Mathematical Data Science

Numerical Linear Algebra for Big Data

Martin van Gijzen

Delft University of Technology

March 29, 2018

Outline

Introduction

- Last time we discussed the SVD, and image compression as an example

Introduction

- Last time we discussed the SVD, and image compression as an example
- Today will be devoted to tomography as an example of least-squares problems

Introduction

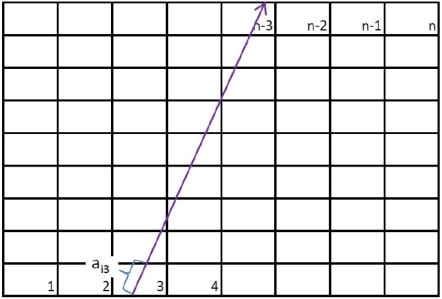
- Last time we discussed the SVD, and image compression as an example
- Today will be devoted to tomography as an example of least-squares problems
- The SVD is again a useful tool to solve such problems
  However, only suited for small problems

Introduction

- Last time we discussed the SVD, and image compression as an example
- Today will be devoted to tomography as an example of least-squares problems
- The SVD is again a useful tool to solve such problems
  However, only suited for small problems
- We will also discuss a technique that us applicable to large problems: LSQR.

Tomography

The goal of tomography is to reconstruct the interior of a body from projections.



Examples include

- Geophysical tomography
- Medical tomography

Geophysical tomography

Suppose that the travel time of an earth quake and that the angle between surface receiver and epi-center are known. Then for each reception of a wave one can set-up a linear equation

$$T_t = \sum_{i=1}^{n_p} a_i s_i$$

in which

- T_t is the travel time (measured)
- a_i is the travel length through pixel i (known, mostly zero)
- s_i inverse of wave speed through pixel i (not known)

The goal is to compute s_i from many measurements

Medical tomography

In a CT-scanner beams of X-rays are transmitted. For each ray the decay is measured, which yields a linear equation

$$d_t = \sum_{i=1}^{n_p} a_i d_i$$

in which

- d_t is the absorption (known)
- a_i is the travel length through pixel i (known, mostly zero)
- d_i is the absorption in pixel i

The goal is to compute d_i from many measurements

Tomography: least-squares problem

The tomography problem leads to a linear system

$$Ax = b$$

with $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$.
Furthermore,

- The system may be inconsistent ($b \notin \mathcal{R}(A)$).
- Usually $m \neq n$.
- The rank of A may be smaller than n.

Tomography: least-squares problem

The tomography problem leads to a linear system

$$Ax = b$$

with $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$.
Furthermore,

- The system may be inconsistent ($b \notin \mathcal{R}(A)$).
- Usually $m \neq n$.
- The rank of A may be smaller than n.

Solutions to this problem satisfy the normal equations

$$A^T A x_{LS} = A^T b$$

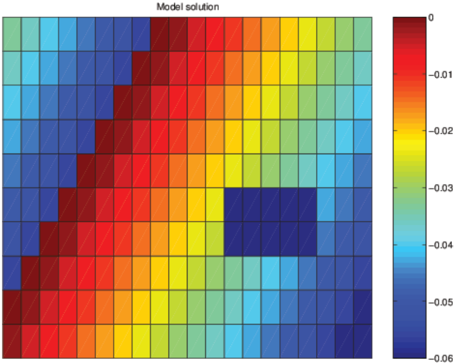
But the problem is ill-posed.

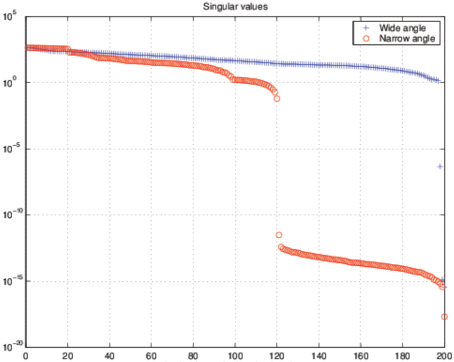
Application to geophysical tomography

We will illustrate the theory on a small geophysical test case: Nolet's problem.

- Nolet's problem collects measurements from 20 earth quakes on 20 receivers.
- The domain is subdivided in 20×10 pixels.
- This leads to problem with 400 equations and 200 unknowns.
  In practice the number of equations can be billions.
- We will consider two test cases: narrow angle and wide angel.
- The right-hand-side vector is either consistent (no noise) or perturbed with noise.

Nolet's problem





- Wide angle case: arrival angle between -35^o and 5^o
- Narrow angle case: arrival angle between -5^o and -1^o

Least-squares problems

Suppose $rank(A) < n$ and x_{LS} is a least-squares solution. Then

$$\hat{x} = x_{LS} + y \quad \text{with} \quad y \in \mathcal{N}(A)$$

is also a least squares solution.
A unique least-squares solution x_{LSMN} is the one with minimum norm, which is the solution of the constrained problem

$$\min_x \|Ax - b\|_2 \quad \text{subject to} \quad x \perp \mathcal{N}(A) \ .$$

The Singular Value Decomposition

Let $A \in \mathbb{R}^{m\times n}$ be a matrix of rank r. Then there exist unitary matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ such that

$$A = U\Sigma V^T, \quad \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix}$$

where $\Sigma \in \mathbb{R}^{m\times n}$ and $\Sigma_r = diag(\sigma_1, \sigma_2, \cdots, \sigma_r)$, and

$$\sigma_1 \geq \sigma_1 \geq \cdots > 0.$$

The σ_i are called the singular values of A.

The SVD and the LSMN solution

The least-squares minimum norm solution can be computed using the SVD by

$$x_{LSMN} = V \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T b$$

The matrix

$$A^+ = V \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T$$

is called the pseudoinverse or the Moore-Penrose inverse of A.

The SVD and the LSMN solution (2)

Proof that $x_{LSMN} = A^+ b$:

$$z = V^T x = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \quad c = U^T b = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

where $z_1, c_1 \in \mathbb{R}^r$. Then

$$\|b - Ax\|_2 = \|U^T(b - AVV^T x)\|_2 =$$

$$= \| \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} - \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \|_2 = \| \begin{pmatrix} c_1 - \Sigma_r z_1 \\ c_2 \end{pmatrix} \|_2$$

Hence $\|b - Ax\|_2$ is minimized by $z_1 = \Sigma_r^{-1} c_1$ and $\|x\|$ by $z_2 = 0$.

Noisy problems

In least-squares problems b often corresponds to measured date, which means that we are actually solving the noisy problem

$$Ax = b + \delta b \ .$$

Moreover, small singular values typically correspond to the noise.

These small singular values have a dramatic effect on the LSMN-solution (why?)!!!

This is an example of a so-called ill-posed problem: small perturbations in the data give a large perturbation in the solution.

Regularization

Limiting this effect is called regularization. Several regularization methods have been proposed:

- Set small singular values to 0. This requires the explicit calculation of the SVD, which is not possible for large scale problems.
- (Tykhonov regularisation) Solve the damped least squares problem:

$$\min_x \left\| \begin{pmatrix} A \\ \tau I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2$$

- Use an iterative method (reason: convergence to small singular values is slow)

LSQR

LSQR (Paige and Saunders) is derived by applying Lanczos bi-diagonlisation to

$$\begin{pmatrix} I & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} r \\ x \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix} .$$

with starting vector $u_1 = \frac{1}{\|b\|} \begin{pmatrix} b \\ 0 \end{pmatrix}$

LSQR (2)

Bidiagonalisation algorithm (Golub and Kahan)

$\beta_1 u_1 = b \quad \alpha_1 v_1 = A^T u_1$

FOR $i = 1, \cdots$ DO

$$\beta_{i+1} u_{i+1} = A v_i - \alpha_i u_i$$
$$\alpha_{i+1} v_{i+1} = A^T u_{i+1} - \beta_{i+1} v_i$$

END FOR

with $\alpha_i > 0$ and $\beta_i > 0$ such that $\|u_i\| = \|v_i\| = 1$.

LSQR (3)

With $U_k = [u_1, u_2, \cdots, u_k]$, $V_k = [v_1, v_2, \cdots, v_k]$ and

$$B_k = \begin{bmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \beta_3 & \ddots & \\ & & \ddots & \alpha_k \\ & & & \beta_{k+1} \end{bmatrix},$$

it follows that

$$\begin{aligned} \beta_1 U_{k+1} e_1 &= b \\ A V_k &= U_{k+1} B_k \\ A^T U_{k+1} &= V_k B_k^T + \alpha_{k+1} v_{k+1} e_{k+1}^T \end{aligned}$$

LSQR (4)

Now construct solution vectors $x_k = V_k y_k$. Then we get for $r_k = b - Ax_k$:

$$\begin{aligned} r_k &= \beta_1 U_{k+1} e_1 - A V_k y_k \\ &= \beta_1 U_{k+1} e_1 - U_{k+1} B_k y_k \\ &= U_{k+1}(\beta_1 e_1 - B_k y_k) \\ &= U_{k+1} t_k \end{aligned}$$

LSQR (5)

Substitution in the augmented system and using the Gallerkin condition gives

$$\begin{pmatrix} U_{k+1}^T & 0 \\ 0 & V_k^T \end{pmatrix} \begin{pmatrix} I & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} U_{k+1}t_{k+1} \\ V_k y_k \end{pmatrix} = \begin{pmatrix} U_{k+1}^T b \\ 0 \end{pmatrix},$$

which leads to the reduced system

$$\begin{pmatrix} I & B_k \\ B_k^T & 0 \end{pmatrix} \begin{pmatrix} t_{k+1} \\ y_k \end{pmatrix} = \begin{pmatrix} \beta_1 e_1 \\ 0 \end{pmatrix}.$$

LSQR (6)

This last equation is equivalent to the least squares problem

$$\min \|\beta_1 e_1 - B_k y_k\|_2$$

In LSQR this problem is solved using the QR-algorithm.

- LSQR is mathematically equivalent to CG applied to the normal equations
- It minimizes the residual over $K^k(A^T A; A^T b)$
- LSQR is famous for its robustness
- It solves the LSMN-problem

Final remarks

Today we have seen some algorithms for tomography

Other widely used algorithms include

- Filtered Back Projection
- ART
- SIRT

Although LSQR is prefered from a mathematical point of view, these simple methods are the ones used in practice.