Lecture Notes 20: Linear Least Squares Problems

CPSC 302: Numerical Computation for Algebraic Problems

Jessica Bosch
jbosch@cs.ubc.ca
http://www.cs.ubc.ca/~jbosch

University of British Columbia
Department of Computer Science

2017/2018 Winter Term 1

Outline

Goals of this Chapter

- Introduce and solve the linear least squares problem, ubiquitous in data fitting applications.
- Introduce algorithms based on orthogonal transformations.
- Evaluate different algorithms and understand what their basic features translate into in terms of a tradeoff between stability and efficiency.
- Introduce SVD use for rank-deficient and highly ill-conditioned problems.

Chapter 5: Linear Least Squares Problems

- Motivation
- Data Fitting

- Normal Equations
- QR Decomposition
- Householder and Gram-Schmidt
- SVD and Truncated SVD (TSVD)

Outline

Linear Least-Squares

- Throughout this chapter we consider the problem

$$\min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_2,$$

  where A is $m \times n$, with $m > n$.

- So, it is an overdetermined system of equations: we have more rows, for instance corresponding to data measurements, than columns, where $\mathbf{x}$ corresponds to unknown model parameters.

- In general, there is no $\mathbf{x}$ satisfying $A\mathbf{x} = \mathbf{b}$, hence we seek to minimize a norm of the residual $\mathbf{r} = \mathbf{b} - A\mathbf{x}$. The ℓ_2-norm is the most convenient to work with, although it is not suitable for all purposes, and it enjoys rich theory.

- Assume A has linearly independent columns. Then there is a unique solution to this problem, as we'll soon see.

Outline

Application: Data Fitting

Given measurements, or observations

$$(t_1, b_1), (t_2, b_2), \ldots, (t_m, b_m) = \{(t_i, b_i)\}_{i=1}^{m},$$

want to fit a function

$$v(t) = \sum_{j=1}^{n} x_j \phi_j(t),$$

- $\phi_1(t), \phi_2(t), \ldots, \phi_n(t)$ are known linearly independent basis functions
- $x_1, \ldots, x_n$ are coefficients to be determined s.t.

$$v(t_i) \approx b_i, \quad i = 1, 2, \ldots, m.$$

Data Fitting (cont.)

Define $a_{ij} = \phi_j(t_i)$. Want $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m-1,1} & a_{m-1,2} & \cdots & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}, \ \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{m-1} \\ b_m \end{pmatrix}.$$

Data Fitting (cont.)

Define $a_{ij} = \phi_j(t_i)$. Want $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m-1,1} & a_{m-1,2} & \cdots & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}, \ \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{m-1} \\ b_m \end{pmatrix}.$$

1. If $m = n$ get interpolation problem. Solution $\mathbf{x} = A^{-1}\mathbf{b}$.

Data Fitting (cont.)

Define $a_{ij} = \phi_j(t_i)$. Want $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m-1,1} & a_{m-1,2} & \cdots & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{m-1} \\ b_m \end{pmatrix}.$$

1. If $m = n$ get interpolation problem. Solution $\mathbf{x} = A^{-1}\mathbf{b}$.
2. If $m > n$ we can't generally set $\mathbf{r} = \mathbf{b} - A\mathbf{x} = \mathbf{0}$. So relax requirement: we want, e.g., $\min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_2$. Obtain a least-squares data fitting problem.

Data Fitting (cont.)

Define $a_{ij} = \phi_j(t_i)$. Want $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m-1,1} & a_{m-1,2} & \cdots & a_{m-1,n} \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{m-1} \\ b_m \end{pmatrix}.$$

1. If $m = n$ get interpolation problem. Solution $\mathbf{x} = A^{-1}\mathbf{b}$.
2. If $m > n$ we can't generally set $\mathbf{r} = \mathbf{b} - A\mathbf{x} = \mathbf{0}$. So relax requirement: we want, e.g., $\min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_2$. Obtain a least-squares data fitting problem.

Note: even if we can increase n so that A becomes square, there may be reasons not to want this:

1. A smaller n gives fewer parameters to control and may better describe global trend of data.
2. If the data contains noise, don't want to over-fit it.

Example: Linear Regression

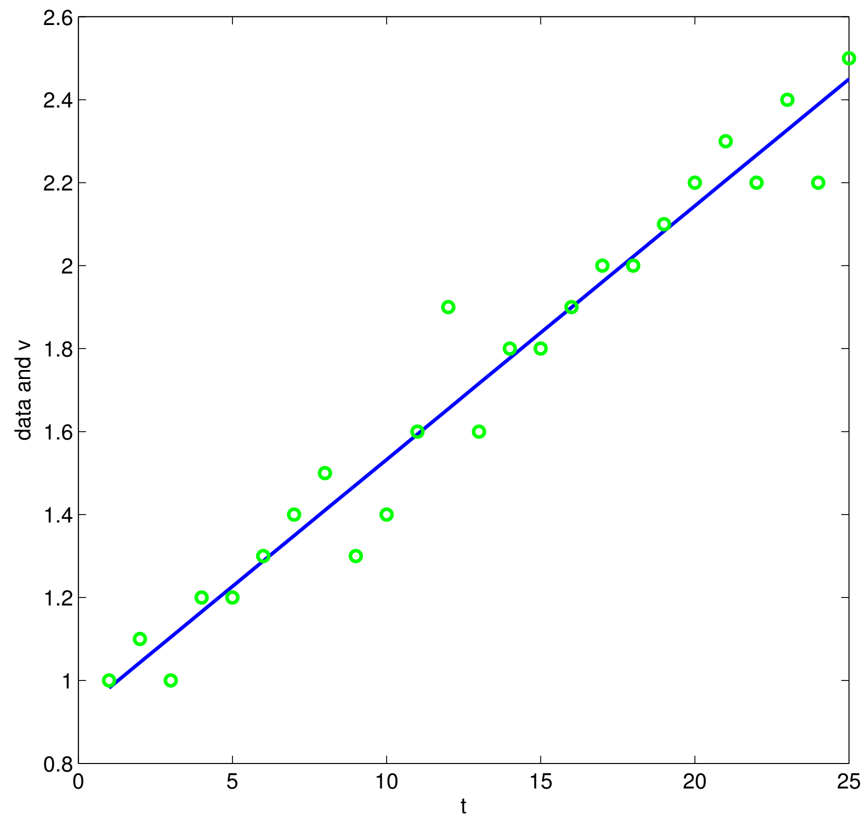

Figure: Linear regression curve (in blue) through green data points. Here $m = 25$ and $n = 2$.

Data Fitting Example

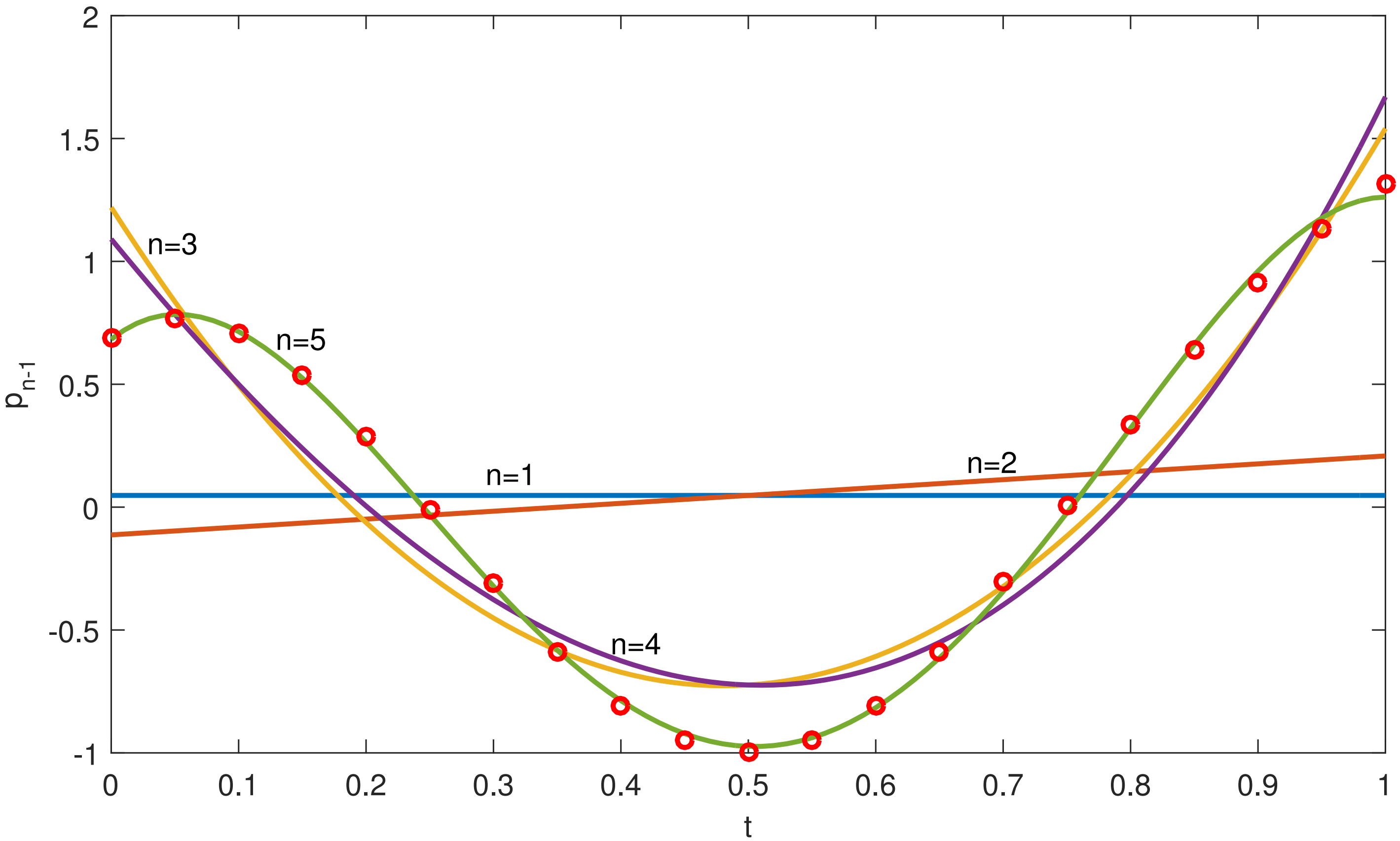

Figure: First 5 best polynomial approximations to $f(t) = \cos(2\pi t) + 10(t - .5)^5$ sampled at $0 : 0.05 : 1$. Data values at red circles.

Other Data Fitting Problems

- When $m > n$, i.e., A "long and skinny", we have an over-determined system.

Other Data Fitting Problems

- When $m > n$, i.e., A "long and skinny", we have an over-determined system.
- Instead of least-squares, may consider

$$\min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_1 \quad \text{or} \quad \min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_\infty$$

  for the over-determined problem. Both of these lead to linear programming formulations. The ℓ_1-norm is good against outliers in data.

Other Data Fitting Problems

- When $m > n$, i.e., A "long and skinny", we have an over-determined system.
- Instead of least-squares, may consider

$$\min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_1 \quad \text{or} \quad \min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_\infty$$

  for the over-determined problem. Both of these lead to linear programming formulations. The ℓ_1-norm is good against outliers in data.
- If $m < n$ we have an under-determined system. Now there are many solutions to $A\mathbf{x} = \mathbf{b}$: want to pick one wisely. For instance,

$$\min_{\mathbf{x}} \{\|\mathbf{x}\|_2 \;\; s.t. \;\; A\mathbf{x} = \mathbf{b}\}$$

- Alternatively, do it in ℓ_1

$$\min_{\mathbf{x}} \{\|\mathbf{x}\|_1 \;\; s.t. \;\; A\mathbf{x} = \mathbf{b}\}$$

  Obtain a sparse solution: $x_j = 0$ for at least $m - n$ components. Again, this leads to a linear programming formulation, further discussed in AG, Chapter 9.