ENGSCI 721

INVERSE PROBLEMS

Oliver Maclaren
oliver.maclaren@auckland.ac.nz

MODULE OVERVIEW

Inverse Problems (*Oliver Maclaren*) [~8 lectures/2 tutorials]

1. *Basic concepts* [4 lectures]
   Forward vs inverse problems. Well-posed vs ill-posed problems. Algebra and calculus of inverse problems (matrix calculus, generalised inverses etc). Regularisation and trade-offs.
2. *More regularisation* [3 lectures]
   Higher-order Tikhonov regularisation, truncated singular value decompositions, iterative regularisation.

MODULE OVERVIEW

3. *Preview of the statistical view of inverse problems*
   [1 lectures] Bayesians, Frequentists and all that. Basic frequentist analysis.

LECTURE 2: INVERSES I

Topics:

- The algebra and calculus of inverse problems
- Tall and wide systems
- Least squares solutions to tall systems
- Matrix calculus

EngSci 721 : Lecture 2. Inverses I

- Algebra & Calculus of inverse problems
  - ↳ Resolving lack of $\boxed{\text{existence}}$ (today) &/or uniqueness (tomorrow)
    - ↳ non-square linear systems
    - ↳ non 1-1/onto nonlinear systems
  - ↳ Formulating & solving as optimisation problems
    - ↳ $\boxed{\text{least squares}}$ & least norm solutions
    - ↳ matrix calculus
  - ↳ Algebraic characterisation of various types of inverse: preview

Algebra & Calculus of Inverse Problems

Generalised Inverses

Our basic problem can be defined as:

$$\boxed{\text{'solve', ie 'invert', equations like } F(x) = y \text{ for } x, \text{ given } y}$$

where:

- x & y could be vectors, functions, images etc
- solutions might not exist, might not be unique &/or might not be stable

---

Note: mappings, measurements & vectors

We might have an 'exact' model of the form

$$\boxed{y = a + bx}, \text{ for } \boxed{\text{scalars}}\ x, y$$

If we take a series of 'noisy' $\boxed{\text{measurements}}$ we get eg:

$$y_{obs,i} = (a + bx)_i + e_i$$

$$\Rightarrow \quad y_{obs,i} = a + x_i + e_i$$

for $i = 1, \ldots, n$.

These lead to $\boxed{\text{vector}}$ eqns in terms of the noisy observations/realisations:

$$\begin{bmatrix} y_{obs,1} \\ \vdots \\ y_{obs,n} \end{bmatrix} = a \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + b \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

eg $\boxed{\bar{y}_{obs} \approx a\bar{1} + b\bar{x}}$ } $\bar{y}, \bar{x}, \bar{1}$ vectors

(I usually drop the explicit overbars on vectors)

Linear or nonlinear? Finite or Infinite?

- We will discuss some basic algebra of the problem in the linear & finite dimensional setting

  → ie using Linear Algebra in $\mathbb{R}^n$

- This can be considered as the 'algebra of linear mappings' in $\mathbb{R}^n$

→ An important question is:

do these results carry over to the nonlinear &/or infinite-dimensional setting(s)?

Short answer: yes!

(key concepts & as long as careful)

Linear or nonlinear? Finite or Infinite?

- Algebra of arbitrary mappings?

  ↳ Category theory

  (see eg Nashed / Maclaren & Nicholson for generalised inverses in category theory setting
  → existence ≡ axiom of choice!)

- Discontinuous infinite dimensional linear mappings ('ill-posed')

  ↳ ill-conditioned finite dimensional matrices

→ Theory is a bit beyond scope, but most of the key tools are applicable to nonlinear setting
(we'll solve some of these!)

Linear Setting

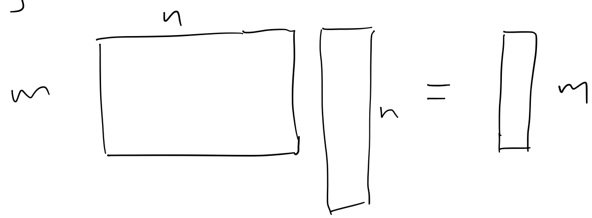
$$\mathbb{R}^n \underset{?}{\overset{A}{\rightleftarrows}} \mathbb{R}^m$$

Consider the system of equations

$$Ax = y$$

- A is $m \times n$ matrix
- $x \in \mathbb{R}^n$ vector
- $y \in \mathbb{R}^m$ vector

eg



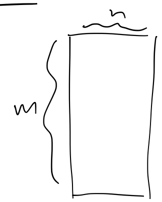
rows : eq^{n}s

cols : unknowns

How do we solve when $m \neq n$?

→ $m > n$, more rows than cols.

existence?



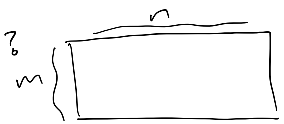
→ eq^{n}s > unknowns

→ possibly inconsistent / overdeterm.

→ 'more data than parameters'

→ $m < n$, more cols than rows

uniqueness?



→ unknowns > eq^{n}s

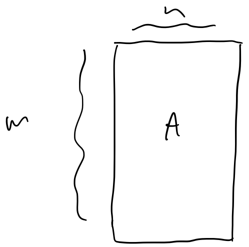
→ possibly many sol^{n}s

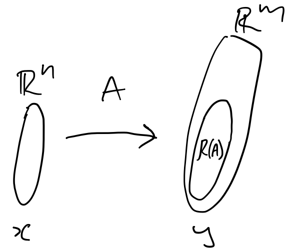
→ 'more param. than data'

Tools developed to 'solve' each case

Case 1: possibly inconsistent: more equations than unknowns.

Consider $Ax = y$, A is $m \times n$ & $m > n$





→ Assume w.l.o.g. that all n columns are linearly independent

- $y \in R(A) \iff$ there is a unique solution

  $R(A)$: range/col space/image of A

  → see handout

- $y \notin R(A) \Rightarrow$ no (exact) solution

→

1. Inconsistent equations: approximate solution

→ Define $r = y - Ax$ } residual 'error'

→ Measure size of error with a norm $\|\cdot\|$ (see handout for diff. types)

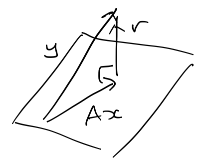
→ Typically assume $\|\cdot\|_2$

New problem:

$$\underset{x \in \mathbb{R}^n}{\text{minimise}} \|y - Ax\|, \quad A \text{ \& } y \text{ given}$$

- "best approximation"
- "closest approx"

etc.



Minimiser of $\| \ \|$ & minimiser of $\| \ \|_2^2$ are same ($x \mapsto x^2$ is monotonic for $x \geq 0$)

⇒ Least squares approximation:

$$\underset{x}{\text{minimise}} \|y - Ax\|_2^2 \quad \text{(equiv. problem)}$$

Norms, products, summation etc

- $\|x\|_2^2 = \langle x, x \rangle$
$$= x^T x$$
$$= \sum_i x_i^2$$

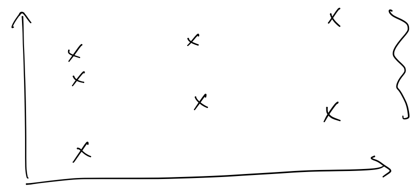
- $Ax = \sum A_{ij} x_j$
$$= A_{ij} x_j$$ (Einstein summation convention → sum over repeated indices)

So:

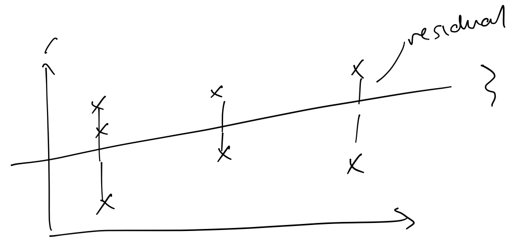
- $y^T A x = y_i A_{ij} x_j$

See handout (?) (we'll do some practice!)

Illustration



no line can fit all → inconsistent



least squares

Derivation via calc (can do geometrically too!)

$$\min_x \|y - Ax\|_2^2 = \min_x \langle y - Ax, y - Ax \rangle$$
$$= \min_x f(x)$$

where $f(x) = \langle y - Ax, y - Ax \rangle$
$$[= (y - Ax)^T (y - Ax)]$$

same, different notation.
$$= \langle y, y \rangle - 2 \langle y, Ax \rangle + \langle Ax, Ax \rangle$$
$$= y^T y - 2 y^T A x + x^T A^T A x$$
$$= y_i y_i - 2 y_i A_{ij} x_j + (A_{ij} x_j)(A_{ik} x_k)$$

etc

Differentiating vectors, matrices, tensors wrt vectors, matrices, tensors ...

Requires:

'Matrix calculus', 'tensor calculus' etc
→ multiple conventions/notation
(see eg Wiki page on Matrix Calculus)

I will sketch here, but then give you three key rules you can use instead of remembering details!

Conventions

→ I'll use $d_x f$ for derivative of f wrt x, regardless of whether f, x are scalar, vector etc.

→ use 'Jacobian layout', eg vector f, x:

$$[d_x f]_{ij} = \frac{\partial f_i}{\partial x_j} = \begin{array}{c|c} & x_1 \quad \cdots \quad x_n \\ \begin{array}{c} f_1 \\ \vdots \\ f_m \end{array} & \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \end{array}$$

Note, if f is scalar-valued, this implies $d_x f$ is a row vector:

$$\boxed{d_x f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}}$$

Hence this is sometimes written

$$\boxed{d_x f = \frac{\partial f}{\partial x^T}}$$

to indicate the 'layout' direction is 'row like' in x.

We can define the gradient of f as

$$\boxed{\nabla_x f = (d_x f)^T = \frac{\partial f^T}{\partial x}}$$

(often drop transpose on f)

which again indicates how to 'lay out' results.

3 Key rules: (A, a independent of x)

Constant

$$d_x(a) = \frac{\partial}{\partial x^T}(a) = 0^T$$

Linear

$$d_x(Ax) = \frac{\partial}{\partial x^T}(Ax) = A$$

Quadratic

$$d_x(x^T A x) = \frac{\partial}{\partial x^T}(x^T A x) = x^T(A + A^T)$$

Know these!

Extra:

Additional principles for 'deriving' results (see handout):

- Derivative of scalar, possibly dep. on x:

$$d_x(a(x)) = d_x(a) = d_x(a^T)$$

- (Multivariable vector) chain rule

$$d_x(f(h(x), g(x))) = \frac{\partial}{\partial x^T} f(h, g)$$

$$= (d_g f)\cdot(d_x g) + (d_h f)\cdot(d_x h)$$

$$= \frac{\partial f}{\partial g^T}\frac{\partial g}{\partial x^T} + \frac{\partial f}{\partial h^T}\frac{\partial h}{\partial x^T}$$

Imply eg 'product rule' of form:

$$d_x(h^T g) = h^T d_x g + g^T d_x h$$

$$= h^T \frac{\partial g}{\partial x^T} + g^T \frac{\partial h}{\partial x^T}$$

Back to least squares!

$$\min_x f(x) = y^T y - 2y^T A x + x^T A^T A x$$

$\Rightarrow$ set $d_x f = 0^T$ (1) (row vector)

(or $\nabla_x f = 0$ ...)

3 Rules:

$$d_x(y^T y) = 0^T$$

$$d_x(-2y^T A x) = -2y^T A$$

$$d_x(x^T A^T A x) = x^T(A^T A + (A^T A)^T)$$

$$= 2x^T A^T A$$

(1) $\Rightarrow$ $-2y^T A + 2x^T A^T A = 0^T$

$\Rightarrow$ $-A^T y + A^T A x = 0$

$\Rightarrow$ $\boxed{A^T A x = A^T y}$

Least squares approximation

We have seen this leads to....

The normal equations:

$$\boxed{A^T A x = A^T y}$$

[normal? $A^T(Ax - y) = 0$ geometric]

→ Since we assume the n cols of A are linearly independent then $A^T A$ is invertible (see handout) & so get unique approximate sol^n

$$\boxed{x^* = (A^T A)^{-1} A^T y}$$

(we will look at what happens if not LI soon!).