Inverse Problems 52506: Project work description

Mikko Kaasalainen and Sampsa Pursiainen, Tampere University of Technology

May 7, 2015

The goal of the project work is to examine different approaches and aspects of the reconstruction of the underlying model from noisy observations simulated by, e.g., a convolution-type measurement model.

Choose one of the following four cases. In the first two, the idea is to create the setup and inversion programs from scratch for a proper understanding of the problem. The measurement models are linear, so you can use the regularization techniques of Part I (many of which are particularly simple to code using ready-made Matlab functions). You can start by adding noise to the measurements and noting that the problem matrix is ill-posed. Then you can try, e.g., SVD, Tikhonov regularization, or the Bayesian version of linear Gaussian models (Appendix; choose a suitable a priori correlation matrix). All of these are simple matrix computations.

The last two are related to the statistical inversion techniques of Part II. The instructions are more detailed, so the problems are more straightforward than the first two; on the other hand, ”more work” is involved in the computations. In the first two, your own setup and design requires more work than the actual computations.

1. Convolution in $\mathbb{R}^2$: design original signals and convolution masks, create data with various noise levels, and then try reconstruction schemes (for formulation see lecture notes, Part I)
2. Tomography in $\mathbb{R}^2$: design a density distribution in some domain, create various tomographic line integrals and add noise, then reconstruct the distribution (for formulation see lecture notes, Part I)
3. (Part II) Filtering. Let $\mathbf{z}$ be an image including *a priori* well-localized objects, for example pointlike stars, to be recovered and let $\mathbf{y}$ be a blurred image obeying a linear model $\mathbf{y} = \mathbf{A}\mathbf{z} + \mathbf{n}$ with an anisotropic blurring operator $\mathbf{A}$ (moving object blur) and additive and independent Gaussian noise $\mathbf{n}$. Use a hierarchical conditionally Gaussian prior with gamma and/or inverse gamma hyperprior to deblur the original image $\mathbf{z}$. Try recovery in a filtered form in which $\mathbf{z}$ is related to the unknown $\mathbf{x}$ of the inverse problem via (a) $\mathbf{z} = \mathbf{B}\mathbf{x}$ and (b) $\mathbf{z} = \mathbf{I}\mathbf{x}$ with $\mathbf{B}$ denoting an isotropic blur (slight one) and $\mathbf{I}$ is an indentity matrix. Explore the results yielded by a *maximum a posterior* (MAP) estimate and/or the conditional mean (CM) with low and high noise as well as suitably

chosen values of the scaling parameter θ_0. Set the shape parameter β equal to 1.5.

4. (Part II) Metropolis-Hastings (M-H) transition rules. Finding a M-H transition function $T(x, y)$ that fits to both the local and global dynamics of the target (invariant) density π can be somewhat challenging. Here, the aim is to compare using Matlab the performances of the following three random-walk transition rules:

   (i) $y = x + \sigma w$ (standard),

   (ii) $y = x + \sigma_\ell w$ with ℓ randomly 1 or 2 (variable step length),

   (iii) $y = x + (\sigma^2/2)\nabla(\log \pi(x)) + \sigma w$ (Langevin algorithm).

   In all of these, w is picked from $\mathcal{N}(0, I)$. Apply the rules (i)–(iii) to find the mean of both

   (a) the tri-modal density $\pi(x) \propto \sum_{i=1}^{3} \exp[\frac{1}{2\sigma^2}\|x - c_i\|^2]$ with $\sigma = 7$, $x = (x_1, x_2)$, $c_1 = (1, 2.5)$, $c_2 = (3, 4)$, and $c_3 = (4, 1)$;

   (b) the posterior density $\pi(x \mid c)$ in which $x = (x_1, x_2)$, $x_1 > 0$, $x_2 > 0$ is a vector of chemical reaction rate constants and y is a data vector containing observations of a single reactant concentration as a function of time. The data prediction (forward) model for the concentration is $c(t) = 1 + \exp(-x_1 t) + \exp(-x_2 t) + n(t)$ in which $t \geq 0$ is time and $n(t)$ is a realization of a zero-mean Gaussian noise term with a time dependent standard deviation $\sigma_{noise}(t)$. The observation vector is given by $y = (c(1), c(2), c(3)) = (1.8, 1.4, 1.2)$ and the corresponding standard deviations by $(\sigma_{noise}(1), \sigma_{noise}(2), \sigma_{noise}(3)) = (0.2, 0.1, 0.05)$. Use as a prior a Gaussian density with the mean $(1.5, 0.7)$ and covariance $\sigma_{pr}^2 I$. Try both $\sigma_{pr} = 0.5$ and $\sigma_{pr} = 1.5$.

   In each test, generate a sample ensemble of at least 100000 points. Compare the sample-based mean to a reference value obtained via standard numerical integration by e.g. measuring the distance between those points. Visualize in a single image the target density, sample points as well as the reference and the sample-based estimate for the mean. Illustrate and analyze also sampling histories. Base your conclusions on more than just one test run per transition rule. Choose the proposal standard deviations so that the acceptance rate will be between 25 and 35 % for the rules (i) and (ii), and around 57 % for the Langevin algorithm (iii). Notice that in Langevin algorithm, utilizing the gradient of $\log \pi$, the transition function is asymmetric, i.e. $T(x, y) \neq T(y, x)$. Which method seems to be the winner? Discuss other possible situations in which the random walks (i)–(iii) might perform well or poorly, e.g. regarding the dimensionality of the target distribution. Consider also how appropriate are the choices of σ_{pr} in (b) based on the resulting conditional mean estimates.

Report

The project work report (≥ 7 pages) should be organized in the standard form of a scientific study/paper, including at least the following sections:

- A short introduction and background
- Your problem and simulation setup
- Methods and results
- Analysis of results, conclusions, discussion
- References

Use any software packages and programming languages you like. Use your imagination: just about all "typical" theoretical and practical aspects of inverse problem solving are encountered in these problems. Approach this as something studied for the first time: what would we like to know? What approaches should we try? Plot, for example,

- your solutions vs. original signal
- model fits vs. data.

You can get up to 12 extra points with this work. The project can be done in groups of ≤ 3 people.

Appendix: Linear Bayes inference

If we assume Gaussian distributions, our linear model

$$y = Ax + \epsilon \tag{1}$$

yields a simple maximum likelihood estimate $\hat{x}$ for the posterior distribution. Let our a priori distribution be concentrated around x_0 (with tightness and correlation given by the matrix Σ_0):

$$p_{pr}(x) \sim \exp[-\frac{1}{2}(x - x_0)^T \Sigma_0^{-1}(x - x_0)]. \tag{2}$$

On the other hand, the conditional distribution of the measurement is

$$p(y|x) \sim \exp[-\frac{1}{2}(y - Ax)^T \Sigma_1^{-1}(y - Ax)], \tag{3}$$

where Σ_1 is the covariance matrix of the measurements ($\Sigma_{ij} = \langle \epsilon_i \epsilon_j \rangle$). Then Bayes says

$$p(x|y) \sim p_{pr}(x)p(y|x) \tag{4}$$

so we obtain

$$p(x|y) \sim \exp\{-\frac{1}{2}[x^T Q x - 2x^T Q Q^{-1}(\Sigma_0^{-1} x_0 + A^T \Sigma_1^{-1} y) + ...]\} \tag{5}$$

where the Fisher information matrix is

$$Q = (\Sigma_0^{-1} + A^T \Sigma_1^{-1} A). \tag{6}$$

Thus our posterior distribution is the exponential quadratic form

$$p(x|y) \sim \exp[-\frac{1}{2}(x - \hat{x})^T Q (x - \hat{x})] \tag{7}$$

with the centre of the posterior distribution $\hat{x}$ at

$$\hat{x} = Q^{-1}(\Sigma_0^{-1} x_0 + A^T \Sigma_1^{-1} y). \tag{8}$$

The formal error estimates (squared) for the parameters $\hat{x}_j$ are the diagonal elements of Q^{-1}. If the a priori part is uniform, we simply get the old LSQ normal equations.