CS 357: Numerical Methods

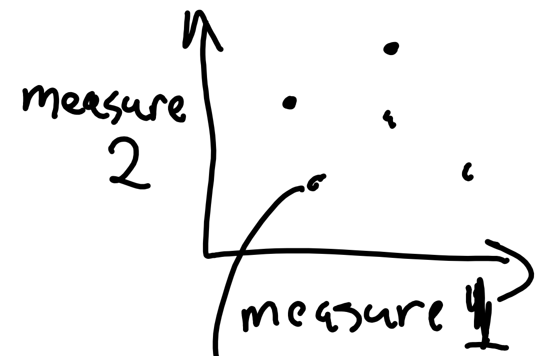
Principal Component Analysis
Interpolation

Eric Shaffer

Principal Component Analysis Revisited

Consider n trials

Each trial produces an m dim. point



$P_j = (x_{1j}, x_{2j})$

$X = \begin{pmatrix} \quad \end{pmatrix}$ ↓ i measure

$j \to$ trial

$$\begin{array}{c} \begin{array}{ccc} P_1 & & \end{array} \\ \begin{pmatrix} x_{11} & x_{12} & \cdots \\ x_{21} & x_{22} & \end{pmatrix} \end{array}$$

Principal Component Analysis Revisited

① Compute estimate of mean

$$u_i = \frac{1}{n} \sum_j x_{ij}$$

row-wise average

② $y_{ij} = x_{ij} - u_i = Y$

③ $C = \frac{1}{n-1} \left(Y Y^T \right)$

Principal Component Analysis Revisited

Principal Component
are columns of U
in $\left(\frac{1}{\sqrt{n-1}}\right) Y = U \Sigma V^T$

Interpolation

- Suppose we have the following discrete data

t	-2	0	1
y	-27	-1	0

- Common things we wish to do include:
  - Draw a smooth curve through the data points
  - Infer values between the points
  - Predict future values
  - Approximate the function that generated the values
    - In way that allows us to find a derivative or integral
    - Compute new values of the function quickly

Interpolation

- Interpolation is the process of fitting a function to data such that the function matches the given data values ***exactly***.
- Such a function is called an ***interpolant***.

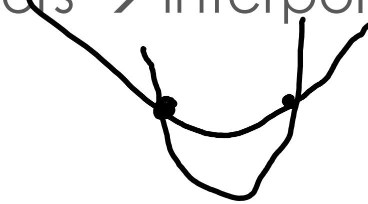
- Consider computing a line $y=x_1 + x_2 t$ to the data points (t1,y1) and (t2,y2)
- What does the line give us that the points don't?

What form should an interpolant have?

- Depends on a number of factors
  - How easy should it be to evaluate at other points?
  - Does the application require certain behavior?
    - Periodicity
    - Convexity
    - Monotonicity
- Common choices include
  - Polynomials and piecewise polynomials
  - Trigonometric functions
  - Exponential functions

Existence and Uniqueness

- Existence and uniqueness are determined by
  - The number of data points
  - The number of parameters in the interpolant
- Too few parameters → interpolant does not exist
  - Example? Fit $x_1 t + x_2$ to 1,000,000
- Too many parameters → interpolant is not unique
  - Example?



Basis Functions

Given a set of data points an interpolant is chosen from a set of basis functions $\phi_1(t), \dots, \phi_n(t)$

that span a space of functions.

The interpolant is expressed as a linear combination of the basis functions

$$\underset{=}{f(t_i)} = \sum_{j=1}^{n} \underbrace{x_j}_{\text{weight}} \underbrace{\phi_j(t_i)}_{\text{basis function}} = \underset{=}{y_i}$$

Determining the interpolant parameters

- Requiring that f interpolate (t_i, y_i) means

$$f(t_i) = \sum_{j=1}^{n} x_j \phi_j(t_i) = y_i$$

- This is a system of linear equations $\quad Ax = y$

$$a_{ij} = \phi_j(t_i)$$

y_i are known

x_i are unknown coefs.

Existence and Uniqueness

$Ax = y$

- If the number of parameters = number of data points
  - And what else? A is nonsingular

  We have a unique interpolant

- If the number of parameters > number of points
  - Interpolant is not unique
  - Why might that be useful?

Choice of basis functions

- The choice of basis functions affects
  - The conditioning of A
  - The amount of work to solve the system
  - Anything else?

Computational cost of evaluating the interpolant

The monomial basis

$\phi_j(t) = t^{j-1}$ for $j = 1, \ldots, n$

for n data points

$p_{n-1}(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$

polynomial of degree $n-1$

have n parameters

The Vandermonde matrix

$$Ax = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & & t_2^{n-1} \\ \vdots & \vdots & \vdots & & \\ 1 & t_n & t_n^2 & \cdots & t_n^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

---

if $Az = 0 \to$ poly w/ coefs given by z has n zeros $\to z = 0$

Example

$(-2,-27) \quad (0,-1) \quad (1,0)$

fit $p(t) = x_1 + x_2 t + x_3 t^2$

$$Ax = \begin{bmatrix} 1 & -2 & 4 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} x = \begin{bmatrix} -27 \\ -1 \\ 0 \end{bmatrix}$$

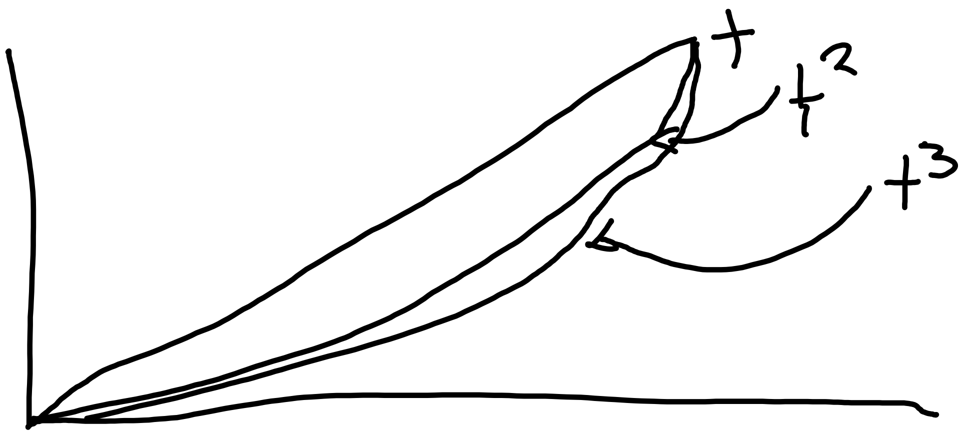
Solve w/ GE

$$x = \begin{matrix} -1 \\ 5 \\ -4 \end{matrix}$$

Example

Complexity and Conditioning

- What is the complexity of solving the system? $O(n^3)$
- The Vandermonde matrix is often ill-conditioned



as n gets larger

columns get closer to linearly dependent

Evaluating a polynomial

- What is cost of evaluating polynomial of degree n-1?

$P_{n-1} = x_1 + x_2 t + \cdots + x_n t^{n-1}$

$O(n)$ additions

mults $= 1 + 2 \cdots + n-1$

$O(n^2)$

Horner's Method

$$p_{n-1} = x_1 + t(x_2 + t(x_3 + t(\ldots (x_{n-1} + t x_n)))$$

$O(n)$ adds

$O(n)$ mults

Example

$$1 - 4t + 5t^2 - 2t^3 + 3t^4$$

$$= 1 + t(-4 + t(5 + t(-2 + 3t)))$$