Inverse Problems in Imaging, GV08, M078, 2016

Answer 3 questions

Marks for each part of each question are indicated in square brackets

Calculators are permitted

1. This question is about ill-posedness in inverse problems

   a. Discuss the meaning of the terms *existence*, *uniqueness* and *stability*, using as an example the solution of a linear problem $\boldsymbol{y} = \mathsf{A}\,\boldsymbol{f}$.

   [6 marks]

   b. Define the singular value decomposition (SVD) of a linear operator A and how it can be used to determine whether the properties of existence, uniqueness and stability are present. Give the definition of the Moore-Penrose inverse of A in terms of its SVD.

   [8 marks]

   c. Show that the solution of an inverse problem possessing non-uniqueness can be posed as the optimisation problem

   $$\text{find } \boldsymbol{f}_* = \arg\min \Psi(\boldsymbol{f}) \quad \text{subject to } \boldsymbol{y} = \mathsf{A}\,\boldsymbol{f}\,.$$

   Discuss the meaning of the function $\Psi(\boldsymbol{f})$ and how it relates to prior knowledge of the expected distribution of possible solutions. What form of Ψ is implied if the Moore-Penrose inverse is used for the solution of the inverse problem?

   [6 marks]

d. Show that, in general, the solution of an inverse problem possessing any of the properties non-existence, non-uniqueness, or instability can be posed as the optimisation problem

$$\text{find } \boldsymbol{f}_* = \arg\min \left[\mathcal{D}(\boldsymbol{y}, \mathsf{A}\boldsymbol{f}) + \Psi(\boldsymbol{f})\right] .$$

Discuss the meaning of the function $\mathcal{D}(\boldsymbol{y}, \mathsf{A}\boldsymbol{f})$ and how it relates to prior knowledge of the expected distribution of noise in experimental measurement. What form of $\mathcal{D}$ is implied if the Moore-Penrose inverse is used for the solution of the inverse problem?

[6 marks]

e. Discuss the meaning of *sparsity* as a possible constraint for the distribution of solutions of inverse problems. How is the sparsity constraint implemented in terms of the function Ψ introduced in parts c) and d)? What computational difficulties arise in the use of constraints of this type?

[7 marks]

[Total 33 marks]

2. This question is about image denoising.

Consider the problem of recovering an image f from a noise corrupted measurement $g = f + n$ where n is assumed to be a zero mean white noise process.

a. If the problem is posed as the optimisation of a functional

$$f_* = \arg\min_f \left[\Phi(f) = \int_{\mathbb{R}^2} \frac{1}{2}(g-f)^2 + \alpha\psi(|\nabla f|)\mathrm{d}x\mathrm{d}y \right] ,$$

where α is a regularisation parameter and ψ is an ordinary function of one variable, show that the gradient of $\Phi(f)$ is given by

$$\Phi'(f) = (g-f) - \alpha\nabla \cdot \left(\frac{\psi'(|\nabla f|)}{|\nabla f|} \right) \nabla f \, .$$

[8 marks]

b. Find the explicit form of the gradient $\Phi'(f)$ for the following regularisation functionals

$i)$ First order Tikhonov regulariser $\psi(|\nabla f|) = \frac{|\nabla f|^2}{2}$,

$ii)$ Perona-Malik regulariser $\psi(|\nabla f|) = \frac{T^2}{2} \log \left[1 + \left(\frac{|\nabla f|}{T} \right)^2 \right]$.

Explain in each case the intended effect of the regulariser and the role played by the parameter T in case ii) .

[6 marks]

c. Suppose that the image f is sampled on a regular grid of pixels $f_{ij} = f(x_i, y_j); i = 1 \ldots n_x, j = 1 \ldots n_y$ and reordered as a vector $\boldsymbol{f}$ of length $n_x n_y$. Show that a local quadratic approximation to $\Phi(\boldsymbol{f})$ can be expressed as

$$\Phi(\boldsymbol{f} + \boldsymbol{h}) \simeq \Phi(\boldsymbol{f}) + \boldsymbol{h}^{\mathrm{T}} \left(\alpha \mathsf{L} \boldsymbol{f} - (\boldsymbol{g} - \boldsymbol{f})\right) + \frac{1}{2} \boldsymbol{h}^{\mathrm{T}} \left(\mathsf{I} + \alpha \mathsf{L}\right) \boldsymbol{h} \ ,$$

where L is a sparse $n_x n_y \times n_x n_y$ matrix with at most 5 non-zero entries per row. Show in detail how to construct the entries in L for the regularisation functionals given in part b).

[8 marks]

d. By considering your answer to part c), show that an iterative Gauss-Newton denoising scheme for the image can be written as

$$\boldsymbol{f}^{(k+1)} = \boldsymbol{f}^{(k)} + (\mathsf{I} + \alpha \mathsf{L})^{-1} \left[(\boldsymbol{g} - \boldsymbol{f}^{(k)}) - \alpha \mathsf{L} \boldsymbol{f}^{(k)} \right] \ ,$$

and show that for the first order Tikhonov regulariser this scheme can be solved in one step, but not for the other regulariser introduced in part b). Discuss whether line-search should be applied for the Perona-Malik regulariser.

[6 marks]

e. By considering the limit as $\alpha \to 0$ or otherwise, show that the scheme in part d) can also be considered as an evolution

$$\frac{\partial \boldsymbol{f}}{\partial t} = -\mathsf{L}(\boldsymbol{g} - \boldsymbol{f}) \, .$$

Discuss the relative merits of using an evolution method versus a Gauss-Newton method.

[5 marks]

[Total 33 marks]

3. This question is about Maximum Likelihood methods.

Positron Emission Tomography (PET) is a medical imaging modality in which a radioactive tracer is inserted in the body and decays with rate f. Data is measured in terms of emitted photons which are counted at a rate g. The data and image are related by a model

$$g = \mathcal{A} f$$

where $\mathcal{A}$ is a linear operator denoting the probabilty density function that a photon emitted at spatial location $\boldsymbol{r}$ inside the body is detected by an element of the detector.

a. Assume that the distribution f is discretised into N voxels and that the measurements g are sampled by a total of M detectors, so that $\mathcal{A}$ can be represented as a $M \times N$ matrix A, with $\boldsymbol{f}$ a $N-$dimensional vector and $\boldsymbol{g}$ a $M-$dimensional vector. What is the probability distribution for the random variable $\boldsymbol{g}$, and what are its mean and covariance ? Explain why the sum of each column of A must be positive and less than or equal to one.

[7 marks]

b. Write down an expression for the negative log-likelihood of observing the measurements $\boldsymbol{g}$ given a positron source image $\boldsymbol{f}$ and find the gradient and Hessian of this expression with respect to the elements of $\boldsymbol{f}$.

[8 marks]

c. Describe, and explain the derivation, of the *Maximum-Likelihood* (ML) algorithm for finding $\boldsymbol{f}$ from the measurements $\boldsymbol{g}$. How is regularisation included in this algorithm ?

[10 marks]

d. A friend suggests that you use a least squares method to solve the inverse problem because “all probability distributions end up as Gaussian due to the Central Limit Theorem”. Describe how this would be done. Do you think it is a good idea ?

[8 marks]

[Total 33 marks]

4. This question is about constrained optimisation.

a. Consider this problem

$$\text{minimise} \quad \Phi(\boldsymbol{x}) = \quad 10 - (2x_1 + 3x_2 + 4x_1^2 + 2x_1x_2 + x_2^2)$$

Write this problem as a quadratic form $\Phi(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\mathsf{P}\boldsymbol{x} + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{x} + c$ and thus find its unique solution.

[8 marks]

b. Consider the problem in part a) to be augmented by the following inequality constraints

$$\begin{aligned} x_1 - x_2 &\geq 0 \\ x_1 + x_2 &\leq 4 \\ x_1 &\leq 3 \\ x_2 &\geq -1 \end{aligned}$$

Draw a diagram illustrating the feasible region for this problem and the location of your solution to part a).

[6 marks]

c. State the Karush-Kuhn-Tucker (KKT) conditions that must be satisfied for an optimal solution to the constrained problem, and explain how they can be implemented by an active set method. Show what solution will be found by the active set method for this problem.

[12 marks]

d. Discuss the *log barrier* method for constrained optimisation and its advantages and disadvantages compared to active sets.

[7 marks]

[Total 33 marks]

5. This question is about posterior sampling and estimation

a. State Bayes' theorem expressing the posterior distribution of a function $\pi(\boldsymbol{f}|\boldsymbol{g})$ given a *likelihood model* for measured data $\boldsymbol{g}$ and a *prior model* for the distribution $\pi(\boldsymbol{f})$. Explain the meaning of the terms likelihood and prior in your answer as well as any other terms that you use.

[8 marks]

b. Explain the difference between a *Maximum A-Posteriori* (MAP) estimate of $\pi(\boldsymbol{f}|\boldsymbol{g})$ and a *Conditional Mean* (CM) estimate. Give examples where these would be i) the same, ii) different.

[6 marks]

c. Describe in reasonable detail the Monte Carlo Markov Chain (MCMC) technique for obtaining estimates of the posterior density; include in your answer a discussion of different proposal generating schemes. Is MCMC a good way to obtain the MAP estimate? Is it a good way to obtain the CM estimate? For what other estimates is MCMC a suitable technique?

[12 marks]

d. Define the mixture of Gaussians (MoG) prior and give an example of an inverse problem where it might be used. Discuss the relative merits of using deterministic versus stochastic solution methods for solving such an inverse problem.

[7 marks]

[Total 33 marks]

END OF PAPER