Engsci 721 Inverse Problems and Learning From Data

Oliver Maclaren (oliver.maclaren@auckland.ac.nz)

Lecture 6: Regularisation

Problem 1

Define and contrast the optimisation problems a) satisfied by the generlised inverse and b) used to define Tikhonov-regularised solutions.

Problem 2

Write down as many formulations of the Tikhonov-regularised linear least squares problem that you can (I can think of at least 3-5).

Explain the interpretation of any relevant regularisation parameters.

Sketch trade-off curves for some of these and how solutions vary with the associated regularisation parameter.

Which of these formulations are applicable to both linear and nonlinear problems? Show how to express nonlinear problems for these formulations.

Problem 3

Consider the standard Tikhonov form of the regularised least squares problem

$$\min_x ||Ax - y||^2 + \lambda ||x||^2$$

- Rewrite this in the form of a new 'augmented' but 'standard' least squares problem.
- Explain how the previous procedure enables us to convert e.g. an underdetermined problem into an overdetermined problem.
- Justify the linear independence of the columns of the matrix in the resulting overdetermined problem.
- Explain why this enables us to 'solve' the resulting problem and derive an explicit expression for the 'Tikhonov inverse'. You may state/use the normal equations without proof.
- Describe the solutions corresponding to the two limits $\lambda \to 0$ and $\lambda \to \infty$.

Problem 4

Describe a method for choosing the regularisation parameter λ in Tikhonov regularisation.

Lecture 7: Higher-order Tikhonov

Problem 5

Consider the standard Tikhonov form of the regularised least squares problem

$$\min_x ||Ax - y||^2 + \lambda ||x||^2$$

where $||\cdot||$ is the L_2 norm. Suppose you instead wanted to obtain solutions that emphasise 'smoothness' rather than small norm $||x||$. Show how to formulate an optimisation problem to do this. Define and give expressions for any new operators you introduce.

Lecture 8: L_1-norm and sparsity

Problem 6

Consider the standard Tikhonov form of the regularised least squares problem

$$\min_x ||Ax - y||^2 + \lambda ||x||^2$$

where $||\cdot||$ is the L_2 norm. Suppose you instead wanted to obtain solutions that preserve *sharp* transitions or features in a in x rather than small L_2 norm.

- Modify the above optimisation problem to be suitable for this task.
- Briefly explain why your formulation is appropriate for this task.
- Sketch typical solutions that you might obtain from this approach compared to an approach emphasising smooth solutions.

Problem 7

Outline how to use IR/WLS to solve an L_2 penalised problem.

Lectures 9: Singular value decomposition

Problem 8

Contrast the problem that must be solved to obtain eigenvalues/eigenvectors of a matrix A to that which must be solved to obtain singular values/singular vectors.

Problem 9

Write down the full, reduced and series expansion forms of the SVD.

Problem 10

Give explicit forms of the generalised inverse of A in terms of its SVD component matrices, both reduced and full, and in terms of an SVD series expansion. Verify, using one of these forms, that it is a left/right inverse when it should be.

Problem 11

Suppose you have a general linear forward model, A, with four observations (i.e. number of rows $m = 4$) and three parameters (i.e. number of columns $n = 3$). You may assume the columns are linearly independent.

- Write down the general expressions for the *model resolution* operator and the *data resolution* operator, respectively, using matrices appearing in the SVD of A. State which spaces they map between.
- Which of these do you expect to be the identity in this case and which not? Why?
- Give a simple example of U_r, V_r compatible with the basic problem description above (hint: if in doubt, use standard bases!).
- Next, check your answer to the question about which resolution operator is expected to be the identity (up to numerical rounding...) on the case of polynomial regression for projectile motion with observations of the form $\{(t_i, y_i)\}, i = 1, 2, 3, 4$. In particular, suppose that $t_1 = 1, t_2 = 3, t_3 = 5, t_4 = 13$ and hence the forward operator is:

$$A = \begin{bmatrix} 1. & 1. & -0.5 \\ 1. & 3. & -4.5 \\ 1. & 5. & -12.5 \\ 1. & 13. & -84.5 \end{bmatrix}.$$

This corresponds to a model with four observations and three parameters (the coefficients of constant, linear and quadratic terms, respectively). You can use Python/Matlab etc to compute the (reduced) SVD e.g. np.linalg.svd(A, full_matrices=False).

Hint: you should find that the (reduced) matrices of left and right singular vectors are given by:

$$U_r = \begin{bmatrix} -0.00769051 & 0.29107872 & 0.89685522 \\ -0.05693158 & 0.58920415 & 0.1049587 \\ -0.15174676 & 0.73868339 & -0.42646699 \\ -0.98674848 & -0.14986157 & 0.05253846 \end{bmatrix}, \; V_r = \begin{bmatrix} -0.01388544 & 0.35601275 & 0.93437793 \\ -0.15886432 & 0.92181289 & -0.3535861 \\ 0.98720278 & 0.15334901 & -0.04375794 \end{bmatrix}$$

respectively.

Problem 12

What would you expect the singular spectrum of the forward operator appearing in a typical ill-posed problem to look like?

Problem 13

Explain how to use truncated SVD to obtain regularised solutions.

Problem 14

Look up principle components analysis (PCA) and its relation to SVD.

Lecture 10: Iterative regularisation

Problem 15

Reformulate the least squares normal equations

$$A^T A x = A^T y$$

as a fixed point problem and show how to derive a simple iterative solution scheme. Hint: if you get stuck, use a 'theory of everything' or two.

Problem 16

In reference to the problem above, explain the concept of 'iterative regularisation' (and/or 'semi-convergence') and how you might use it to obtain a 'good' solution to an ill-posed inverse problem. Make sure to describe what you'd expect/want the solutions to look like a) at the very begininning of iteration, b) in the 'middle' phase of iteration, and c) in the very last phase of iteration.

Lecture 11: Stochastic/minibatch gradient descent

Problem 17

Describe the minibatch version of gradient descent and contrast it with standard gradient descent. What is the minibatch size for a) stochastic gradient, and b) standard gradient descent. What is another name for standard gradient descent in machine learning terminology.

Problem 18

Suppose we carry out minibatch gradient descent for three epochs. How many times has the full training data set been used? If the minibatch sizes are 10 samples, how many parameter updates (gradient descent iterations) have occurred after three epochs?