present

```
1 Enter cell code...
```

Gentle introduction to inverse problems

Václav Šmídl, et. al.

1. Faculty of Electrical Engineering, University of West Bohemia, Pilsen
2. Institute of Information Theory and Automation, Prague

First steps...

The first encounter with an inverse problem is at primary school:

$$6 = 2x$$

What is the value of x?

- Later, you will learn that division is the inverse operation of multiplication.

$$x = 6/2$$

- Even later, you will learn that inverse function f^{-1} undoes operations of function f,

$$x = f^{-1}(f(x))$$

- In multi-dimensional problems, we have inverse matrices

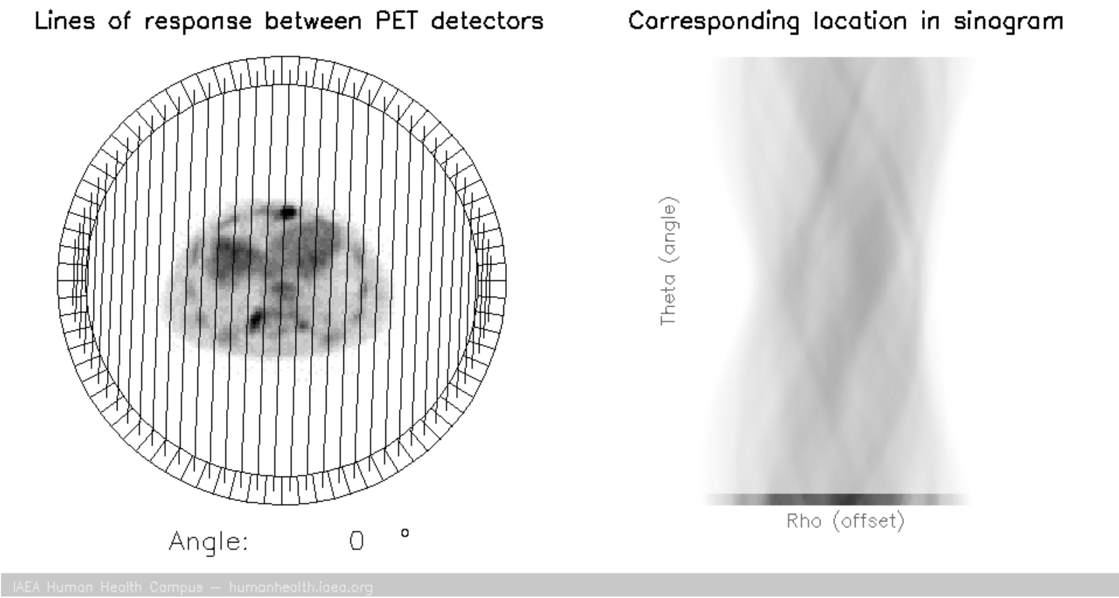
$$x = A^{-1}Ax$$

There is always a 'probelm'!

- division by zero, $6 = 0x$, $0 = 0x$, bijective, low-rank matrix A
- your teacher told you not to use inversion under such conditions

Why? Applications

Computer tomography



Blind Image Debluring

- you take a blurry image and you want to recover the sharp image,
- mathematical operation is convolution,
- blind means that we do not know how it was deblured,

Change in mathematics:

1. Inversion

$$y = f(x) \quad \Rightarrow \quad x = f^{-1}(y)$$

Allow for an error in the model:

$$y = f(x) + e$$

Minimize the error

2. Optimization approach:

$$x^* = \arg\min_x D(y, f(x)), \qquad D(y, f(x)) = ||y - f(x)||,$$

3. Probabilistic approach:

$$x^* = \arg\max_x p(y|f(x))$$

Close relation of optimization and probability:

Based on two functions: log and exp

Gaussian distribution of probability:

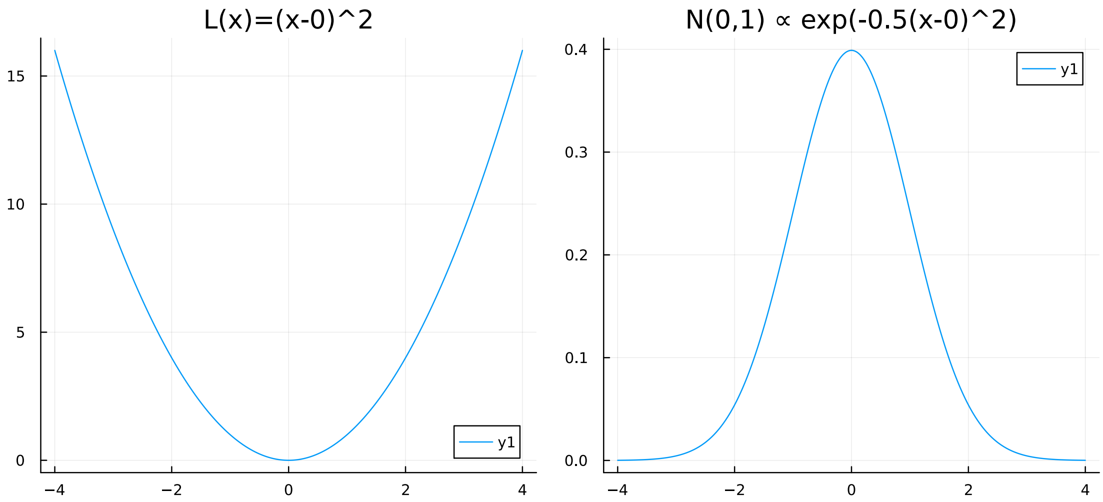
$$p(y|x) = N(x,1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-y)^2\right)$$

Its maximum is found by optimizing the negative log-likelihood:

$$y^* = \arg\max p(y|x) = \arg\min -\log p(y|x) = \arg\min (x-y)^2$$

And also the other way around:

$$y^* = \arg\min D(y,x) \qquad \Rightarrow \qquad p(y|x) \propto \exp(-D(y,x))$$



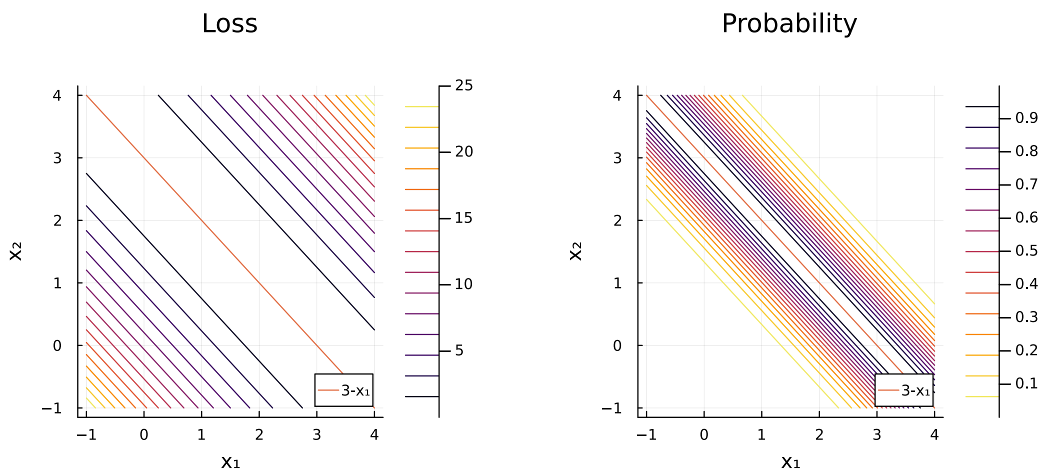
Linear Inversion model: Toy

Consider a toy example:

$$3 = x_1 + x_2 \qquad = [1,1]\mathbf{x}$$

being solved as

$$x^* = \arg\min (3 - (x_1 + x_2))^2$$



$\log_{10}\alpha$: -8, p: 2.0, surface:

Additional information

add *regularization* function to the loss:

$$x^* = \arg\min D(y, f(x)) + \alpha g(x)$$

where α is a coefficient of the penalization function $g(x)$.

The regularization adds loss to any solution that we do not want. The common choice is Lp norm:

$$g(x) = ||x||_p^p$$

With important special case:

- $p = 2$: essentially the Euclidean norm : $g(x) = x^\top x$

  Solutions are penalized for vector amplitude Also known as:
  - Tichonov
  - Ridge regression
  - weight decay (NN, ADAMW)
- $p = 1$: $g(x) = \sum_i |x_i|$

  penalizes absolute value for each dimension
  - leads to sparse solution (as many zeros as possible)
  - LASSO
- $p < 1$: for $p < 1$ is is nonconvex (super Gaussian)
  - sparsifies aggresively (approaches L0 norm)

The same works in probabilities

- Square distance from the origin:

$$g(x) = (x_1^2 + x_2^2) \quad \Rightarrow \quad p(x) = N(0, \alpha^{-1}I)$$

- generally any p-norm: $p(x) = GN_p(0, \alpha I)$

Yeah - almost the same, right?

$$g(x) = -\log p(x)$$

We can do more:

- get rid of α

$$p(x|y) = \int p(x|y, \alpha)p(\alpha)d\alpha$$

- analytical solutions are available only in a few cases (Gaussian, Dropout,....)
- Monte Carlo Techniques
- Variational Lower Bound (ELBO)

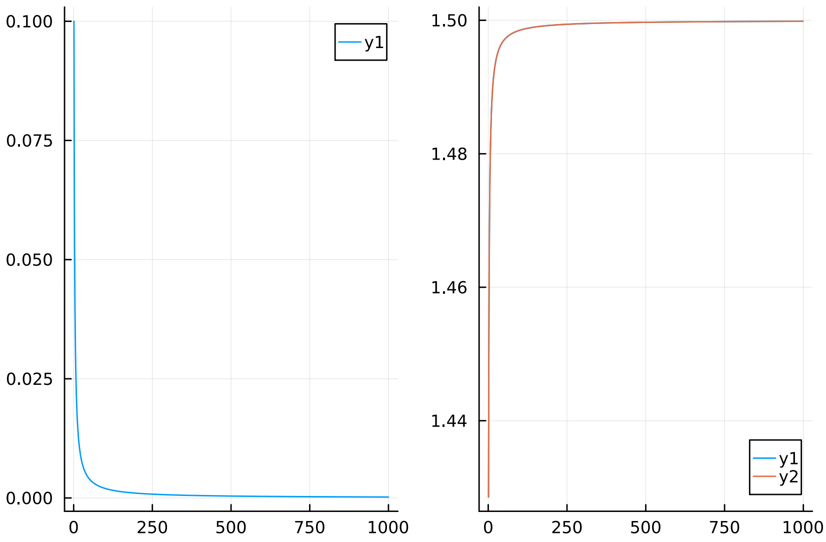
$$p(x|y) \geq E_{q(\alpha|x)}\left[\frac{p(x, \alpha)}{q(\alpha|x)}\right]$$

Variational Bayes demo

For unknown α in ridge regression, optimizing elbo results in the iterated algorithm:

- set α_0, e.g. $\alpha_0 = 0.1$

1. Solve OLS: $x = (A^\top A + \hat{\alpha}_i I)^{-1}(A^\top y)$
2. Set new regularization: $\hat{\alpha}_{i+1} = 2/(\hat{x}^\top \hat{x} + \text{tr}(S))$ where $S = (A^\top A + \hat{\alpha}_i I)^{-1}$



Variational Sparse regression

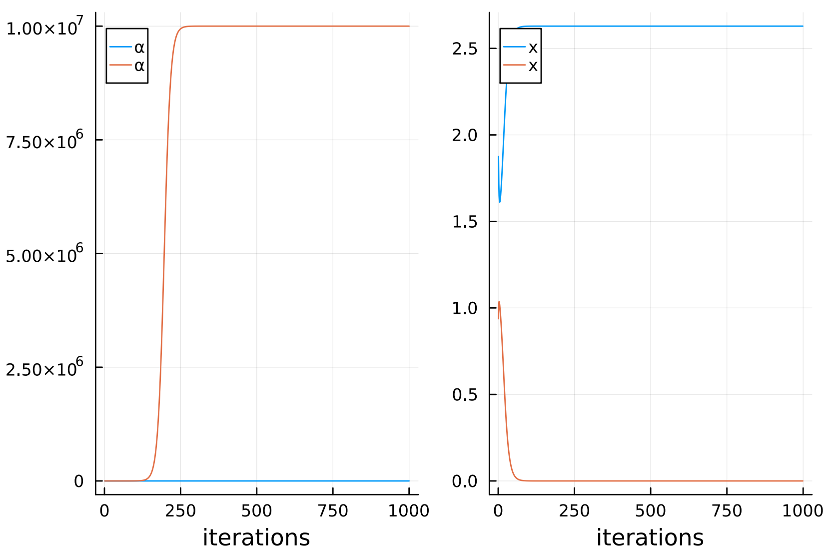
A very simple change allowing α_1 for x_1 and α_2 for x_2

$$x^* = \arg\min(y - (x_1 + x_2)^2) + \alpha_1 x_1^2 + \alpha_2 x_2^2$$

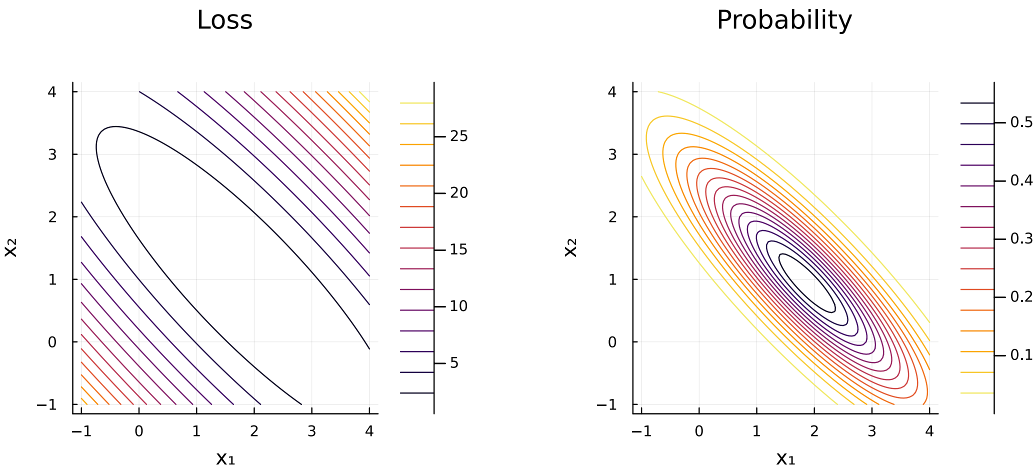
Optimizing elbo results in the iterated algorithm:

- set α_0, e.g. $\alpha_0 = 0.1$

1. Solve OLS: $x = (A^\top A + \text{diag}([\hat{\alpha}_1, \hat{\alpha}_2]))^{-1}(A^\top y)$
2. Set new regularization: $\hat{\alpha_1} = 1/(\hat{x}_1^2 + S_{1,1})$, and $\hat{\alpha_2} = 1/(\hat{x}_2^2 + S_{2,2})$



Step: 1



Toy example of multiplicative model

Multiplicative noise:

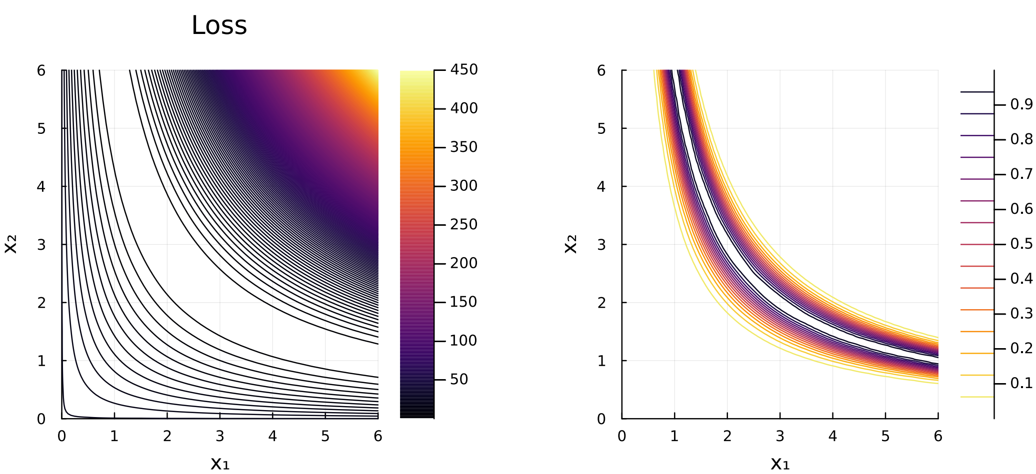
$$y = x_1 x_2$$

The problem is clearly ill-posed, since any transformation

$$x_1 x_2 = (x_1 t)(x_2 / t)$$

yields the same measurement.

$\log_{10}\lambda$: -3



People prefer natural numbers.

Penalize differences from natural numbers:

$$g(x) = (x - \text{round}(x))^2$$

Blind Image Debluring

A sharp image u is observed through a convolution operator h, yielding a blurred image, g:

$$g = u \circledast h$$

which can also be written in matrix forms: $g = Uh = Hu$.

Both u and h are unknown. (A fancy version of multiplicative decomposition).

Common prior: sparse differences of images $||\nabla u||_p$

1. Classical MAP approaches had to tune (schedule) regularization coefficients.
2. Variational Bayes managed to tune it automatically.
    - could also adjust space varying model error [1]

[1] Kotera, J., Šmídl, V. and Šroubek, F., 2017. Blind deconvolution with model discrepancies. IEEE transactions on image processing, 26(5), pp.2533-2544.

Take home message

- inverse problems are very common
- solved as an optimization probelm (= probabilistic)
- solution has to have two parts:
    1. additional formulation of preferences (regularization / prior)
    2. optimization-friendly way to define it