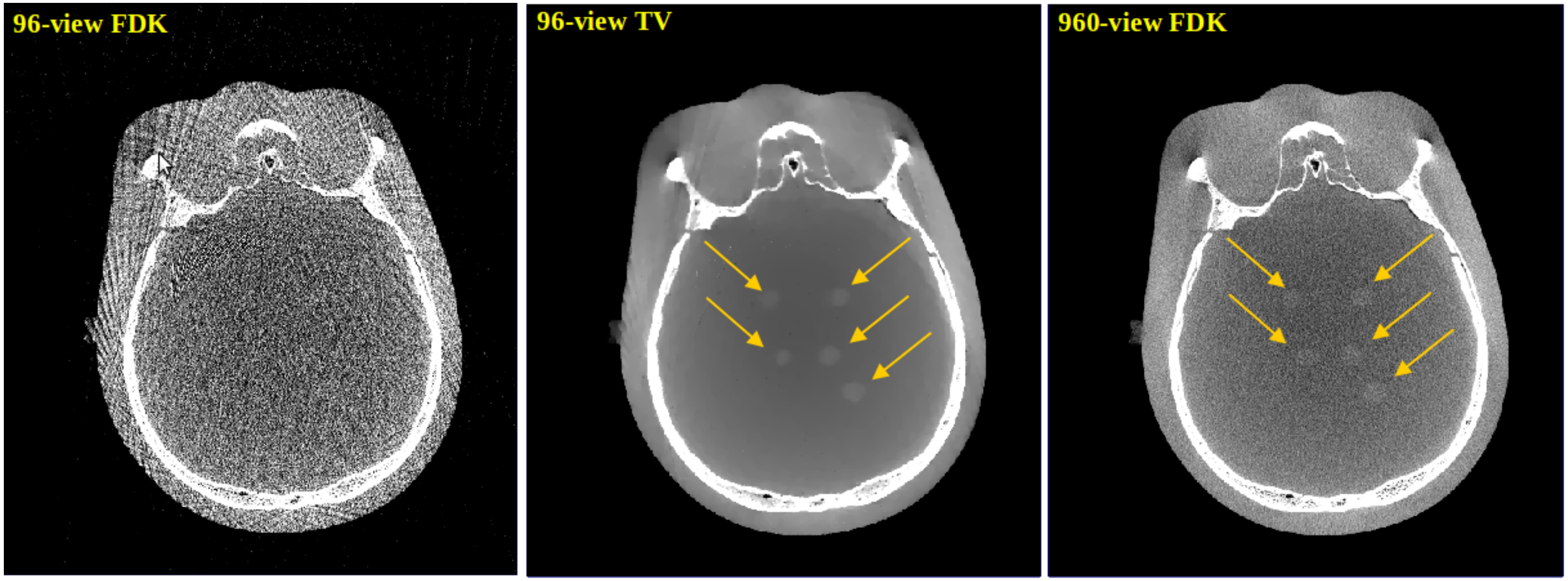
Image

Projections

Noise

$$\bar{b} = A\, \bar{x}$$

Comparison (Cone Beam + Head Phantom)



J. Bian et al., Phys. Med. Biol. 55 (2010), 6575

Analytic reconstruction (e.g., FBP, FDK):

- Fast, limited memory, many years of experience.
- May need lots of data.

Algebraic reconstruction (e.g., TV):

- Potential for better reconstructions.
- Slow, memory demanding, limited understanding of parameters

Tomographic Imaging = Inverse Problem

Inverse problems are *ill posed*.

Discretizations of inverse problems are *ill conditioned*.

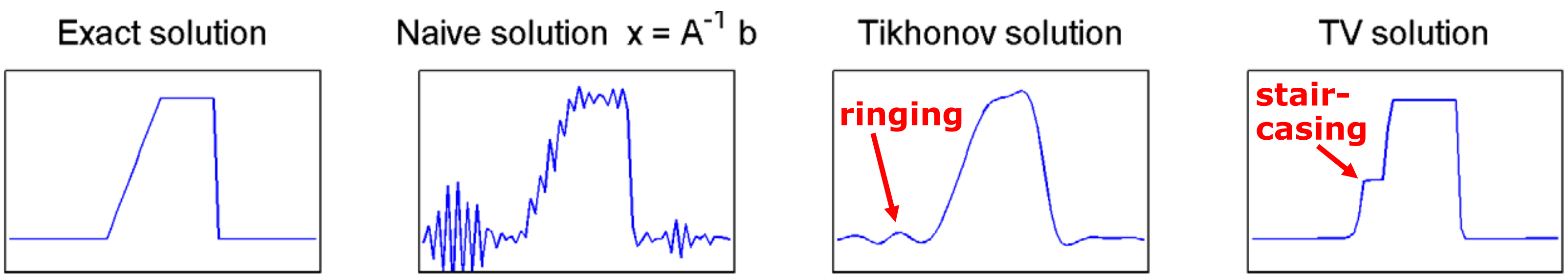
We must use *regularization* to define/compute a stable solution.

No regularization:

- reconstruction has artifacts (often severe) from the noise.

Regularization:

- reduces noise artifacts but introduces other types of artifacts!



Regularization = Incorp. of Prior Information

The regularized problem is an optimization problem of the form

$$\boxed{\min_{x \geq 0} f(x), \qquad f(x) = \tfrac{1}{2}\|A\,x - b\|_2^2 + \Omega(x).}$$

The regularization function expresses *prior information* about the solution. 1-D examples:

Solution is smooth (*compressible* repr. in spectral basis):

$$\Omega(x) = \|x\|_2^2$$

$$\Omega(x) = \|D\,x\|_2^2 \qquad D = \text{some deriv. op.}$$

Solution is piecewise smooth (*sparse* repr. of the gradient):

$$\Omega(x) = \|D_1 x\|_1 \qquad D_1 = 1.\ \text{deriv. op.}$$

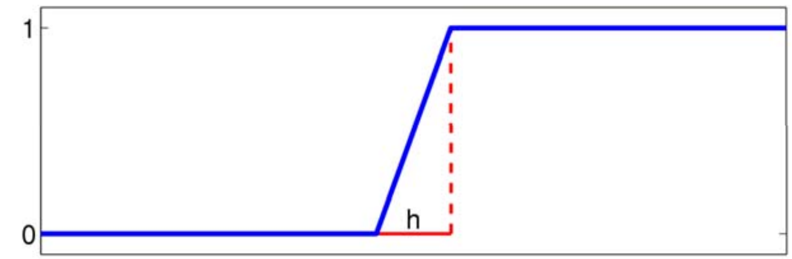
Total Variation Allows Steep Gradients

1-D continuous formulation:

$$TV(g) = \|g'\|_1 = \int_\Omega |g'(t)|\, dt$$

Example (2-norm penalizes steep gradients, TV doesn't):

$$\begin{aligned} TV(g) &= 1 \\ \|g'\|_2^2 &= \int_\Omega g'(t)^2 dt = 1/h \end{aligned}$$



2-D and 3-D continuous TV formulations:

$$TV(g) = \big\| \, \|\nabla g\|_2 \, \big\|_1 = \int_\Omega \|\nabla g(\mathbf{t})\|_2 \, d\mathbf{t}$$

TV Produces a Sparse Gradient Magnitude

Underlying assumption or prior knowledge: the image consists (approx.) of regions with constant intensity.

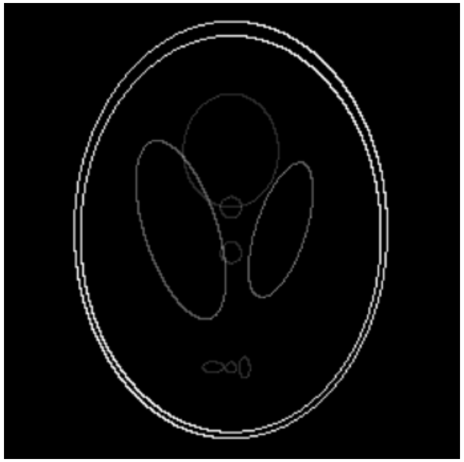
Hence the gradient magnitude (2-norm of gradient in each pixel) is sparse.

image

TV = 1-norm of the gradient magnitude,

= sum of 2-norm of gradients.

gradient magnitude



3 % non-zeros

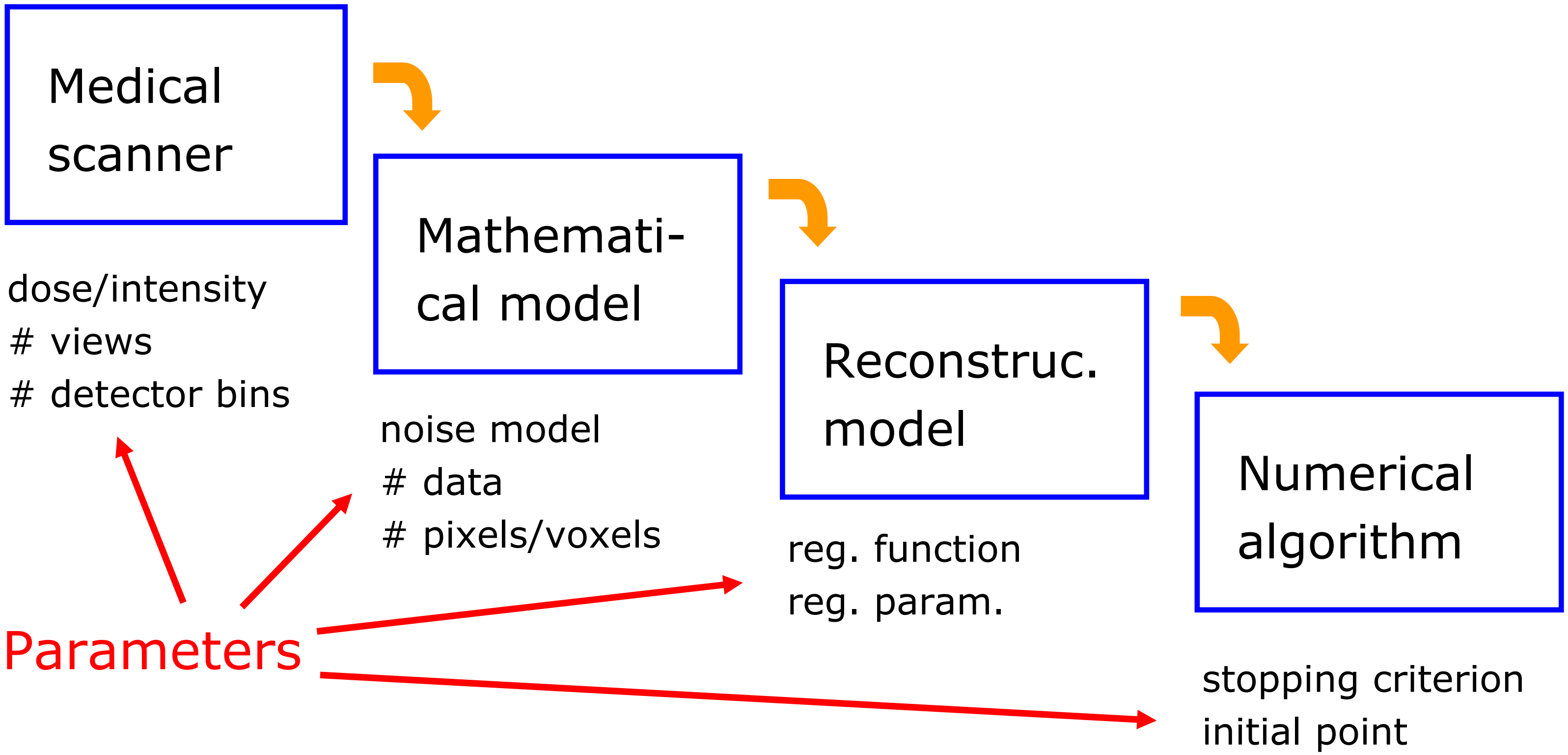
Experience shows that the TV prior is often so “strong” that it can compensate for a reduced amount – or quality – of data.

This talk: a closer study of this claim.

Is TV Really Better? No Simple Answer!

Is the TV prior so strong that it can compensate for a reduced amount – or quality – of data?

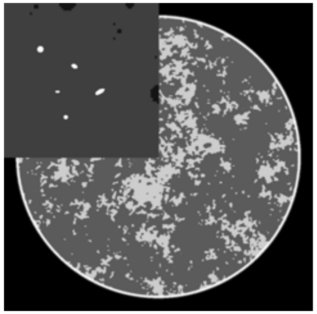
A careful study must consider all steps in the solution process.



Case Studies – Micro-Calcifications in Breasts

Simulated test image: cross section of a female breast, four tissue types with different gray level intensities

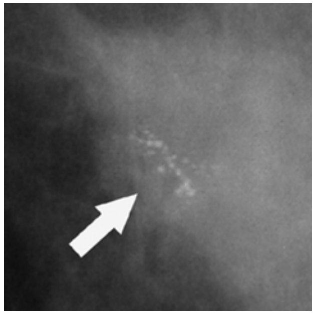
1. skin,
2. fat,
3. fibro-glandular tissue (having a realistic complex structure),
4. micro-calcifications.



Our simulated image, with ”region of interest” inserted

Micro-calcifications $\rightarrow$ an early indicator of a developing cancer.

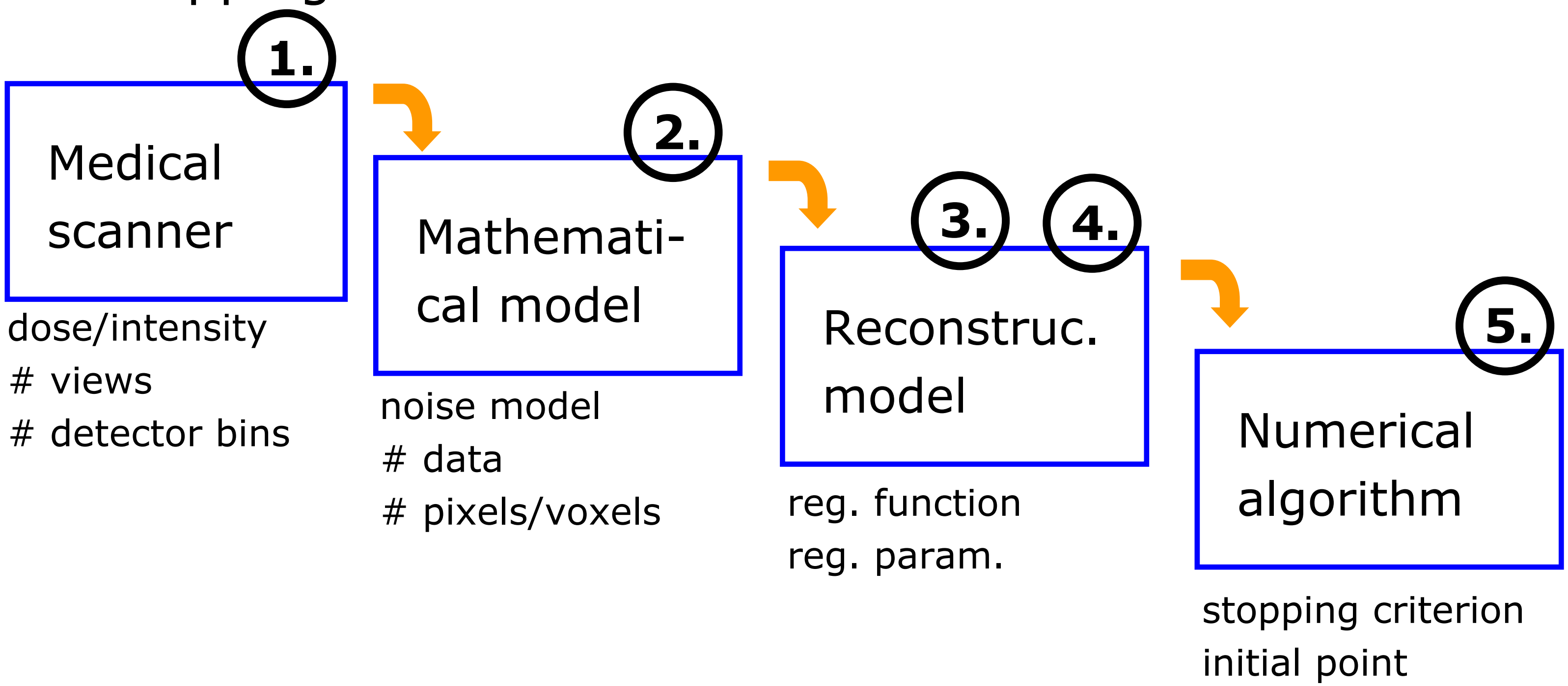
Their tiny size and high contrast make accurate imaging a challenge.



© Geneva Foundation for Medical Education and Research

Simulation Studies

1. Dose and number of views
2. Number of views and bins on detector
3. TV versus 2-norm regularization
4. Regularization parameter
5. Stopping criterion



1: Dose and Number of Views

Max accumulated X-ray dose:

- safety requirements in a medical scan,
- material limitations in nondestructive testing.

The product of dose intensity and #views is constant.

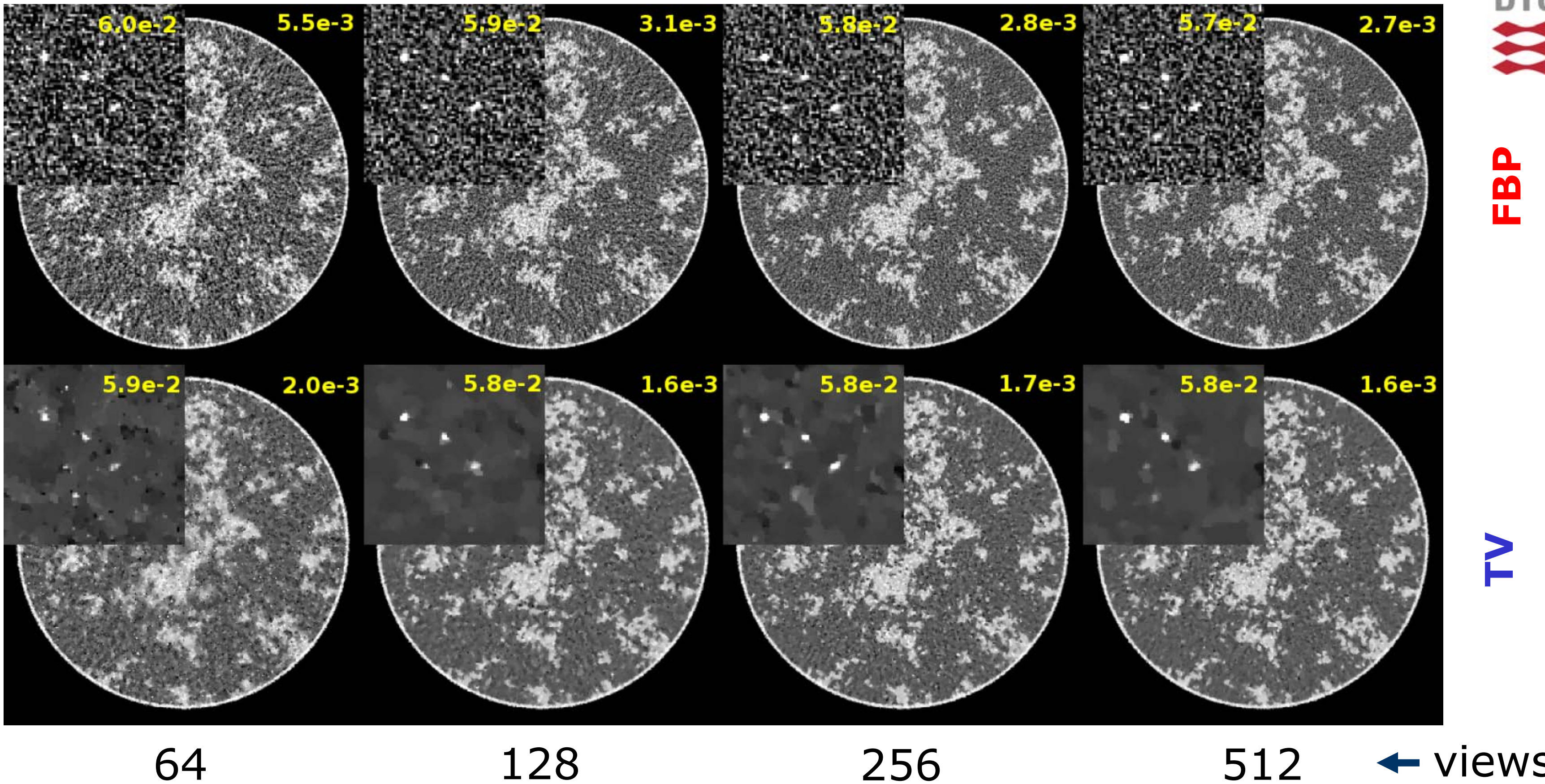
Measurements are photon counts – *Poisson statistics:*

$$\mathrm{SNR}_{\mathrm{pixel}} = \frac{\mathrm{count}}{\mathrm{st.dev.}} \approx \frac{\mathrm{count}}{\sqrt{\mathrm{count}}} = \sqrt{\mathrm{count}},$$

i.e., high dose $\rightarrow$ high SNR.

What is better:

- few views with high dose and high SNR, or
- many views with low dose and lower SNR?



FBP: Results improve slightly with #views; lots of HF-structure noise.

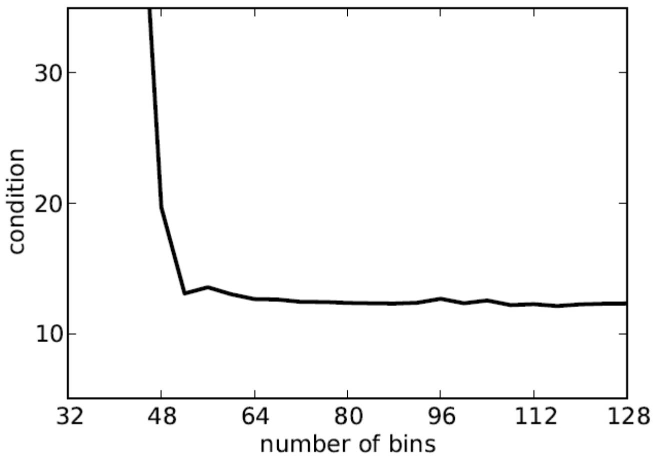
TV: Visual appearance varies significantly with #views; cartoon artifacts dominate. As the SNR deteriorates:

- size of piecewise-constant regions increases,
- while their number decreases.

Conclusion: very different images!

2: Number of Views and Bins on Detector

The condition number reflects the "difficulty" of the problem.



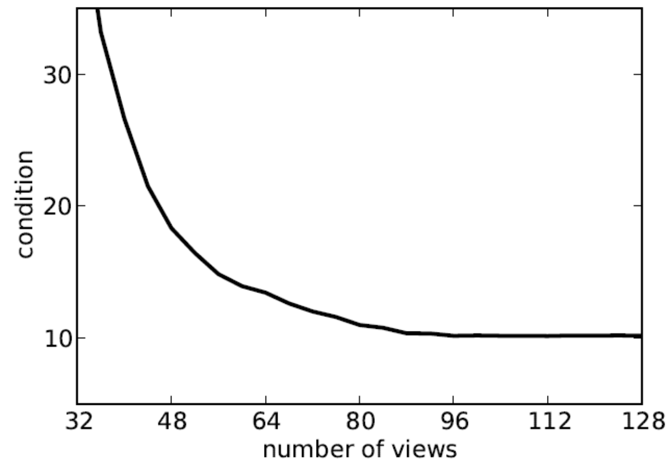


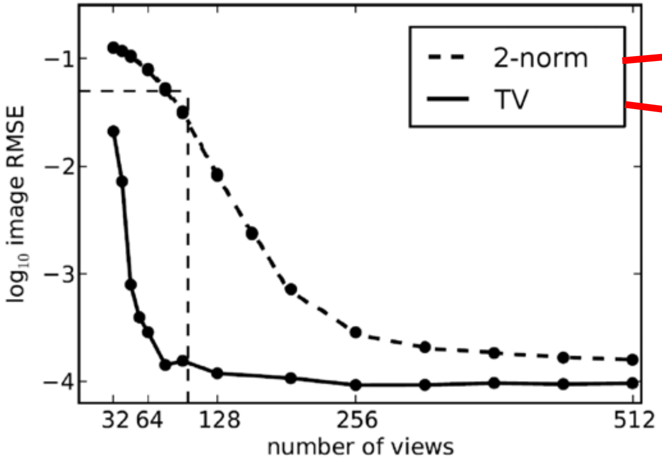
Image is $N \times N$ with $N = 32$.

Our observations for TV reconstruction:

- cond(A) decays fast with increasing #bins and slower with #views.
- Suggests the choice <u>#bins ≅ 2N</u> & <u>#views ≅ 2N</u> for this problem.
- Increasing parameters further reduces cond(A) marginally.

Conclusion: "small" no. of views or bins gives well-cond. problem.

3. TV versus 2-Norm Regularization



$$\Omega(x) = \|x\|_2$$

$$\Omega(x) = \textstyle\sum_i \|D_i\, x\|_2$$

RMSE = || exact image − reconstruction $||_2$

Two models → two different types of artifacts
→ RMSE does not tell the whole story.

Main observations:

- TV model gives much lower RMSE than the 2-norm model as #views decreases.
- RMSE for TV is almost independent of #views as long as #views > 100, and increases dramatically for fewer views.
- RMSE for 2-norm increases steadily as #views decreases.

Conclusion: TV represents a strong prior which is able to compensate for the reduction in the amount of data.

4. The TV Regularization Parameter

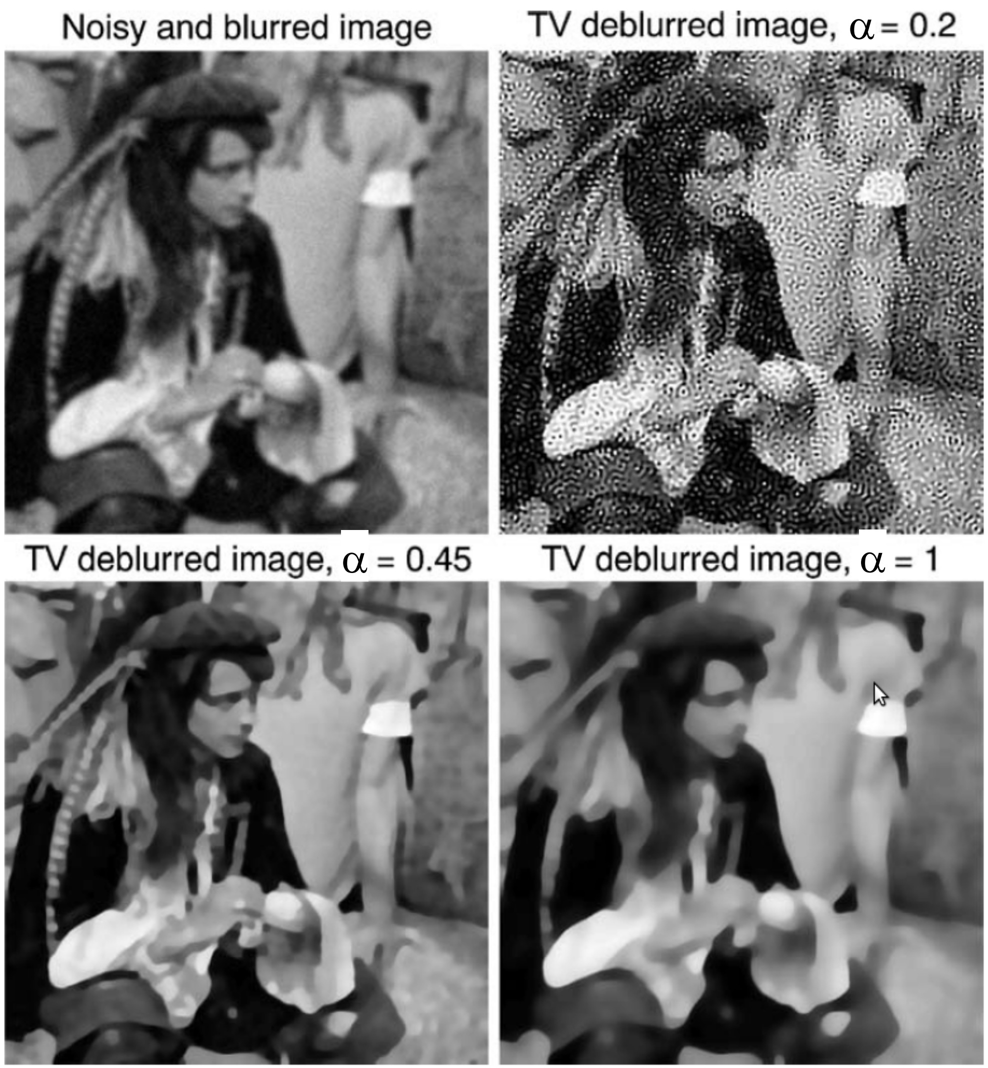


Image deblurring example illustrates the role of the regu-larization parameter.

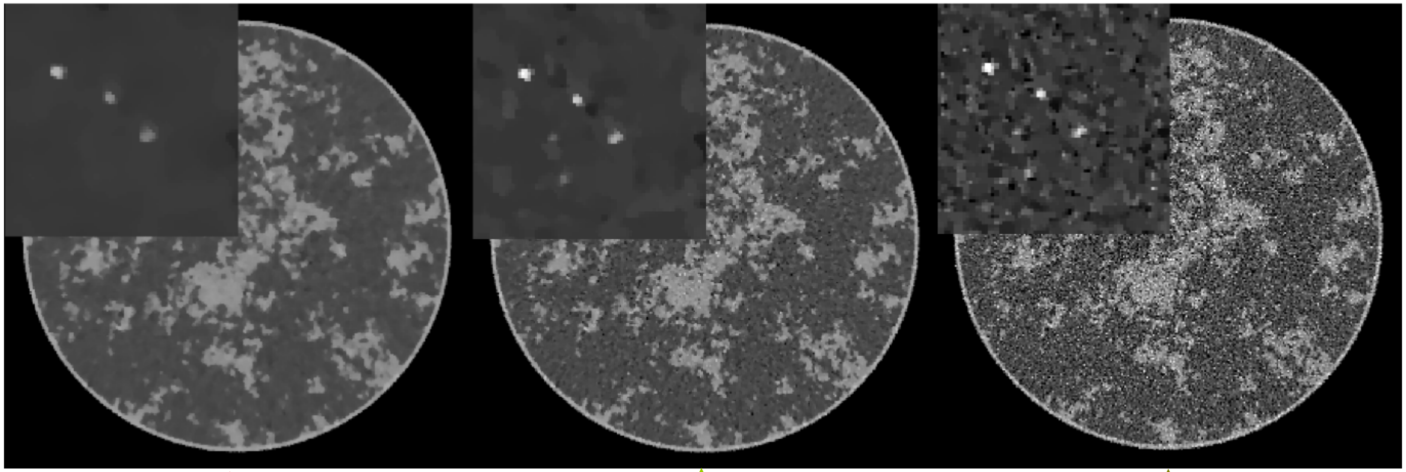
Small α:

- noise dominates

Large α :

- cartoon artifacts

4a. The Influence of α on the Reconstruction



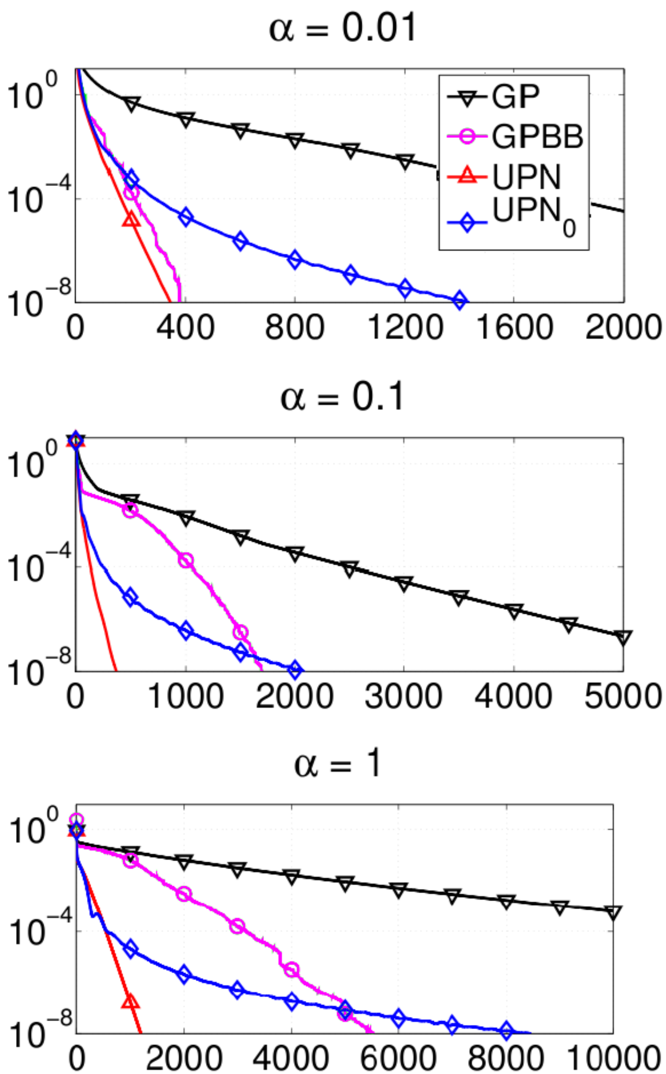
α is too large: the inverted noise is suppressed but the regions of constant intensity are too large.

α is too small: $f(x)$ is dominated by the residual norm, and the solution is dominated by inverted noise.

Many details without being influenced by the noise.

Conclusion: must choose a good α.

4b. The Influence of α on the Problem



GP Standard gradient projection algorithm.

GPBB GP with Barzilai-Borwein acceleration.

$\mathbf{UPN_0}$ Optimal first-order method.

UPN Ditto that exploits strong convexity.

The four methods differ by the amount of information about the problem they exploit.

$\mathbf{UPN_0}$ and **UPN** estimate the needed information during the iterations.

As α increases, the TV regularization term in $f(x)$ becomes increasingly important and the problem becomes more difficult to solve.

$\rightarrow$ increasing no. of iterations for all methods.

Conclusions:

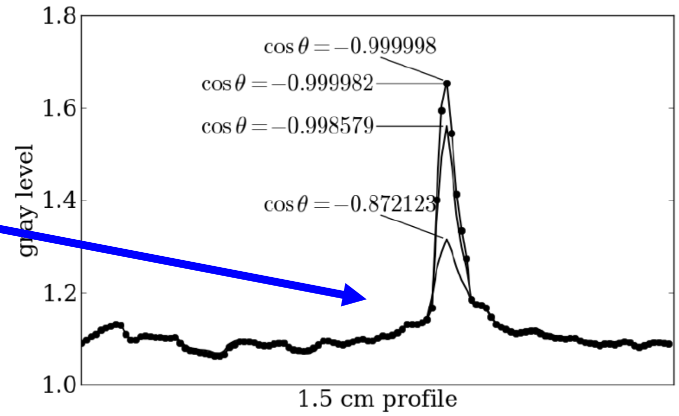
- GPBB is always superior to GP.
- For large α (harder problems) the optimal first-order methods are even faster.

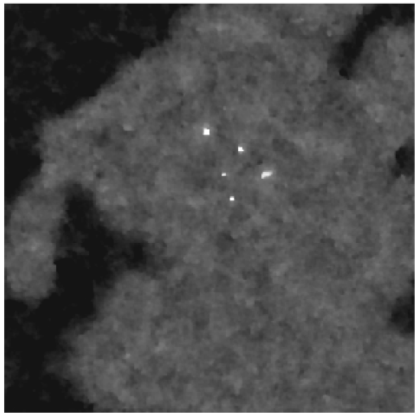
5. The Stopping Criterion

Angle θ between $\nabla_x \|A\,x - b\|_2^2 = A^T(A\,x - b)$ and $\nabla_x \Omega(x)$.

At the solution we have $\cos\theta = -1$.

Profile through a single micro-calcification





- As the number of iterations increases the sharp peak gets better resolved.
- Low-frequency components are captured after a small number of iterations.
- Many more iterations are needed to capture the peak's shape and magnitude.

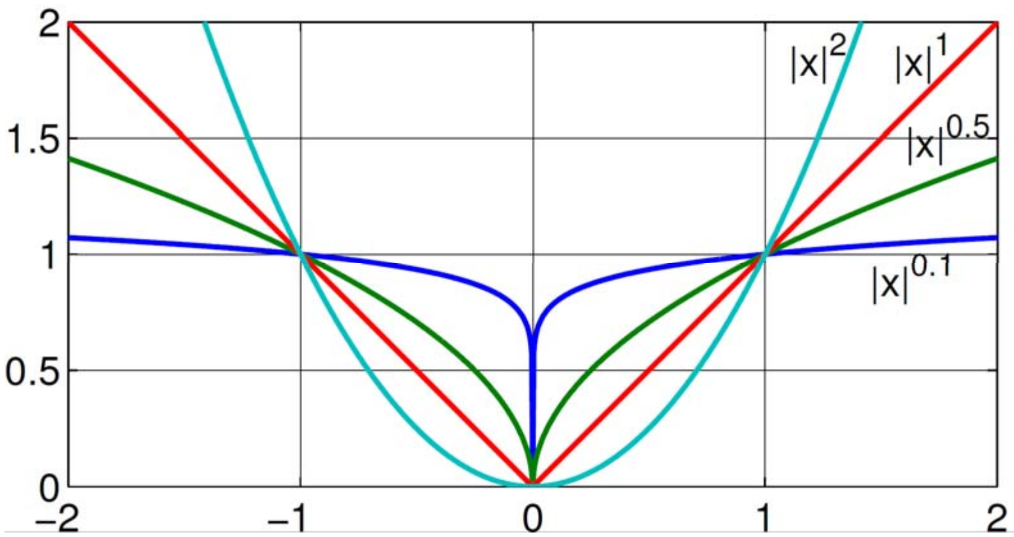
Conclusion: The TV model focuses on providing an accurate representation of the gradient, and it is important to be close to the minimum of $f(x)$.

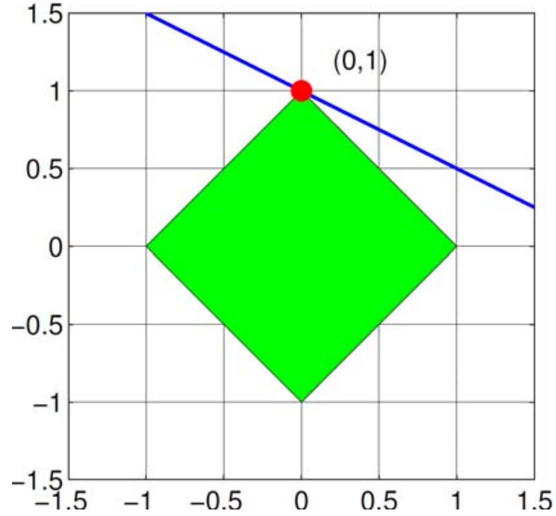

That's all Folks!

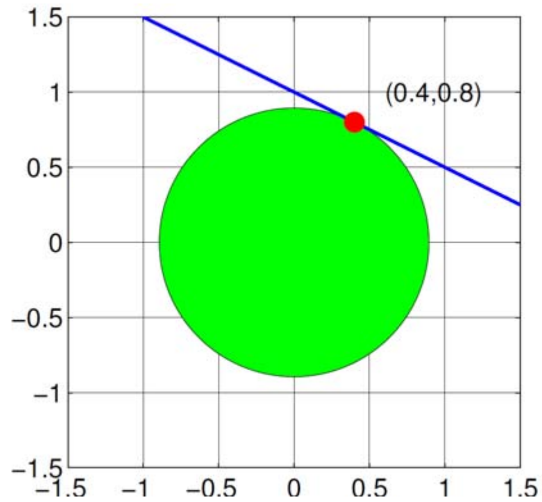
1-norm: Enforcing sparsity

- Last decade: Explosion in signal proc. Work using L1-norm as sparsity inducing prior, e.g. Compressed Sensing, Sparse Approximation.

- Core idea: Use L1-norm penalty on the signal or some transform or function of the signal.
- p smaller than 1: More sparsity enforcing
- p larger than or equal to 1: Convex opt. prob.
- p = 1: Good compromise.







"Nitty Gritty" Details of TV

Discrete TV is a sum of 2-norms. For the 2-D case:

$$TV(x) = \sum_{\text{pixels}} \|D_i\, x\|_2,$$

$D_i \in \mathbb{R}^{2\times n}$ is a difference approximation to the gradient at pixel i.

The Euclidean norm is not differentiable at the origin; use a smooth version, e.g., the Huber function:

$$TV_\tau(x) = \sum_{\text{pixels}} \phi_\tau(\|D_j\, x\|_2),$$

$$\phi_\tau(z) = \begin{cases} |z| - \frac{\tau}{2} & \text{if } |z| \geq \tau \\ \frac{1}{2\tau}\,|z|^2 & \text{else.} \end{cases}$$