Statistical Computing

Class Notes Fall 2019

Stat 771

Vivak Patel

Floating Point Numbers and Finite Arithmetic

Overview
1. Positional Numeral System
2. Floating Point Number + IEEE standards
3. Floating Point Approximations and Errors
4. Finite Arithmetic Operations + Fundamental Axiom

I. Positional Numeral System

Define: A positional numeral system is a way of representing numbers and is characterized by
1. Base, $\beta \in \{2,3,4,\dots\}$
2. A sequence of digits $d_0.d_1d_2d_3\dots$, $d_j \in \{0,1,\dots,\beta-1\}$
3. An exponent $e \in \mathbb{Z}$

The number $d_0.d_1d_2\dots \times \beta^e$

$$(d_0 + d_1\beta^{-1} + d_2\beta^{-2} + \cdots)\beta^e$$

Question 1

II. Floating Point Numbers + IEEE Standards

A. General Floating Point Numbers

Definition: A $(\beta, p, e_{min}, e_{max})$ floating point number is a real number that can be exactly represented in a base $\beta \in \{2,3,4,\dots\}$ with a fixed precision p (i.e., number of digits), and whose exponent is between e_{min} and e_{max} (inclusive, integer)

$$\{(d_0 + d_1\beta^{-1} + \cdots + d_{p-1}\beta^{-p+1})\beta^e : d_j \in \{0,1,\dots,\beta-1\},\ e \in \{e_{min}, e_{min}+1, \dots, e_{max}\}\}$$

Example: Suppose $\beta=10$, $p=3$, $e_{min}=-1$, $e_{max}=1$.
Then 0.1 is a floating point number
- 0.01 and $e=1$
- 0.10 and $e=0$
- 1.00 and $e=-1$

Definition: A floating point representation is normalized if $d_0 \neq 0$.

Example: The total number of non-negative floating point numbers that can be represented in base $\beta, p, e_{min}, e_{max}$ and are normalized
- $d_0 \in \{1,\dots,\beta-1\}$, $\beta-1$
- $d_1,\dots,d_{p-1}$: β^{p-1}
- e: $(e_{max}-e_{min}+1)$

Example: $\beta=3$, $p=2$, $e_{min}=-1$, $e_{max}=0$. 12 non-negative

$3^{-1}(1+0(3^{-1})) = 3/9$
$3^{-1}(1+1(3^{-1})) = 4/9$
$3^{-1}(1+2(3^{-1})) = 5/9$
$3^{-1}(2+0(3^{-1})) = 6/9$
$3^{-1}(2+1(3^{-1})) = 7/9$
$3^{-1}(2+2(3^{-1})) = 8/9$
$3^{0}(1+0(3^{-1})) = 1$
$3^{0}(1+1(3^{-1})) = 4/3$
$5/3$
$6/3$
$7/3$
$8/3$

Question 2

B. IEEE Standards

Define: A bit is a single position of storage on a computer.

1. IEEE 16 bit standard (Half-precision)
- 1 bit is sign
- 5 bits is exponent. 00000 special, 11111
- 11 bits are significand

a. Exponent

$2^5 - 2 = 30$, 00001_2 to 11110_2; $1\cdot2^4 + 1\cdot2^3 + 1\cdot2^2 + 1\cdot2^1 + 0\cdot2^0 = 30$

$e_{min} = 00001_2 - 01111_2 = -01110_2 = -14$
$e_{max} = 11110_2 - 01111_2 = 01111_2 = 15$

$e = 00000$

$2^{-14}(0.\_\ 2^{-1}\_\ 2^{-2}\_ + 2^{-3} + \cdots)$

Largest Value:
$$2^{15}(1 + 1\cdot2^{-1} + 1\cdot2^{-2} + \cdots + 1\cdot2^{-10}) = 65504$$

Smallest Value (normalized)
- Sign bit 0
- Exponent: 00001_2
- 0's in significand

$2^{-14}(1+0) = 6.10\times10^{-5}$

Smallest Subnormal Value:
- sign bit: 0
- Exponent: 00000 (2^{-14})
- 0 ... 1 (implicit, last)

$2^{-14}(0 + \cdots + 0 + 1\cdot2^{-10}) = 2^{-24} \approx 5.96\times10^{-8}$

2. IEEE 32-bit (Single Precision) Standard
- 1 bit is sign
- 8 bits for exponent. 00000000 special 0, 11111111 ∞
- 24 bits for significand (first is not stored)

32 bit Standard
- 1 bit sign
- 8 bits exponent
- 24 bits significand (1 is implicit)

(1) Largest normalized number? $2^{127}(1 + 1(2^{-1}) + 1(2^{-2}) + \cdots + 1\cdot2^{-23}) = 2^{128}(1-2^{-24}) \approx 3.4\times10^{38}$

(2) Smallest normalized number? $2^{-126}(1+0\cdots+0) = 2^{-126} \approx 1.18\times10^{-38}$

(3) Smallest subnormal number? $2^{-126}(0 + \cdots + 1(2^{-23})) = 2^{-149} \approx 1.40\times10^{-45}$

III. Floating Point Approximation + Errors.

0 10101 0000000000 = 64 ×
 −01111

Next number 0 10101 0000000001 $= 64 + (2^{-10})2^6 = 64 + 2^{-4}$

$[64, 64+\frac{2^{-4}}{2}] \Rightarrow 64$
$(64+\frac{2^{-4}}{2}, 64+2^{-4}] \Rightarrow 64+2^{-4}$

0 01111 0000000000 = 1.0
0 01111 0000000001 $= 2^0(1 + \cdots + 1\cdot2^{-10}) = 1 + 2^{-10}$

Proposition Let S be the set of (positive) normalized floating numbers for $(\beta, p, e_{min}, e_{max})$. Then for any real number $z \in \left[\beta^{e_{min}}, \beta^{e_{max}}\left(\frac{1-\beta^{-p}}{1-\beta^{-1}}\right)\right]$

(1) $\exists s \in S$ with exponent e s.t. $|z-s| \le 0.5\cdot\beta^{e-p+1}$ (absolute error)

(2) $\frac{|z-s|}{z} \le 0.5\,\beta^{-p+1}$ (relative error)

Proof $s_1 = \max\{s\in S : s \le z\}$
$s_2 = \min\{s \in S : s \ge z\}$

(1) If $s_1 = s_2$ then $z \in S$, error is 0.
(2) If $s_1 < s_2$. e be exponent of s_1 and $d_0.d_1\dots d_{p-1}$ be the digits
(a) $\exists j$ s.t. $d_j < \beta-1$.
$r = s_1 + \beta^{e-p+1}$
Fact: $r > s_1$
Fact? $r \le s_2$
Fact? $z \le r$
$r = s_2$
$z \in [s_1, s_1+\beta^{e-p+1}]$
$|z-s| \le \frac{\beta^{e-p+1}}{2}$
(b) $\forall j$, $d_j = \beta-1$
$s_1 = \beta^e(\beta-1 + (\beta-1)\beta^{-1} + \cdots + (\beta-1)\beta^{-p+1}) = \beta^e(\beta-1)\left(\frac{1-\beta^{-p}}{1-\beta^{-1}}\right) = \beta^{e+1}(1-\beta^{-p})$
$r = \beta^{e+1}$
$z \in [s_1, s_1+\beta^{e-p+1}]$

Relative Error $\frac{|z-s|}{z} \le \frac{\beta^{e-p+1}}{2s_1} \le \frac{\beta^{e-p+1}}{2\beta^e}$

See Question 5 + Added Question 1

Definitions (Units in the Last Place (ULP))
- For a floating point number with base β, exponent e and precision p, 1 ULP is equal to β^{e-p+1}

Definition (Machine Epsilon)
For a base β with precision p, machine epsilon/precision, ε, is either
(1) β^{-p+1} (C, C++, Matlab, Julia, Python) ×
(2) $0.5\beta^{-p+1}$ (LAPACK)

Example Largest absolute error for representing z, is 0.5 ULP
Largest relative error, is 0.5ε

IV. Floating Point Arithmetic

$10^{-3} + 2^{-6}$ in 16 bit

4.1 Addition
(1) Same exponent
(2) Truncate
(3) Summation

Example: Base 10, $p=3$

	e	sign.	true value	inf p	closest
Number 1	2	1.34	134	134	
Number 2	→ 0	2.70	2.70	2.70	
Convert	2	0.027	2.70		
truncate	2	0.02	2.00		
Sum	2	1.37	137	136.70	1.37, e=2

1 ULP $= 10^{2-3+1} = 1$. 0.3 ULP

See Question 4

4.2 Subtraction and Guard digit

	e	sign	value	inf. precision	closest
num. 1	1	1.01	10.1	10.1	
num. 2	0	9.93	9.93	9.93	
Convert	1	0.993	9.93		
Truncate	1	0.99	9.9		
Subtract	1	0.02	0.2	0.17	1.70 e=−1

1 ULP $= 10^{1-3+1} = 0.1$, 0.3 ULP

$\frac{0.03}{0.17} \approx 0.176 = 17.6\varepsilon$

See Question 6

4.3 Multiplication + Division
1. Add exponents
2. Multiply significand
3. Truncate

4.4 Properties of finite precision addition, multiplication
1. Add/mult are commutative
2. Add/mult are not associative
3. Multiplication over a sum is not distributive

V. Fundamental Axiom

Axiom There exists a $u>0$ s.t. for any operation, denoted by op, the corresponding floating point operation, flop, between floating point numbers x and y satisfies

$x \text{ flop } y = (x \text{ op } y)(1+\delta)$ $|\delta| \le u$

Example: Matrix Storage
Suppose $A \in \mathbb{R}^{n\times m}$. flop(A)
$|\text{flop}(A) - A| \le u|A|$

Question 7

5.1 Dot Product (Exact Storage)

```
function flopdot(x,y)
    n = length(x)
    n == length(y) || @error
    s = 0
    for i = 1:n
        s += x[i]*y[i]
    end
    return s
end
```

Lemma: Let $x,y \in \mathbb{R}^n$. Assume fund. axiom holds on arithmetic operations, exact representation, $(n+1)u < 0.01$

$$|\text{flopdot}(x,y) - \text{dot}(x,y)| \le 1.01(n+1)u\,|x|'|y|$$

Proof Let s_p be the value of s after iteration $i=p$.
Then for $|\delta_j| \le u$ and $|\varepsilon_j| \le u$
$s_1 = \text{flop}(\times)(x_1,y_1) = x_1y_1(1+\delta_1)$
$s_p = \text{flop}(+)(s_{p-1}, \text{flop}(\times)(x_p,y_p))$
$= (s_{p-1} + x_py_p(1+\delta_p))(1+\varepsilon_p)$
$= \sum_{i=1}^p x_iy_i(1+\delta_i)\prod_{j=i}^p(1+\varepsilon_j)$

$|\sum_{i=1}^n x_iy_i - s_n| = |\sum_{i=1}^n x_iy_i[1 - (1+\delta_i)\prod_{j=i}^n(1+\varepsilon_j)]|$
$\le \sum_{i=1}^n |x_i||y_i|\,|1-(1+\delta_i)\prod_{j=i}^n(1+\varepsilon_j)|$

Induction $|\alpha_j| \le u$
1. Base case: $|(1+\alpha_1) - 1| \le (1+u) - 1$
2. Hypothesis: $|\prod_{i=1}^j(1+\alpha_i) - 1| \le (1+u)^j - 1$
3. Final: $|(1+\alpha_{j+1})\prod_{i=1}^j(1+\alpha_i) - 1|$
$= |(1+\alpha_{j+1})\prod_{i=1}^j(1+\alpha_i) - (1+\alpha_{j+1}) + \alpha_{j+1}|$
$= |(1+\alpha_{j+1})(\prod_{i=1}^j(1+\alpha_i) - 1) + \alpha_{j+1}|$
$\le |1+\alpha_{j+1}|((1+u)^j - 1) + |\alpha_{j+1}|$
$\le (1+u)[(1+u)^j - 1] + u$

$|1 - (1+\delta_i)\prod_{j=i}^n(1+\varepsilon_j)| \le (1+u)^{n-i+2} - 1$
$\le (1+u)^{n+1} - 1$
$\le \exp((n+1)u) - 1$
$\le \sum_{j=1}^\infty \frac{[(n+1)u]^j}{j!}$
$\le \sum_{j=1}^\infty \frac{[(n+1)u]^j}{2^j}$
$\le \frac{(n+1)u}{1-(n+1)u/2}$
$\le \frac{1}{1-0.005}(n+1)u$
$\le 1.01(n+1)u$

$|\sum_{i=1}^n x_iy_i - s_n| = |\sum_{i=1}^n x_iy_i[1-(1+\delta_i)\prod_{j=i}^n(1+\varepsilon_j)]|$
$\le \sum_{i=1}^n |x_i||y_i|\,1.01(n+1)u$

5.2 Solving Linear System (Precise solver, Inexact Storage)

Def: Let A be invertible, and let $\|A\|_p = \max_{\|x\|_p=1}\|Ax\|_p$

We define condition number of A
$\kappa_p(A) = \|A\|_p\|A^{-1}\|_p$

Theorem: Let A be invertible, $b\neq0$ and let $x = A^{-1}b$.
- Solve linear exactly
- $A+E$, $\|E\|_\infty \le u\|A\|_\infty$
- $b+e$, $\|e\|_\infty \le u\|b\|_\infty$

And $u\kappa_\infty(A) < 1$ then $\hat{x}$ that solves $(A+E)\hat{x} = b+e$

$$\frac{\|x-\hat{x}\|_\infty}{\|x\|_\infty} \le \frac{2u\kappa_\infty(A)}{1-u\kappa_\infty(A)}$$

Proof:
1. $A+E$ is invertible.
$A(I + A^{-1}E)$
$\|A^{-1}E\|_\infty \le \|A^{-1}\|_\infty\|E\|_\infty \le u\|A\|_\infty\|A^{-1}\|_\infty = u\kappa_\infty(A) < 1$
Then by Question 8, $A+E$ is invertible
2. $\|\hat{x}\|_\infty$ $\|x\|_\infty$
$\|\hat{x}\|_\infty = \|(I+A^{-1}E)^{-1}A^{-1}(b+e)\|_\infty$ $(A+E)\hat{x} = b+e$
$= \|(I+A^{-1}E)^{-1}(x + A^{-1}e)\|_\infty$ $Ax = b$
$\le \|(I+A^{-1}E)^{-1}\|_\infty(\|x\|_\infty + u\|A^{-1}\|_\infty\|b\|_\infty)$ $\|e\|_\infty \le u\|b\|_\infty$
$\le \frac{1}{1-\|A^{-1}E\|_\infty}\left(\|x\|_\infty + u\kappa_\infty(A)\frac{\|b\|_\infty}{\|A\|_\infty}\right)$
$\le \frac{1+u\kappa_\infty(A)}{1-u\kappa_\infty(A)}\|x\|_\infty$ $\|A\|_\infty\|x\|_\infty \ge \|b\|_\infty$
3. $\hat{x} - x = A^{-1}e - A^{-1}E\hat{x}$
$\|\hat{x} - x\|_\infty \le u\kappa_\infty(A)(\|x\|_\infty + \|\hat{x}\|_\infty)$

Question 9

Orthogonalization

Outline
I. Motivating Problems
II. QR decomposition
III. Solve all motivating problem
IV. Existence of QR
V. Stable computation of QR decomp.
VI. Large-data least squares

I. Motivating Problems

Problem 1: A consistent linear system
$A \in \mathbb{R}^{n\times m}$ that has rank m
$b \in \text{range}(A)$
Find x s.t. $Ax=b$

Problem 2: Least Squares
Suppose $A\in\mathbb{R}^{n\times m}$, rank m
$b\in\mathbb{R}^n$
Find x s.t. $x \in \arg\min_{y\in\mathbb{R}^m}\|Ay-b\|_2$

Problem 3: Underdetermined Linear System
$A\in\mathbb{R}^{n\times m}$ rank$(A) \le n < m$
$b \in \text{range}(A)$
Find $x \in \arg\min_{z\in\mathbb{R}^m}\{\|z\|_2 : Az = b\}$

Problem 4: Underdetermined least squares
$x \in \arg\min_{z\in\mathbb{R}^m}\{\|z\|_2 : z \in \arg\min_y \|Ay-b\|_2\}$

Problem 5: Constrained Least Squares
$A\in\mathbb{R}^{n\times m}$, $n \ge m \ge \text{rank}(A)$, $b\in\mathbb{R}^n$
$C\in\mathbb{R}^{p\times m}$, rank$(C) = p$, $d\in\mathbb{R}^p$
$x \in \arg\min_y \|Ay-b\|_2$ s.t. $Cy = d$

II. QR decomposition

Def: A permutation matrix is a square matrix s.t. each column and each row has exactly one entry that takes value 1 and the rest are 0

Example: • Identity matrix • $\begin{bmatrix}0&1\\1&0\end{bmatrix}$ } Question 11

Def A square matrix Q is orthogonal if $Q'Q = QQ' = I$

Def : A matrix
- R is called upper triangular if $R_{ij} = 0$ $\forall i>j$
- L is called a lower triangular matrix if $L_{ij}=0$ $\forall j > i$

Question 12 $3x+2y=4$, $8y=1$

Theorem Let $A\in\mathbb{R}^{n\times m}$ and let $r = \text{rank}(A)$. There exists
- An $m\times m$ permutation Π
- An $n\times n$ orthogonal Q
- An $r\times r$ upper triangular R (w/ nonzeros on diagonal)
- An $r\times(m-r)$ matrix S ($r=m$, S does not matter)

s.t. $A = Q\begin{bmatrix}R & S\\0&0\end{bmatrix}\Pi'$ (Q: $n\times n$, R: $r\times r$, S: $r\times(m-r)$, $(n-r)\times r$, $(n-r)\times(m-r)$, Π': $m\times m$)

called the QR decomposition.

III. Solutions to Problems

3.1. Consistent Linear System.
$A\in\mathbb{R}^{n\times m}$ and rank$(A)=m$
$b \in$ range(A).
$Ax=b$
$Q\begin{bmatrix}R\\0\end{bmatrix}\Pi'x = b$
$\begin{bmatrix}R\\0\end{bmatrix}\Pi'x = Q'b = \begin{bmatrix}c\\0\end{bmatrix}$ $\}m$, $\}n-m$
$R\Pi'x = c$
$\Pi'x = R^{-1}c$
$x = \Pi(R^{-1}c)$

3.2. Least Squares
Recall $A\in\mathbb{R}^{n\times m}$ rank$(A)=m$
$\|Ay-b\|_2 = \|Q\begin{bmatrix}R\\0\end{bmatrix}\Pi'y - b\|_2$
$= \|\begin{bmatrix}R\\0\end{bmatrix}\Pi'y - Q'b\|_2$
$= \|\begin{bmatrix}R\\0\end{bmatrix}\Pi'y - \begin{bmatrix}c_1\\c_2\end{bmatrix}\|_2$
$= \|\begin{bmatrix}R\Pi'y - c_1\\-c_2\end{bmatrix}\|_2$
$= \sqrt{\|R\Pi'y - c_1\|_2^2 + \|c_2\|_2^2}$

Question 15
$x \in \arg\min_y \|R\Pi'y - c_1\|_2^2 + \|c_2\|_2^2$
$x = \Pi R^{-1}c_1$

Question 14
Question 18

3.3. Underdetermined Linear System
$A\in\mathbb{R}^{n\times m}$ and rank$(A) \le n < m$
$b\in$ range(A)
$x \in \arg\min_{y\in\mathbb{R}^m}\{\|y\|_2 : Ay = b\}$

1. $Ay = b$
$Q\begin{bmatrix}R&S\\0&0\end{bmatrix}\Pi'y = b$
$\begin{bmatrix}R&S\\0&0\end{bmatrix}\Pi'y = Q'b$
$\begin{bmatrix}R&S\\0&0\end{bmatrix}\begin{bmatrix}z_1\\z_2\end{bmatrix} = \begin{bmatrix}c\\0\end{bmatrix}$
$Rz_1 + Sz_2 = c$
$z_1 = -R^{-1}Sz_2 + R^{-1}c$

2. $\|y\|_2^2 = \|\begin{bmatrix}z_1\\z_2\end{bmatrix}\|_2^2 = \|\begin{bmatrix}R^{-1}c - R^{-1}Sz_2\\z_2\end{bmatrix}\|_2^2$
$= \|R^{-1}c - R^{-1}Sz_2\|_2^2 + \|z_2\|_2^2$
$-2S'(RR')^{-1}c + 2S'(RR')^{-1}Sz_2 + 2z_2$
$S'(RR')^{-1}c = (S'(RR')^{-1}S + I)z_2$

Question 15.
$(S'(RR')S + I)^{-1}S'(RR')^{-1}c = z_2$

3.4. Undetermined Least Squares
$A\in\mathbb{R}^{n\times m}$ w/ rank$(A) \le n < m$.
$b\in\mathbb{R}^n$.
Find $x \in \arg\min\{\|z\|_2 : z \in \arg\min_y \|Ay-b\|_2\}$

1. $\|Ay-b\|_2^2 = \|Q\begin{bmatrix}R&S\\0&0\end{bmatrix}\Pi'y - b\|_2^2$
$= \|\begin{bmatrix}R&S\\0&0\end{bmatrix}\Pi'y - \begin{bmatrix}c_1\\c_2\end{bmatrix}\|_2^2$ ($\Pi'y = \begin{bmatrix}z_1\\z_2\end{bmatrix}$)
$= \|Rz_1 + Sz_2 - c_1\|_2^2 + \|c_2\|_2^2$
$z_1 = -R^{-1}Sz_2 + R^{-1}c_1$

2. $\|y\|_2^2 = \|\Pi'y\|_2^2 = \|\begin{bmatrix}z_1\\z_2\end{bmatrix}\|_2^2 = \|-R^{-1}Sz_2 + R^{-1}c_1\|_2^2 + \|z_2\|_2^2$

3.5. Constrained Least Squares
Question 17

IV. Existence of QR decomposition

4.1 Gram-Schmidt Procedure

Lemma: Let $r\in\mathbb{N}$ and given a set of linearly independent vectors $\{a_1,\dots,a_r\}$, there exists an orthonormal set of vectors, $\{q_1,\dots,q_r\}$, that span $\{a_1,\dots,a_r\}$

Proof
1. Base case: $r=1$
$R_{11} = \|a_1\|_2$, $q_1 = \frac{1}{R_{11}}a_1$, $a_1 = q_1R_{11}$
2. $r=2$. q_2:
(a) $\|q_2\|_2 = 1$ ✓
(b) $q_1'q_2 = 0$ ✓
(c) R_{12}, R_{22}, s.t. $a_2 = q_1R_{12} + q_2R_{22}$ ✓
$s = a_2 - q_1'a_2\cdot q_1$
$a_2 = q_1\underbrace{(q_1'a_2)}_{R_{12}} + \underbrace{\frac{s}{\|s\|_2}}_{q_2}\underbrace{\|s\|_2}_{R_{22}}$
$q_1's = q_1'a_2 - (q_1'a_2)\cdot q_1'q_1 = 0$
3. Induction Hypothesis
Suppose $\{a_1,\dots,a_{r-1}\}$ is linearly independent then there is an orthonormal $\{q_1,\dots,q_{r-1}\}$ that spans $\{a_1,\dots,a_{r-1}\}$
$a_j = \sum_{i=1}^j q_iR_{ij}$ for $j=1,\dots,r-1$
4. $\{a_1,\dots,a_r\}$ are lin. ind.
$\{q_1,\dots,q_{r-1}\}$ that are orthonormal and span $\{a_1,\dots,a_{r-1}\}$
(a) $\|q_r\|_2 = 1$
(b) $q_r'q_j = 0$ $j=1,\dots,r-1$
(c) $a_r = \sum_{i=1}^r q_iR_{ir}$
$s = a_r - \sum_{i=1}^{r-1}q_i(q_i'a_r)$
$a_r = \sum_{i=1}^{r-1}q_i\underbrace{(q_i'a_r)}_{R_{ir}} + \underbrace{\frac{s}{\|s\|_2}}_{q_r}\underbrace{\|s\|_2}_{R_{rr}}$

$q_i's = q_i'a_r - \sum_{j=1}^{r-1}q_i'q_j(q_j'a_r)$
$= q_i'a_r - q_i'a_r = 0$

Remember
$a_j = \sum_{i=1}^j q_iR_{ij} = [q_1\ q_2 \cdots q_j]\begin{bmatrix}R_{1j}\\R_{2j}\\R_{3j}\\\vdots\\R_{jj}\end{bmatrix}$
$= [q_1\ q_2\cdots q_j\ q_{j+1}\cdots q_r]\begin{bmatrix}R_{1j}\\R_{2j}\\\vdots\\R_{jj}\\0\\\vdots\\0\end{bmatrix}$

$$\underbrace{[a_1\ a_2 \cdots a_r]}_{A} = \underbrace{[q_1\ q_2\cdots q_r]}_{\text{orthogonal matrix}}\underbrace{\begin{bmatrix}R_{11}&R_{12}&R_{13}&\cdots&R_{1r}\\0&R_{22}&R_{23}&\cdots&R_{2r}\\0&0&R_{33}&\cdots&\\\vdots&&&\ddots&\\0&0&&&R_{rr}\end{bmatrix}}_{\text{Upper triangular matrix}}$$

4.2 Proof of Existence

Theorem: (QR decomposition)
Let $A\in\mathbb{R}^{n\times m}$ and let $r=\text{rank}(A)$. Then there exist
- An $m\times m$ permutation matrix Π
- An $n\times n$ orthogonal matrix Q
- An $r\times r$ upper tri. matrix R
- An $r\times(m-r)$ matrix S $(m>r)$

s.t. $A = Q\begin{bmatrix}R&S\\0&0\end{bmatrix}\Pi'$

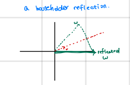
Proof Let Π be a permutation matrix s.t. $A\Pi$'s first r columns are linearly independent.
① $A\Pi$'s first r columns by $[a_1\cdots a_r] = [q_1\cdots q_r]R$
② If $m>r$ $a_{r+1},\dots,a_m \in \text{span}\{q_1,\dots,q_r\}$
$a_{r+j} = \sum_{i=1}^r q_i\underbrace{(q_i'a_{r+j})}_{S_{ij}}$
$A = [q_1\cdots q_r][R\ \ S]\Pi'$ Skinny/Thin QR
③ Full QR decomposition?
$n-r$ more columns of $[q_1,\dots,q_r]$ * See Added Question 2

Some considerations:
1. If A has full column rank, $\Pi = I$.
2. Numerical considerations, we still want $\Pi \neq I$.
See Question 19.
3. A is not full column rank. Π (Added Question 2)
4. Pivoting Question 20. Dan Negrut

V. Stable Computation QR decomposition

5.1 Householder Reflections

Definition. A matrix $H = I - 2vv'$, where $\|v\|_2=1$ is called a householder reflection.



Lemma: A householder reflection is an orthogonal matrix.

Proof : $HH' = I - 4vv' + 4v(v'v)v' = I$

Lemma: (Question 21) Product of orthogonal matrices is orthogonal

$H_nH_{n-1}\cdots H_1A = R$ an upper triangular matrix
$\|a_1\|e_1 = H_1a_1 = a_1 - 2v_1v_1'a_1$
$v_1(2v_1'a_1) = a_1 - \|a_1\|e_1$
$v_1 = \frac{a_1 - \|a_1\|e_1}{\|a_1 - \|a_1\|e_1\|}$
$H_1A = [H_1a_1\ H_1a_2\cdots H_1a_m] = [\|a_1\|e_1\ H_1a_2\cdots H_1a_m] = \begin{bmatrix}\|a_1\| & c_1'\\0 & A_1\end{bmatrix}$ ($n-1$, $n-1\times m-1$)
$H_2 = \begin{bmatrix}1&0'\\0&I-2v_2v_2'\end{bmatrix}$
$H_2H_1A = \begin{bmatrix}\|a_1\| & c_1'\\0&(I-2v_2v_2')A_1\end{bmatrix}$
$\begin{bmatrix}I&0\\0&\Pi_j\end{bmatrix}$

5.2. Givens Rotations

Definition: A matrix $G^{(i,j)}$ is a givens rotation if it is the identity matrix:
1. $[i,i]$, $[j,j]$ are set to λ
2. $[i,j]$ σ
3. $[j,i]$ $-\sigma$

Example: $G^{(2,4)} = \begin{bmatrix}1&0&0&0\\0&\lambda&0&\sigma\\0&0&1&0\\0&-\sigma&0&\lambda\end{bmatrix}$

In Question 21, $\sigma^2+\lambda^2 = 1$, then $G^{(i,j)}$ is orthogonal

$B = G^{(i,j)}M$ where M is an arbitrary matrix
$B_{ik} = \lambda M_{ik} + \sigma M_{jk}$
$B_{jk} = -\sigma M_{ik} + \lambda M_{jk}$

$\underbrace{G_n\cdots G_1}_{\text{orthogonal } Q'}A$ we end up with an upper triangular matrix

$G_1a_1 = \begin{bmatrix}\bar{a}\\0\\a_1[3]\\\vdots\\a_1[n]\end{bmatrix}$

$G_1 = G^{(1,2)}$ $0 = -\sigma a[1] + \lambda a_1[2]$
$\sigma^2+\lambda^2 = 1$.
$\lambda = \frac{a_1[1]}{\sqrt{a_1[1]^2 + a_1[2]^2}}$
$\sigma = \frac{a_1[2]}{\sqrt{a_1[1]^2 + a_1[2]^2}}$
$\bar{a} = \sqrt{a_1[1]^2 + a_1[2]^2}$

$G_nG_{n-1}\cdots G_1A = \begin{bmatrix}\|a_1\|_2 & c_1'\\0&A_1\end{bmatrix}$ ($n-1$, $n-1\times m-1$)

VI. Large Data least Squares

$A\in\mathbb{R}^{n\times m}$ is a large matrix $(n\gg m)$

Gentleman (1973, 1974)

1. $QR=A$ $(A'A)^{-1}A'b = (R'R)^{-1}R'(Q'b) = R^{-1}(Q'b)$
2. qr decomposition $[A\ b]$
$\begin{matrix}m\{\\n-(m+1)\end{matrix}\begin{bmatrix}R & Q'b\\0&\rho\\0&0\end{bmatrix}$ $m+1\times m+1$

1. $[a_1\ b_1] \leftarrow$ first rows of $[A\ b]$ ($1\times m+1$)
$[R_1\ c_1] \leftarrow 1\times m+1$
2. $\begin{bmatrix}R_1&c_1\\a_2&b_2\end{bmatrix}\}$ 2 rows $\rightarrow [R_2\ c_2]$
⋮
3. $\begin{bmatrix}R_m&c_m\\a_{m+1}&b_{m+1}\end{bmatrix}\}$ $m+1$ rows $\rightarrow \begin{bmatrix}R_{m+1}&c_{m+1}\\0'&\rho_{m+1}\end{bmatrix}$ ($m\times m$, $m\times1$, $1\times m$, 1×1)

Rotational Characterizations of Matrices

Outline:
1. Motivating Problems
2. Eigenvalue (Schur) and Singular Value Decompositions
3. Solve the Problems
4. Existence of EVD and SVD
5. Properties of EVD and SVD
6. QR iteration for computing EVD
7. Large scale data reduction

I. Motivating Problems

Problem 1: Underdetermined Least Squares Problem
Let $A\in\mathbb{R}^{n\times m}$ with rank$(A) \le \min\{m,n\}$.
Let $b\in\mathbb{R}^n$.
$x\in\arg\min\{\|y\|_2 : y\in\arg\min\|Az-b\|_2^2\}$

Problem 2: Computing the condition number
$\kappa_2(A) = \|A\|_2\|A^{-1}\|_2$

Problem 3: Finding clusters in N node undirected graph
N nodes
A be the matrix of 1's and 0's that specify connections between nodes
$[1\ 2\ 3\ \cdots\ N] \longrightarrow$ independent clusters

Problem 4: Stationary distribution of Markov Chain
Suppose P is the transition matrix for an irreducible Markov Chain between N states
prob of transitioning from $i\to j$, $j\to i$

Problem 5: Low rank Matrix Approximation (Netflix Problem)
Suppose $A\in\mathbb{R}^{n\times m}$. Find
$X \in \arg\min_{\text{rank}(Y)=k}\|A-Y\|_F^2$

II. Eigenvalue Value Decomp. and Singular Value Decomp.

If A symmetric EVD (Schur decomposition)

Theorem (Schur Decomposition)
Suppose $A\in\mathbb{R}^{n\times n}$ symmetric. Then there exist
- An orthogonal matrix P
- A diagonal matrix $\Lambda\in\mathbb{R}^{n\times n}$

s.t. $A = P\Lambda P'$
Moreover, $Ap_i = \lambda_ip_i$

If A arbitrary

Theorem (SVD) Suppose $A\in\mathbb{R}^{n\times m}$ and let $n\ge m$
There exist
- $U\in\mathbb{R}^{n\times n}$ orthogonal
- $V\in\mathbb{R}^{m\times m}$ orthogonal
- $\sigma_1\ge\sigma_2\ge\cdots\ge\sigma_m\ge0$, $\Sigma = \begin{bmatrix}\sigma_1&&\\&\ddots&\\&&\sigma_m\\&0&\end{bmatrix}$

s.t. $A = U\Sigma V'$

Definition
- Columns of U are left singular vectors
- Columns of V are right singular vectors
- σ_j are singular values of A

Question 24
Corollary 1: (Added problem 3)
Added question 4 is relationship between A and symmetric variants

III. Solutions to motivating problems

3.1. Problem 1: Underdetermined Minimum Norm Solution
$A\in\mathbb{R}^{n\times m}$, rank$(A)\le\min\{m,n\}$. $b\in\mathbb{R}^n$
$x\in\arg\min\{\|y\|_2 : y\in\arg\min\|Ay-b\|_2^2\}$

Suppose A symmetric. $A = P\Lambda P'$
$\|Ay-b\|_2^2 = \|P\Lambda P'y - b\|_2^2$
$= \|\Lambda\underbrace{P'y}_{z} - \underbrace{P'b}_{c}\|_2^2$
$= \|\Lambda z - c\|_2^2$
$\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_n$, $Z = \{i:\lambda_i=0\}$
$\|\Lambda z - c\|_2^2 = \sum_{i\notin Z}(\lambda_iz_i - c_i)^2 + \sum_{i\in Z}c_i^2$
$z_i = c_i/\lambda_i$ $i\notin Z$, z_i can be arbitrary $i\in Z$
$\|z\|_2^2 = \|P'y\|_2^2 = \|y\|_2^2$
$z_i = \begin{cases}0 & \text{if } i\in Z\\ c_i/\lambda_i & \text{if } i\notin Z\end{cases}$

Let A be arbitrary. $A = U\Sigma V'$
$\|Ay-b\|_2^2 = \|U\Sigma V'y - b\|_2^2$
$= \|\Sigma\underbrace{V'y}_{z} - \underbrace{U'b}_{c}\|_2^2$
$= \|\Sigma z - c\|_2^2$
$r = \text{rank}(A)$. $\sigma_{r+1} = \cdots = \sigma_{\min(m,n)} = 0$
$= \sum_{i=1}^r(\sigma_iz_i - c_i)^2 + \sum_{i=r+1}^{\min(m,n)}c_i^2$
$z_i = c_i/\sigma_i$ for $i=1,\dots,r$ z_i are arbitrary for $i=r+1,\dots,\min(m,n)$
$\|z\|_2^2 = \|P'y\|_2^2 = \|y\|_2^2 \Rightarrow$ arbitrary $z_i = 0$

Problem 2: Condition Number
$\kappa_2(A) = \|A\|_2\|A^{-1}\|_2$

Lemma: Let D be a diagonal matrix (real). Then $\|D\|_2 = \max_i|D_{ii}|$

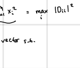
Proof $\|D\|_2^2 = \max_{\|x\|_2=1}\|Dx\|_2^2 = \max_{\|x\|_2=1}\sum_{i=1}^n D_{ii}^2x_i^2$
$\le \max_i|D_{ii}|^2\max_{\|x\|_2=1}\sum_{i=1}^n x_i^2 = \max_i|D_{ii}|^2$
Let $j\in\arg\max_i|D_{ii}|$. Let z be a vector s.t.
$z_i = \begin{cases}0 & i\neq j\\ \text{sign}(D_{ii}) & i=j\end{cases}$
$\max_i|D_{ii}|^2 = |D_{jj}|^2 = \|Dz\|_2^2 \le \max_{\|x\|_2=1}\|Dx\|_2^2$
$= \|D\|_2^2 \le \max_i|D_{ii}|^2$

Suppose A is symmetric (see added question 3)
$\kappa_2(A) = \|A\|_2\|A^{-1}\|_2 = \|P\Lambda P'\|_2\|P\Lambda^{-1}P'\|_2$
$= \|\Lambda\|_2\|\Lambda^{-1}\|_2$
$= \frac{\max_i|\lambda_i|}{\min_i|\lambda_i|}$

Suppose A is arbitrary
$\kappa_2(A) = \|A\|_2\|A^{-1}\|_2 = \|\Sigma\|_2\|\Sigma^{-1}\|_2 = \frac{\sigma_1}{\sigma_n}$

Problem 3: Finding clusters in undirected graph
N nodes, A is adjacency matrix
$A_{ij} = A_{ji} = \begin{cases}1 & \text{if node } i,j \text{ are connected}\\0&\text{if node } i,j \text{ are not (directly) connected}\end{cases}$

Graph:
(1) Node 1 connects to nodes 2,3
(2) Nodes 4 and 5 are connected
(3) Node 6 is on its own



There exists a partition $\mathcal{P}$ of $\{1,\dots,N\}$ s.t. $S_1, S_2\in\mathcal{P}$ and for all $i\in S_1$, $j\in S_2$, $A_{ij}=0$
$\mathcal{P} = \{(1,2,3),(4,5),(6)\}$

$\Pi'A\Pi = \begin{bmatrix}C_{11}&C_{12}&C_{13}&\cdots&C_{1J}\\C_{21}&C_{22}&C_{23}&\cdots&C_{2J}\\\vdots&&&\ddots&\vdots\\C_{J1}&C_{J2}&&\cdots&C_{JJ}\end{bmatrix}$ ($|S_1|$, $|S_2|$, $|S_J|$)

Eigenvalues of $\Pi'A\Pi$ are the eigenvalues of C_{ii}
v_i is an eigenvector of C_{ii}
$\Pi'A\Pi\begin{bmatrix}0\\\vdots\\0\\v_i\\0\\\vdots\\0\end{bmatrix}\}|S_i| = \begin{bmatrix}0\\\vdots\\0\\C_{ii}v_i\\0\\\vdots\\0\end{bmatrix}$

Problem 4: Stationary Distribution
P is transition matrix for N-state irreducible markov chain.
$P'\pi = \pi$

Problem 5: Dimensionality Reduction
$A\in\mathbb{R}^{n\times m}$ find a matrix
$X\in\arg\min_{\text{rank}(Y)=k}\|A-Y\|_F^2$

If rank$(A)\le k$ then just use A
$r = \text{rank}(A) > k$. Using Question 25
$A = U\Sigma V'$
$\|A-Y\|_F^2 = \|\Sigma - \underbrace{U'YV}_{Z}\|_F^2 = \|\Sigma - Z\|_F^2$
rank(Z) = rank$(U'YV)$ = rank(Y)
$\|\Sigma - Z\|_F^2 = \sum_{i\neq j}Z_{ij}^2 + \sum_{i=r+1}^{\min(m,n)}Z_{ii}^2 + \sum_{i=1}^r(\sigma_i - Z_{ii})^2$
$Z_{ij} = \begin{cases}0 & \text{if } i\neq j,\ i>j \text{ and } i>m\\ \sigma_i & \text{if } i=j,\ i\le r\end{cases}$

$\Sigma = \begin{bmatrix}\sigma_1&&&0\\&\sigma_2&&\\&&\sigma_3&\\0&&&\sigma_4\end{bmatrix}$ $\Sigma_k = \begin{bmatrix}\sigma_1&&0\\&\sigma_2&\\0&&0\end{bmatrix}$ (k)

$\|\Sigma - \Sigma_k\|_F^2 = \sigma_{k+1}^2 + \cdots + \sigma_r^2$
$\min_{\text{rank}(Z)=k}\|\Sigma - Z\|_F^2 \le \sigma_{k+1}^2 + \cdots + \sigma_r^2$

For any Z s.t. rank$(Z)=k$ there is an orthogonal matrix $[\underbrace{P}_{\mathbb{R}^{m\times k}}\ \underbrace{Q}_{\mathbb{R}^{m\times(m-k)}}]$ s.t.
- the cols of P - basis for row space of Z
- the col's of Q - basis for null space of Z

$\|\Sigma - Z\|_F^2 = \|(\Sigma - Z)[P\ Q]\|_F^2 = \|[(\Sigma-Z)P\ \ (\Sigma-Z)Q]\|_F^2$
$= \|(\Sigma-Z)P\|_F^2 + \|(\Sigma-Z)Q\|_F^2$
$= \|(\Sigma-Z)P\|_F^2 + \|\Sigma Q\|_F^2$
$\ge \|\Sigma Q\|_F^2$

$\min\{\|\Sigma - Z\|_F^2 : \text{rank}(Z)=k\} \ge \min\{\|\Sigma Q\|_F^2 : Q'Q = I_{m-k}\}$
$\mathcal{L}(Q,\Lambda) = \|\Sigma Q\|_F^2 - \text{tr}(\Lambda(Q'Q - I_{m-k}))$, Λ symmetric
$0 = 2\Sigma'\Sigma Q - 2Q\Lambda$
$\Sigma'\Sigma Q = Q\Lambda$
$\Lambda = \Sigma'\Sigma$, Q are eigenvectors of $\Sigma'\Sigma$
Q is m-k columns of the identity matrix
$\min\{\|\Sigma Q\|_F^2 : Q'Q = I_{m-k}\} = \sigma_{k+1}^2 + \cdots + \sigma_r^2$
$\sigma_{k+1}^2 + \cdots + \sigma_r^2 \ge \min\{\|\Sigma - Z\|_F^2 : \text{rank}(Z)=k\}$
$\ge \min\{\|\Sigma Q\|_F^2 : Q'Q = I_{m-k}\}$
$= \sigma_{k+1}^2 + \cdots + \sigma_r^2$

Best rank-k approximation to A is $U\Sigma_kV'$
See Added Question 6

IV. Existence of EVD/SVD

Theorem Suppose $A\in\mathbb{R}^{n\times n}$ symmetric matrix. Then, there exists an orthogonal matrix P and a real diagonal matrix Λ s.t. $A = P\Lambda P'$
Moreover, if λ_i is the ith diagonal of Λ and p_i is the ith column of P, $Ap_i = \lambda_ip_i$

Proof: $f(\lambda) = \det(A-\lambda I) = 0$. By the fundamental theorem of algebra, there exists a $\lambda_1\in\mathbb{C}$ s.t. $f(\lambda_1) = 0$. That is $\exists v_1$ s.t. $Av_1 - \lambda_1v_1 = 0$ (normalized)
$\lambda_1 = \bar{v}_1'\underbrace{Av_1}_{\lambda_1v_1} = \bar{v}_1'\bar{A}v_1 = \bar{\lambda}_1\bar{v}_1'v_1 = \bar{\lambda}_1$
(1) Therefore λ_1 is real
(2) $Av_1 = \lambda_1v_1$
$A(\text{Re}(v_1) + i\,\text{Im}(v_1)) = \lambda_1(\text{Re}(v_1) + i\,\text{Im}(v_1))$
$A(\text{Re}(v_1)) - \lambda_1(\text{Re}(v_1)) = -i(A\,\text{Im}(v_1) - \lambda_1\text{Im}(v_1))$
$A\,\text{Re}(v_1) = \lambda_1\text{Re}(v_1)$
$A\,\text{Im}(v_1) = \lambda_1\text{Im}(v_1)$
$p_1 = \frac{\text{Re}(v_1)}{\|\text{Re}(v_1)\|_2}$

Existence of decomposition: $Ap_1 = \lambda_1p_1$
Let H_1 be the householder reflection such that $H_1'p_1 = e_1$
$\lambda_1e_1 = \lambda_1H_1'p_1 = H_1'(Ap_1) = H_1'AH_1H_1'p_1 = H_1'AH_1e_1$
$H_1'AH_1 = \begin{bmatrix}\lambda_1 & c'\\0&A_1\end{bmatrix} = \begin{bmatrix}\lambda_1&0\\0&A_1\end{bmatrix}$ (symmetric)
$\begin{bmatrix}1&0\\0&H_2'\end{bmatrix}H_1'AH_1\begin{bmatrix}1&0\\0&H_2\end{bmatrix} = \begin{bmatrix}\lambda_1&&\\&\lambda_2&\\&&A_2\end{bmatrix}$

Theorem: Suppose $A\in\mathbb{R}^{n\times m}$ and let $n\ge m$. Then there exist orthogonal $U\in\mathbb{R}^{n\times n}$, $V\in\mathbb{R}^{m\times m}$, nonnegative scalars $\sigma_1\ge\sigma_2\ge\cdots\ge\sigma_m\ge0$ and a matrix $\Sigma = \begin{bmatrix}\text{diag}(\sigma_1,\dots,\sigma_m)\\0\end{bmatrix}$
s.t. $A = U\Sigma V'$

Proof: $\|A\|_2 = \max_{\|v\|_2=1}\|Av\|_2$. There exists $v_1\in\mathbb{R}^m$ and $\sigma_1\ge0$ s.t.
$\sigma_1 := \|A\|_2 = \|Av_1\|_2$. Let $u_1 = \frac{1}{\sigma_1}Av_1$
Now let Q_1 be an orthogonal matrix whose first column is v_1 and let P_1 be an orthogonal matrix whose first column is u_1. Then
$P_1'AQ_1 = \begin{bmatrix}\sigma_1 & c'\\0&A_1\end{bmatrix}$
$\|P_1'A\|_2 = \|AQ_1\|_2 = \|P_1'AQ_1\|_2 = \|A\|_2 = \sigma_1$ (Question 13, Added Q5)
$\sigma_1^2 = \|A\|_2^2 \ge \|P_1'AQ_1\|_2^2$
$= \|\begin{bmatrix}\sigma_1&c'\\0&A_1\end{bmatrix}\|_2^2$
$\ge \frac{\|\begin{bmatrix}\sigma_1&c'\\0&A_1\end{bmatrix}\begin{bmatrix}s_1\\s_2\end{bmatrix}\|_2^2}{\|\begin{bmatrix}s_1\\s_2\end{bmatrix}\|_2^2}$ $\begin{bmatrix}s_1\\s_2\end{bmatrix} = \begin{bmatrix}\sigma_1\\c\end{bmatrix}$
$\ge \frac{\|\begin{bmatrix}\sigma_1^2 + \|c\|_2^2\\A_1c\end{bmatrix}\|_2^2}{\sigma_1^2 + \|c\|_2^2}$
$\ge \frac{(\sigma_1^2 + \|c\|_2^2)^2 + \|A_1c\|_2^2}{\sigma_1^2 + \|c\|_2^2}$
$\ge \sigma_1^2 + \|c\|_2^2$

Therefore, $c=0$
$P_1'AQ_1 = \begin{bmatrix}\sigma_1&0\\0&A_1\end{bmatrix}$
$\begin{bmatrix}1&0\\0&P_2\end{bmatrix}'P_1'AQ_1\begin{bmatrix}1&0\\0&Q_2\end{bmatrix} = \begin{bmatrix}\sigma_1&&\\&\sigma_2&\\&&A_2\end{bmatrix}$

5. Related Properties.

5.1 Projection onto Subspaces
See Question 27

5.2. Inequalities

Lemma: Let $A, E\in\mathbb{R}^{n\times m}$ and let $\sigma_{max}(\cdot)$, $\sigma_{min}(\cdot)$ denote the largest and smallest singular values of a given matrix. Then,
$\sigma_{max}(A+E) \le \sigma_{max}(A) + \|E\|_2$
$\sigma_{min}(A+E) \ge \sigma_{min}(A) - \|E\|_2$

Proof. See question 26

Lemma (Hoffman-Wielandt Inequality)
Let $A, E\in\mathbb{R}^{n\times m}$ and let $\sigma_k(\cdot)$ denote the k^{th} largest singular value of a given matrix. Let $p = \min(n,m)$. Then
$$\sum_{k=1}^p[\sigma_k(A+E) - \sigma_k(A)]^2 \le \|E\|_F^2$$

Proof:
Let $\Sigma_E = \text{diag}(\sigma_1(A+E), \sigma_2(A+E), \dots, \sigma_p(A+E))$
$\Sigma = \text{diag}(\sigma_1(A), \sigma_2(A), \dots, \sigma_p(A))$
U_E, U, V_E, V be orthogonal matrices so that $A+E = U_E\Sigma_EV_E'$, $A = U\Sigma V'$
$\sum_{k=1}^p[\sigma_k(A+E) - \sigma_k(A)]^2 = \text{tr}[(\Sigma_E - \Sigma)'(\Sigma_E - \Sigma)]$
$= \text{tr}(\Sigma_E^2 - 2\Sigma_E'\Sigma + \Sigma^2)$
$= \text{tr}[V_E\Sigma_E'\Sigma_EV_E' - 2\Sigma_E'\Sigma + V\Sigma'\Sigma V']$
$= \text{tr}[V_E\Sigma_E'U_E'U_E\Sigma_EV_E' - 2\Sigma_E'\Sigma + V\Sigma'U'U\Sigma V']$
$= \text{tr}((A+E)'(A+E) - 2\Sigma_E'\Sigma + A'A)$
$= \|A+E\|_F^2 - 2\,\text{tr}(\Sigma_E'\Sigma) + \|A\|_F^2$
$= \|A+E\|_F^2 - 2\,\text{tr}((A+E)'A) + \|A\|_F^2 + 2\,\text{tr}[(A+E)'A] - 2\,\text{tr}(\Sigma_E'\Sigma)$
$= \|\underbrace{A+E-A}_{E}\|_F^2 + 2\,\text{tr}[(A+E)'A] - 2\,\text{tr}(\Sigma_E'\Sigma)$

We need to show $\text{tr}(\Sigma_E'\Sigma) \ge \text{tr}((A+E)'A) = \text{tr}(V_E\Sigma_E'U_E'U\Sigma V')$
$= \text{tr}(\underbrace{V'V_E}_{P}\Sigma_E'\underbrace{U_E'U}_{Q}\Sigma)$

$\max_{P,Q}\ \text{tr}[P\Sigma_E'Q\Sigma]$
s.t. $PP' = I_n$, $Q'Q = I_n$, $QQ' = I$, $P'P = I_n$

See Added Question 7

5.3. Moore-Penrose Pseudoinverse

Definition (Pseudoinverse) For a matrix $A\in\mathbb{R}^{n\times m}$, a pseudo inverse of A, $A^+\in\mathbb{R}^{m\times n}$, satisfies
- $AA^+A = A$
- $A^+AA^+ = A^+$
- $(AA^+)' = AA^+$
- $(A^+A)' = A^+A$

Theorem (Moore-Penrose) For any matrix, the pseudo inverse exists and is unique

Proof. (Uniqueness). Suppose B, C are pseudo inverses of A
$AB = ACAB = (AC)(AB) = (AC)'(AB)' = C'(A'B'A') = C'A' = (AC)' = AC$
$BA = CA$
$B = B(AB) = (BA)C = CAC = C$

(Existence for Arbitrary Matrix) Suppose if Σ is a diagonal matrix, then Σ^+ exists.
$A = U\Sigma V'$, $A^+ = V\Sigma^+U'$
- $AA^+A = U\Sigma V'V\Sigma^+U'U\Sigma V' = U\Sigma\Sigma^+\Sigma V' = U\Sigma V' = A$
- $A^+AA^+ = V\Sigma^+U'U\Sigma V'V\Sigma^+U' = V\Sigma^+\Sigma\Sigma^+U' = V\Sigma^+U' = A^+$
- $(AA^+)' = (U\Sigma V'V\Sigma^+U')' = U(\Sigma\Sigma^+)'U' = U\Sigma\Sigma^+U' = AA^+$
- $(A^+A)' = (V\Sigma^+U'U\Sigma V')' = (V\Sigma^+\Sigma V')' = V\Sigma^+\Sigma V' = A^+A$

(Pseudoinverse for diagonal matrix)
Added Question 8

Added Question 9

VI. Computing the Schur Decomposition.

Added Question 10 (implement)
Added Question 12 - asks you to extend to get SVD

6.1. Largest Eigenvalue

Power Method
- p_0 is arbitrary, $\frac{p_0}{\|p_0\|_2} = p_0$
- $p_k = \frac{Ap_{k-1}}{\|Ap_{k-1}\|_2}$, $k\in\mathbb{N}$
- $\gamma_k = p_k'Ap_k$

Theorem: Suppose $A\in\mathbb{R}^{n\times n}$ symmetric, $Q'AQ = \Lambda$ where Q is orthogonal and Λ is diagonal with entries $\lambda_1,\dots,\lambda_n$ where $|\lambda_1| > |\lambda_2| \ge |\lambda_3| \ge \cdots \ge |\lambda_n|$
Define $\cos(\theta_k) = |p_k'q_1|$
If $\cos(\theta_0)\neq0$
1. $|\sin(\theta_k)| \le \tan(\theta_0)\left|\frac{\lambda_2}{\lambda_1}\right|^k$
2. $|\gamma_k - \lambda_1| \le 2|\lambda_1|\tan(\theta_0)^2\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}$

Proof
By definition
$|\sin(\theta_k)|^2 = 1 - (q_1'p_k)^2 = 1 - \left(\frac{q_1'A^kp_0}{\|A^kp_0\|_2}\right)^2$
Let $a = Q'p_0$. Then
$A^kp_0 = Q\Lambda^kQ'p_0 = Q\Lambda^ka = \sum_{i=1}^n\lambda_i^ka_iq_i$
$|\cos(\theta_0)| = |q_1'p_0| = |a_1| \neq 0$
$|\sin(\theta_k)|^2 = 1 - \frac{\lambda_1^{2k}a_1^2}{\sum_{i=1}^n\lambda_i^{2k}a_i^2} = \frac{\sum_{i=2}^n\lambda_i^{2k}a_i^2}{\sum_{i=1}^n\lambda_i^{2k}a_i^2} \le \frac{\sum_{i=2}^n\lambda_i^{2k}a_i^2}{\lambda_1^{2k}a_1^2}$
$= \frac{1}{a_1^2}\sum_{i=2}^na_i^2\left(\frac{\lambda_i}{\lambda_1}\right)^{2k} \le \frac{1}{a_1^2}\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}\sum_{i=2}^na_i^2$
$= \frac{(1-a_1^2)}{a_1^2}\left(\frac{\lambda_2}{\lambda_1}\right)^{2k} = \tan(\theta_0)^2\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}$

$\gamma_k - \lambda = p_k'Ap_k - \lambda_1 = \frac{p_0'A^{2k+1}p_0}{p_0'A^{2k}p_0} - \lambda_1 = \frac{\sum_{i=1}^na_i^2\lambda_i^{2k+1}}{\sum_{i=1}^na_i^2\lambda_i^{2k}} - \lambda_1$
$= \frac{\sum_{i=2}^na_i^2\lambda_i^{2k}(\lambda_i - \lambda_1)}{\sum_{i=1}^na_i^2\lambda_i^{2k}}$
$|\lambda_i - \lambda_1| \le 2|\lambda_1|$ for any $i\neq1$
$|\gamma_k - \lambda_1| \le \frac{2|\lambda_1|(1-a_1^2)}{a_1^2}\left(\frac{\lambda_2}{\lambda_1}\right)^{2k} = 2|\lambda_1|\tan(\theta_0)^2\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}$

6.2 Largest K Eigenvalues

Orthogonal Iteration
- $P_0\in\mathbb{R}^{n\times r}$ whose columns are orthonormal
- $P_k = \text{orth}(AP_{k-1})$

$\lambda_1,\dots,\lambda_k$ diagonal of $P_k'AP_k$

Lemma: (Added Question 11)
Suppose that V and W are two r-dimensional subspace in $\mathbb{R}^n$
Let Q_V, Q_W be matrices whose columns for orthonormal bases of V and W, resp. Then
$\min\left\{\frac{|v'w|}{\|v\|_2\|w\|_2} : v\in V, w\in W\right\} = \sigma_r(Q_V'Q_W)$

Let N_V, N_W be matrices s.t. $[Q_V\ N_V]$ and $[Q_W\ N_W]$ are orthogonal, P_V, P_W are orthogonal projections onto V and W, then
$\text{dist}(V,W) := \|P_V - P_W\|_2 = \|P_V'N_W\|_2 = \|P_W'N_V\|_2$

Theorem Suppose $A\in\mathbb{R}^{n\times n}$ symmetric, and $Q'AQ = \Lambda$
$|\lambda_1|\ge|\lambda_2|\ge\cdots\ge|\lambda_r| > |\lambda_{r+1}| \ge \cdots \ge |\lambda_n|$. Let Q_α be the first r columns of Q. If
$\cos(\theta) = \min\left\{\frac{|u'v|}{\|u\|_2\|v\|} : u\in\text{range}(Q_\alpha), v\in\text{range}(P_0)\right\} > 0$
Then, $\text{dist}(\text{range}(Q_\alpha), \text{range}(P_k)) \le \tan(\theta)\left|\frac{\lambda_{r+1}}{\lambda_r}\right|^k$

Proof Let R_k denote the upper triangular matrix s.t. $P_kR_k = AP_{k-1}$
- $P_kR_kR_{k-1} = AP_{k-1}R_{k-1} = A^2P_{k-2}$
- $P_kR_kR_{k-1}\cdots R_1 = A^kP_0$

Now let Q_β have orthonormal columns s.t. $Q = [Q_\alpha\ Q_\beta]$
$Q'P_k = \begin{bmatrix}Q_\alpha'P_k\\Q_\beta'P_k\end{bmatrix} = \begin{bmatrix}V_k\\W_k\end{bmatrix}$
let $\Lambda = \begin{bmatrix}\Lambda_\alpha & 0\\0&\Lambda_\beta\end{bmatrix}$ ($r\times r$, $n-r\times n-r$)
$P_kR_kR_{k-1}\cdots R_1 = A^kP_0 = [Q_\alpha\ Q_\beta]\begin{bmatrix}\Lambda_\alpha^k&\\&\Lambda_\beta^k\end{bmatrix}[Q_\alpha\ Q_\beta]'P_0$
$\begin{bmatrix}V_k\\W_k\end{bmatrix}R_kR_{k-1}\cdots R_1 = \begin{bmatrix}\Lambda_\alpha^k&0\\0&\Lambda_\beta^k\end{bmatrix}\begin{bmatrix}V_0\\W_0\end{bmatrix}$
- $\cos(\theta) = \sigma_r(V_0) = \sqrt{1 - \|W_0\|_2^2}$
- $\text{dist}(\text{range}(Q_\alpha), \text{range}(P_k)) = \|W_k\|_2$
- $\Lambda_\alpha^kV_0 = V_k(R_k\cdots R_1)$
- $\Lambda_\beta^kW_0 = W_k(R_k\cdots R_1)$

Since V_0 is nonsingular by assumption and $R_k\cdots R_1$ is nonsingular
$W_k = \Lambda_\beta^kW_0(R_k\cdots R_1)^{-1} = \Lambda_\beta^kW_0(V_k^{-1}\Lambda_\alpha^kV_0)^{-1}$
$= \Lambda_\beta^kW_0V_0^{-1}\Lambda_\alpha^{-k}V_k$
$\|V_k\|_2 \le 1$
$\|W_k\|_2 \le \frac{|\lambda_{r+1}|^k\|W_0\|\|V_0^{-1}\|}{|\lambda_r|^k}$
$= \left|\frac{\lambda_{r+1}}{\lambda_r}\right|^k\frac{\sin(\theta)}{\cos(\theta)}$

6.3. QR iteration
- P_0 that is orthogonal ($P_0\in\mathbb{R}^{n\times n}$)
- $P_k = \text{orth}(AP_{k-1})$
- $T_k = P_k'AP_k$

QR iteration
$P_{k+1}R_{k+1} = AP_k$
$T_k = P_k'AP_k = P_k'P_{k+1}R_{k+1}$
$T_{k+1} = P_{k+1}'AP_{k+1} = P_{k+1}'A(P_kP_k')P_{k+1}$
$= P_{k+1}'(AP_k)P_k'P_{k+1}$
$= P_{k+1}'(P_{k+1}R_{k+1})P_k'P_{k+1}$
$= R_{k+1}P_k'P_{k+1}$

QR iteration:
- U_0 arbitrary orthogonal matrix. $T_0 = U_0'AU_0$
- $\underbrace{U_k}_{\text{orth}}, \underbrace{S_k}_{\text{upper}} \leftarrow QR(T_k)$
- $T_{k+1} = S_k\cdot U_k$

7. Large-Scale dimensionality reduction

$A\in\mathbb{R}^{n\times m}$, $m \gg n$
$U\Sigma V' = A$, $U\Sigma$
Inexact rank-k approximation to column space of A
$H \approx$ left singular vectors of A (k)
$S \approx$ singular values of A (k)
$R \approx$ right singular vectors $= S^{-1}H'A$
$\|A - HSR\|_F^2 = \|A - HSS^{-1}H'A\|_F^2 = \|A - HH'A\|_F^2$

Theorem: Suppose $A\in\mathbb{R}^{n\times m}$ and H, C are constructed from the algorithm. Then
$\|A - HH'A\|_F^2 \le \|A - A_k\|_F^2 + 2\sqrt{k}\|AA' - CC'\|_F$
where A_k is the best rank k approximation to A

Proof
$\|A - HH'A\|_F^2 = \text{tr}((A - HH'A)'(A - HH'A))$
$= \text{tr}(A'A - 2A'HH'A + A'HH'HH'A)$
$= \text{tr}(A'A - A'HH'A)$
$= \|A\|_F^2 - \|A'H\|_F^2$
$= \|A - A_k\|_F^2 + \|A_k\|_F^2 - \|A'H\|_F^2$
$= \|A - A_k\|_F^2 + \sum_{j=1}^k\sigma_j(A)^2 - \|A'H\|_F^2$
$= \|A - A_k\|_F^2 + (\sum_{j=1}^k(\sigma_j(A)^2 - \sigma_j(C)^2))$ ①
$+ (\sum_{j=1}^k\sigma_j(C)^2 - \|A'H\|_F^2)$ ②

① $|\sum_{j=1}^k\sigma_j(A)^2 - \sigma_j(C)^2| \le \sqrt{k}\sqrt{\sum_{j=1}^k(\sigma_j(A)^2 - \sigma_j(C)^2)^2}$
$= \sqrt{k}\sqrt{\sum_{j=1}^k(\sigma_j(AA') - \sigma_j(CC'))^2}$
$\le \sqrt{k}\|AA' - CC'\|_F^2$

② Problem 28. Let H_j denote the columns of H
$|\sum_{j=1}^k\sigma_j(C)^2 - \|A'H_j\|_2^2| \le \sqrt{k}\sqrt{\sum_{j=1}^k(\sigma_j(CC') - H_j'AA'H_j)^2}$
$= \sqrt{k}\sqrt{\sum_{j=1}^k(H_j'(CC' - AA')H_j)^2}$
$\le \sqrt{k}\|CC' - AA'\|_F$

Iterative Linear Solvers

I. Motivating Problem
II. Splitting Methods
III. Column-Action Methods (Gradient Descent, CG, GMRES)
IV. Row-Action Methods (Kaczmarz, Sketching)

I. Motivating Problem

A gaussian process is a time-continuous stochastic process, $\{X_t : t\in[0,T]\}$ s.t. for any $N\in\mathbb{N}$ and for any distinct $t_1,\dots,t_N\in[0,T]$

$$\begin{bmatrix}X_{t_1}\\X_{t_2}\\\vdots\\X_{t_N}\end{bmatrix} \sim N\left(\underbrace{\begin{bmatrix}m(t_1)\\m(t_2)\\\vdots\\m(t_N)\end{bmatrix}}_{M(1:N)}, \underbrace{\begin{bmatrix}K(t_1,t_1)&K(t_1,t_2)&\cdots&K(t_1,t_N)\\K(t_2,t_1)&K(t_2,t_2)&\cdots&K(t_2,t_N)\\\vdots&&\ddots&\vdots\\K(t_N,t_1)&&\cdots&K(t_N,t_N)\end{bmatrix}}_{K(1:N)}\right)$$

where m is a mean function, and $K(\cdot,\cdot)$ is the covariance function

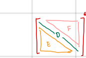
1. Bayesian Statistics. GP is a prior for nonparametric Bayes
2. Weather analysis or Black box models, a GP is used as a model to relate inputs to outputs.
3. Stochastic optimization, GP is used as a model for our objective function
4. Machine learning, GP are often used as models in semi-supervised learning

$K(1:N)y = m(1:N)$
N is very large

II. Splitting Methods

$A\in\mathbb{R}^{n\times n}$, $b\in\mathbb{R}^n$ $Ax = b$
Decomposes $A = A_1 + A_2$ and generates $\{x_k : k\in\mathbb{N}\}$
1. x_0 is arbitrary
2. $x_{k+1} = -A_1^{-1}A_2x_k + A_1^{-1}b$

- $D\in\mathbb{R}^{n\times n}$ diagonal matrix s.t. $D_{ii} = A_{ii}$
- $E\in\mathbb{R}^{n\times n}$ a lower triangular matrix s.t. $A_{ij} = E_{ij}$ $i>j$
- $F\in\mathbb{R}^{n\times n}$ an upper triangular matrix s.t. $A_{ij} = F_{ij}$ if $i<j$



2.1 Methods

2.1.1. Jacobi iteration
We have a current iterate (x_k, x_k), $x_k[i]$
$r_k = Ax_k - b$
$r_k[i] = \sum_{j=1}^nA_{ij}x_k[j] - b[i]$

for $i=1,\dots,n$
$0 = \sum_{j\neq i}A_{ij}x_k[j] + A_{ii}x_{k+1}[i] - b[i]$

If $A_{ii}\neq0$
$x_{k+1}[i] = \frac{1}{A_{ii}}\left(-\sum_{j\neq i}A_{ij}x_k[j] + b[i]\right)$

$0 = (E+F)x_k + Dx_{k+1} - b$
If D is invertible.
$x_{k+1} = -D^{-1}(E+F)x_k + D^{-1}b$
- $G = -D^{-1}(E+F)$
- $f = D^{-1}b$

$x_{k+1} = Gx_k + f$

2.1.2. Gauss-Seidel Method
$i=1$: $x_{k+1}[1] = \frac{-1}{A_{11}}(\sum_{j=2}^nA_{1j}x_k[j] - b[1])$
$i=2$: $x_{k+1}[2] = \frac{-1}{A_{22}}(A_{21}x_{k+1}[1] + \sum_{j=3}^nA_{2j}x_k[j] - b[2])$
i: $x_{k+1}[i] = \frac{-1}{A_{ii}}(\sum_{j=1}^{i-1}A_{ij}x_{k+1}[j] + \sum_{j=i+1}^nA_{ij}x_k[j] - b[i])$

$x_{k+1} = -D^{-1}(Ex_{k+1} + Fx_k - b)$
$Dx_{k+1} = -Ex_{k+1} - Fx_k + b$
$(D+E)x_{k+1} = -Fx_k + b$
$x_{k+1} = \underbrace{-(D+E)^{-1}F}_{G}x_k + \underbrace{(D+E)^{-1}b}_{f}$

$x_{k+1} = Gx_k + f$
Added Question 13 for Backward Gauss-Seidel

2.1.3. Successive Over Relaxation (SOR)
$\omega>0$
$x_{k+1} = -(D+\omega E)^{-1}(\omega F + (\omega-1)D)x_k + \omega(D+\omega E)^{-1}b$
- $G = -(D+\omega E)^{-1}(\omega F + (\omega-1)D)$
- $f = (D+\omega E)^{-1}b$

$x_{k+1} = Gx_k + f$.

Added question 14: implementation + details backward SOR
Question 29: symmetric SOR
Question 31: deeper exploration of splitting methods

2.2. Jordan Canonical Form

Spectral Radius $\rho(G) = \max|\lambda_i(G)|$

2.2.1. Defective Matrices
$\begin{bmatrix}2&1\\0&2\end{bmatrix}$. $0 = (2-\lambda)^2$ $\begin{cases}2v_1 + v_2 = 2v_1\\2v_2 = 2v_2\end{cases}$
$\lambda=2$ $e_1\in\mathbb{R}^2$

Definition: A square matrix A is defective if its eigenvectors do not form a basis for the ambient space

Lemma (Jordan Block) For any $\lambda\in\mathbb{R}$ the matrix $A\in\mathbb{R}^{n\times n}$
$A_{ij} = \begin{cases}\lambda & \text{if } i=j\\1 & \text{if } i+1=j\\0&\text{otherwise}\end{cases}$ $\begin{bmatrix}\lambda&1&&\\&\lambda&1&\\&&\ddots&\ddots\\&&&\lambda\end{bmatrix}$
is defective.

Proof: In added question 15, you prove det(A)
$\det(A) = \sum_{s=1}^nA_{rs}d_{rs}$

where A_{rs} is the matrix A with r^{th} row and s^{th} column removed, and $d_{rs} = \det(A_{rs})$

Base case: $n=2$. $\begin{bmatrix}\lambda&1\\0&\lambda\end{bmatrix}$ so all eigenvalues are λ
Induction hypothesis: up to $n-1$
For case n
$\det(A - \gamma I) = \sum_{s=1}^n\underbrace{B_{ns}(\gamma)}_{B(\gamma)}d_{ns}(\gamma) = \underbrace{(\lambda-\gamma)}_{\lambda}\underbrace{\det(A_{n-1} - \gamma I_{n-1})}_{\lambda}$

Eigenvectors: $(A - \lambda I)v = 0$
$0 = \begin{bmatrix}0&1&&0\\&0&\ddots&\\&&\ddots&1\\0&&&0\end{bmatrix}\begin{bmatrix}v_1\\v_2\\\vdots\\v_n\end{bmatrix} = \begin{bmatrix}v_2\\v_3\\\vdots\\v_n\\0\end{bmatrix}$
$v = e_1$

2.3. Generalized Eigenvector

Definition: For a square matrix A with eigenvalue λ, the generalized eigenvector of rank m is the vector v_m s.t.
$(A - \lambda I)^mv_m = 0$, $(A - \lambda I)^{m-1}v_m \neq 0$

Definition: Let v_m be the generalized eigenvector of rank m of a matrix A for eigenvalue λ. For $j = 1,2,\dots,m-1$, define
$v_j = (A - \lambda I)v_{j+1} = \cdots = (A - \lambda I)^{m-j}v_m$
$\{v_1,\dots,v_m\}$ a Jordan chain.

Lemma: Suppose $\{v_1,\dots,v_m\}$ is a Jordan chain for eigenvalue λ of matrix A. Then
1. Jordan chain is linearly independent
2. v_j is a rank j generalized eigenvector of A corresponding to λ.

Proof $j=1$, $v_{m-1} = (A - \lambda I)v_m$
$(A - \lambda I)^{m-1}v_{m-1} = (A - \lambda I)^mv_m = 0$
$(A - \lambda I)^{m-2}v_{m-1} = (A - \lambda I)^{m-1}v_m \neq 0$
By induction, we have claim 2.

$v_m = \sum_{j=1}^{m-1}\alpha_jv_j$ for scalars α_j not all zero
$0 \neq (A - \lambda I)^{m-1}v_m = \sum_{j=1}^{m-1}\alpha_j(A - \lambda I)^{m-1}v_j = \sum_{j=1}^{m-1}\alpha_j(A - \lambda I)^{m-1-j}(A - \lambda I)^jv_j = 0$

Lemma: Let J be a jordan block of dimension n and eigenvalue λ. Then, there exists a rank n generalized eigenvector, and the corresponding Jordan chain forms a basis for $\mathbb{R}^n$.

Proof For any $k\le n-1$, $(J - \lambda I)^k$ is the matrix with ones in $k+1$ diagonal. For $k\ge n$, $(J - \lambda I)^k = 0$. For any $k>n$, and any v,
$0v = (J - \lambda I)^kv = (J - \lambda I)^{k-1}v$
e_n as a ansatz for a rank n generalized eigenvector.
$0 = \underbrace{(J - \lambda I)^n}_{0}e_n$
$(J - \lambda I)^{n-1}e_n = \begin{bmatrix}0&1\\&0\end{bmatrix}e_n = e_1 \neq 0$

See added question 16.

2.3.1 Jordan Canonical Form

Theorem: Let $A\in\mathbb{R}^{n\times n}$. Let $\lambda_1,\dots,\lambda_r$ denote the distinct eigenvalues of A and let
- μ_i denote the algebraic multiplicity of λ_i. That is, the number of times λ_i appears as a root of $\det(A - \gamma I)$
- δ_i denote the geometric multiplicity of λ_i. That is, the number of linearly independent eigenvectors corresponding to λ_i

Then there exists an invertible matrix M s.t.
$A = M\begin{bmatrix}J_1&&0\\&J_2&\\0&&\ddots\ J_n\end{bmatrix}M^{-1}$
where
- J_i are Jordan blocks
- There are δ_i Jordan blocks corresponding to λ_i and the sum of the dimensions of these blocks is μ_i
- The columns of M that correspond to block J_i are a Jordan chain corresponding to the eigenvalue of J_i

2.3.2. Splitting Method Convergence.

Theorem: Suppose for a matrix G and vector f there exists an x^* with
$x^* = Gx^* + f$
For any x_0, define $x_{k+1} = Gx_k + f$, $k+1\in\mathbb{N}$
The sequence $\{x_k\} \to x^*$ if and only if
$\rho(G) := \max|\lambda_i| < 1$

Proof By construction, $x^* = Gx^* + f$.
$x_{k+1} - x^* = G(x_k - x^*)$
$G = MJM^{-1}$ let $\delta_k = M^{-1}(x_k - x^*)$.
$\delta_{k+1} = J\delta_k = J^{k+1}\delta_0$
$\delta_k\to0 \iff x_k\to x^*$
$J^k\to0$. By Question 30 and Added Question 17,
$\delta_k\to0$ for arbitrary $\delta_0 \iff \rho(G) < 1$

Added Question 18, spectral radius and matrix norms.

3. Column Action Methods

- Gradient Descent
- Conjugate Directions/Gradient
- GMRES

3.1. Gradient Descent
$Ax = b$
$\min_y\frac{1}{2}\|Ay - b\|_2^2$
$f(y) = \frac{1}{2}\|Ay - b\|_2^2$
x_0, $\{x_k : k\in\mathbb{N}\}$, $x_{k+1} = x_k - \alpha_k\dot{f}(x_k)$
$\dot{f}(x) = \nabla_yf(y)|_{y=x}$
$\dot{f}(x) = A'(Ax - b)$
$x_{k+1} = x_k - \alpha_kA'(Ax - b)$

Question 33 - theoretical analysis
Question 34, 35 - implementation

Strategies:
1. α_k to minimize
$\{\|A\underbrace{x_{k+1}}_{x_{k+1}(\alpha)} - b\|_2^2 : x_{k+1} = x_k - \alpha A'(Ax_k - b)\}$
2. If A is invertible, $A^{-1}b = x^*$
$\{\|x_{k+1} - x^*\|_2^2 : x_{k+1} = x_k - \alpha A'(Ax_k - b)\}$
3. Let $r_k = Ax_k - b$
$x_{k+1} = x_k - \alpha_kA'(Ax_k - b)$
$r_{k+1} = Ax_{k+1} - b = Ax_k - b - \alpha_kAA'(Ax_k - b)$
$= r_k - \alpha_kAA'r_k$
$= (I - \alpha_kAA')r_k$
$\|r_{k+1}\|_2 \le \|I - \alpha_kAA'\|_2\|r_k\|_2$
$\min\|I - \alpha AA'\|_2$
4. A^{-1}, $x^* = A^{-1}b$
$x_{k+1} - x^* = x_k - x^* - \alpha_kA'(Ax_k - Ax^*)$
$x_{k+1} - x^* = (I - \alpha_kA'A)(x_k - x^*)$
$\min_\alpha\|I - \alpha A'A\|_2$

Strategy 1:
$\min_\alpha\{\|Ax_{k+1}(\alpha) - b\|_2^2 : x_{k+1}(\alpha) = x_k - \alpha A'(Ax_k - b)\}$
$\min\|Ax_k - \alpha AA'r_k - b\|_2^2$
$\min\|r_k - \alpha AA'r_k\|_2^2$
$\min\|r_k\|_2^2 - 2\alpha\|A'r_k\|_2^2 + \alpha^2\|AA'r_k\|_2^2$
$0 = -2\|A'r_k\|_2^2 + 2\alpha\|AA'r_k\|_2^2$ Question 33(a)
$\alpha_k = \frac{\|A'r_k\|_2^2}{\|AA'r_k\|_2^2}$

Theorem: Suppose A is invertible. Then the sequence of iterates $\{x_k\}$ with x_0 arbitrary and $x_{k+1} = x_k - \alpha_k\dot{f}(x_k)$
with $\alpha_k = \frac{\|A'r_k\|_2^2}{\|AA'r_k\|_2^2}$ converges to x^*

$\|r_k\|_2^2 \le \left(1 - \frac{4\sigma_1^2\sigma_n^2}{(\sigma_1^2 + \sigma_n^2)^2}\right)^k\|r_0\|_2^2$

Proof

$\|r_{k+1}\|_2^2 = \|r_k\|_2^2 - 2\alpha_k \|A'r_k\|_2^2 + \alpha_k^2 \|AA'r_k\|_2^2$

$= \|r_k\|_2^2 - \frac{\|A'r_k\|_2^4}{\|AA'r_k\|_2^2}$

$= \left(1 - \frac{\|A'r_k\|_2^4}{\|r_k\|_2^2\|AA'r_k\|_2^2}\right)\|r_k\|_2^2$

$A = U\Sigma V'$

$1 - \frac{\|\Sigma' U' r_k\|_2^2}{\|U' r_k\|_2^2 \|\Sigma\Sigma' U' r_k\|_2^2} = 1 - \frac{\|w\|_2^4}{\|\Sigma^{-1} w\|_2^2 \|\Sigma w\|_2^2}$

$w = \Sigma' U' r_k$

$1 - \frac{\|A'r_k\|_2^4}{\|r_k\|_2^2\|AA'r_k\|_2^2} \le \max_w 1 - \frac{\|w\|_2^4}{\|\Sigma^{-1}w\|_2^2\|\Sigma w\|_2^2}$

$= 1 - \frac{1}{\max_{\|w\|_2=1} \|\Sigma^{-1}w\|_2^2\|\Sigma w\|_2^2}$

$\le 1 - \frac{4\sigma_1^2\sigma_n^2}{(\sigma_1^2+\sigma_n^2)^2}$ Question 33(b)

$\|r_{k+1}\|_2^2 \le \left(1 - \frac{4\sigma_1^2\sigma_n^2}{(\sigma_1^2+\sigma_n^2)^2}\right)\|r_k\|_2^2$

3.2. Orthogonalized Search Directions

$1 - \frac{4\kappa(A)^2}{(1+\kappa(A)^2)^2}$

3.2.1 Orthogonalizing solution space.

Suppose A is invertible. Let x_0 be arbitrary and let $\{x_k\}$

$x_{k+1} = x_k - \alpha_k s_k$

where $\{s_0, s_1, \dots\}$ form an orthogonal set

$x_n = x_{n-1} - \alpha_{n-1}s_{n-1} = x_{n-2} - \alpha_{n-2}s_{n-2} - \alpha_{n-1}s_{n-1}$

$= x_0 - \sum_{j=0}^{n-1}\alpha_j s_j$

$s_k' x_n = s_k' x_0 - \alpha_k \|s_k\|_2^2$

$\frac{s_k'(x_0 - x_n)}{\|s_k\|_2^2} = \alpha_k$

$x_0 - x^* = \sum_{j=0}^{n-1}\gamma_j s_j$

$\frac{s_k'(x_0-x^*)}{\|s_k\|_2^2} = \gamma_k$

$x_n = x^* \iff \gamma_k = \alpha_k$

$\gamma_k = \frac{s_k' A^{-1}(Ax_0 - b)}{\|s_k\|_2^2} = \frac{s_k' A^{-1} r_0}{\|s_k\|_2^2}$

3.2.2. Conjugation: Orthogonalization in Residual Space

A is symmetric positive definite. Question 36 (a) that we can define an inner product $\langle v,w\rangle_A = v'Aw$

Definition: A set vectors $\{s_0,\dots,s_r\}$ are A-conjugated if they are orthogonal w.r.t. $\langle\cdot,\cdot\rangle_A$

Lemma: If $\{s_0,\dots,s_r\}$ are nonzero and A-conjugated, then they are linearly independent.

Proof Suppose $\{s_0,\dots,s_r\}$ are not linearly independent.

$s_0 = \sum_{i=1}^r \alpha_i s_i$

$s_0'As_0 = \sum_{i=1}^r \alpha_i s_0' A s_i = 0$

$Ax=b$; let x_0 be arbitrary

$x_{k+1} = x_k - \alpha_k s_k$

where s_k are A-conjugated.

$x_n = x_0 - \sum_{j=0}^{n-1}\alpha_j s_j$

$x^* = x_0 - \sum_{j=0}^{n-1}\gamma_j s_j$

Can we compute $\alpha_j = \gamma_j$?

$x_0 - x^* = \sum_{j=0}^{n-1}\gamma_j s_j$

$s_k'A(x_0 - x^*) = \sum_{j=0}^{n-1}\gamma_j s_k'As_j = \gamma_k s_k'As_k$

$\alpha_k = \gamma_k = \frac{s_k'A(x_0-x^*)}{s_k'As_k} = \frac{s_k' r_0}{s_k'As_k}$

To compute s_k we do gram-schmidt.

$\|v\|_A = \sqrt{\langle v,v\rangle_A}$

$\{a_0,\dots,a_n\}$ a set of vectors.

1) $s_0 = \frac{a_0}{\|a_0\|_A}$

2) $s_j = \frac{a_j - \sum_{i=0}^{j-1}\langle a_j, s_i\rangle_A s_i}{\|a_j - \sum_{i=0}^{j-1}\langle a_j,s_i\rangle_A s_i\|_A}$

Algorithm (Method of Conjugate Directions)

(1) we need an x_0 and a vector v_0.

Compute $r_0 = Ax_0 - b$. $\|r_0\|_A$ small? ⇒ stop

(2) $\|r_0\|_A$ not small, $s_0 = v_0/\|v_0\|_A$

$x_1 = x_0 - \alpha_0 s_0$; $\alpha_0 = \frac{r_0's_0}{s_0'As_0}$

(a) r_k small? no continue

(b) v_k, s.t. $\{v_0,\dots,v_{k-1}\}$ are lin. ind.

(c) using Gram-Schmidt and $s_0,\dots,s_{k-1}$

$s_k = \frac{v_k - \sum_{i=0}^{k-1}\langle v_k, s_i\rangle_A s_i}{\|v_k - \sum_{i=0}^{k-1}\langle v_k,s_i\rangle_A s_i\|_A}$

(d) $x_{k+1} = x_k - \alpha_k s_k$; $\alpha_k = \frac{r_0's_k}{s_k'As_k} = r_0's_k$

3.3. Conjugate Gradients

Basic idea: use residuals to find A-conjugate set

(1) Given x_0 compute $r_0 = Ax_0 - b$.

(2) $s_0 = r_0/\|r_0\|_A$

$x_1 = x_0 - \alpha_0 s_0$; $\alpha_0 = \frac{r_0's_0}{s_0'As_0} = \frac{\|r_0\|_2^2}{\|r_0\|_A}$

(a) r_k (is it small?)

(b) using Gram-Schmidt

$s_k = \frac{r_k - \sum_{i=0}^{k-1}\langle r_k,s_i\rangle_A s_i}{\|r_k - \sum_{i=0}^{k-1}\langle r_k,s_i\rangle_A s_i\|_A}$

(c) $x_{k+1} = x_k - \alpha_k s_k$; $\alpha_k = \frac{r_0's_k}{s_k'As_k} = r_0's_k$

Lemma

$r_0's_k = \alpha_k = \arg\min_\alpha \{\|v - x^*\|_A^2 : v = x_k - \alpha s_k\} = \frac{r_k's_k}{\|s_k\|_A^2} = r_k's_k$

$r_k's_j = 0$ for $j = 0,\dots,k-1$.

$s_k = \frac{r_k - \langle r_k, s_{k-1}\rangle_A s_{k-1}}{\|r_k - \langle r_k,s_{k-1}\rangle_A s_{k-1}\|_A}$

Proof

* $r_k = Ax_k - b = Ax_{k-1} - \alpha_k A s_{k-1} - b = r_{k-1} - \alpha_{k-1}As_{k-1}$

\+ $As_{k-1} = \frac{r_{k-1} - r_k}{\alpha_{k-1}}$

(1) Base case: $l=0$. $\alpha_0 = r_0's_0$

(2) Induction hypothesis: we assume that all 3 properties hold for $j \le l$.

(3). Generalize to $l+1$

(a) claim 1:

$r_{l+1} = r_0 - \sum_{j=0}^{l}\alpha_j A s_j$

$s_{l+1}'r_{l+1} = s_{l+1}'r_0 - \sum_{j=0}^{l}\alpha_j s_{l+1}'As_j = s_{l+1}'r_0 = \alpha_{l+1}$

(b) claim 2:

$As_l = \frac{r_l - r_{l+1}}{\alpha_l}$

By conjugation, for $j<l$

$0 = s_j'As_l = \frac{s_j'r_l - s_j'r_{l+1}}{\alpha_l} \iff s_j'r_{l+1} = s_j'r_l$ ($=0$)

$s_j'r_{l+1} = 0$ for $j<l$.

when $j = l$

$1 = s_l'As_l = \frac{s_l'r_l - s_l'r_{l+1}}{\alpha_l}$

$s_l'r_l = \alpha_l = s_l'r_l - s_l'r_{l+1} \Rightarrow s_l'r_{l+1} = 0$

$s_j'r_{l+1} = 0 \quad \forall j \le l$.

(c) Claim 3

By construction: $r_0 \in \text{span}(s_0)$, $r_1 \in \text{span}(s_0,s_1)$, ⋮, $r_j \in \text{span}(s_0,\dots,s_j)$

$r_{l+1}'s_j = 0$ for $j=0,\dots,l$

$r_{l+1}'r_j = 0$ for $j = 0,\dots,l$

† $As_j = \frac{r_j - r_{j+1}}{\alpha_j}$

For $j \le l$

$r_{l+1}'As_j = \frac{r_{l+1}'r_j - r_{l+1}'r_{j+1}}{\alpha_j} = \begin{cases}0 & j+1 \le l\\ -\frac{\|r_{l+1}\|_2^2}{\alpha_l} & j = l\end{cases}$

$s_{l+1} \propto r_{l+1} - \sum_{k=0}^{l}(r_{l+1}'As_k)s_k = r_{l+1} - (r_{l+1}'As_l)s_l$.

1. Given x_0, compute $r_0 = Ax_0 - b$ (If r_0 small stop)
2. $s_0 = r_0/\|r_0\|_A$

$x_1 = x_0 - \alpha_0 s_0$; $\alpha_0 = r_0's_0$

(a) r_k is small?

(b) Use Gram-Schmidt to compute s_k

$s_k = \frac{r_k - (r_k'As_{k-1})s_{k-1}}{\|r_k - (r_k'As_{k-1})s_{k-1}\|_A}$

(c) $x_{k+1} = x_k - \alpha_k s_k$; $\alpha_k = r_k's_k$

$1 - \frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)}+1}$

3.4 Krylov Spaces

- Subspaces $\{K_m\}$ s.t. $K_m \subset K_{m+1}$ ← iterate space
- Subspaces $\{L_m\}$ s.t. $L_m \subset L_{m+1}$ ← residual space
- $x_m - x_0 \in K_m$ and $Ax_m - b \perp L_{m+1}$

Added Question 19

Definition: A Krylov space of a matrix $A\in\mathbb{R}^{n\times n}$ with respect to a vector $v \in \mathbb{R}^n$ of order m is

$K_m(A,v) = \text{span}[v, Av, A^2v, \dots, A^{m-1}v]$

Recall that a monic (i.e, of a single variable) polynomial of degree $q \in \mathbb{Z}_{\ge0}$ is $p(x) = \sum_{j=0}^q \alpha_j x^j$

Let P_q to be the space of all monic, degree-q polynomials then $P_{q-1} \subset P_q$

$p(A) = \sum_{j=0}^q \alpha_j A^j$

Theorem: Let $p(x) = \det(A - xI)$, then $p(A) = 0$.

Lemma: For any $A\in\mathbb{R}^{n\times n}$ and $v \in\mathbb{R}^n$

$K_m(A,v) = \{x = p(A)v : p\in P_{m-1}\}$

Proof: Added question 20(a)

Definition: The minimal polynomial of a vector v w.r.t. to A is the minimal degree, nonzero, monic polynomial s.t. $p(A)v = 0$

The degree of the minimal polynomial is called its grade grade(A,v)

Lemma: For any $A\in\mathbb{R}^{n\times n}$ and $v\in\mathbb{R}^n$, grade$(A,v) \le n$.

Proof: Added Question 20(b)

Lemma: Let q = grade(A,v). Then

1. $K_q(A,v) = K_m(A,v)$ for all $m \ge q$
2. $K_q(A,v)$ is invariant under A.

$\forall x\in K_q(A,v)$, $Ax \in K_q(A,v)$

Proof. By definition of q, there exists a polynomial $p(x) = \sum_{j=0}^{q-1}\alpha_j x^j$ s.t $A^q v - p(A)v = 0$

(Base case) $m=q+1$; $A^q v\in K_q(A,v)$

(Induction) $A^j v\in K_q(A,v)$ for $j = q+1,\dots,m-1$

(Generalize) $A^m v \in K_q(A,v)$. $A^{m-1}v\in K_q(A,v)$

$A^{m-1}v = \sum_{j=0}^{q-1}\beta_j A^j v$

$A^m v = \sum_{j=0}^{q-1}\beta_j A^{j+1}v = \sum_{j=0}^{q-2}\beta_j A^{j+1}v + \beta_{q-1}A^q v$

if $x\in K_q(A,v)$ then $x = \sum_{j=0}^{q-1}\beta_j A^j v$

$Ax = \sum_{j=0}^{q-2}\beta_j A^{j+1}v + \beta_{q-1}A^q v$

Corollary: For any A and v, $\dim(K_m(A,v)) = \min\{m, \text{grade}(A,v)\}$

Proof: Added 20 (c).

Added Question 20 (d).

3.4.1 Find a nicer Basis

If grade$(A,v) = q$ ∃ $\alpha_0,\dots,\alpha_{q-1}$ s.t.

$v = \sum_{j=1}^q \alpha_{j-1}A^j v$

$A^{-1}v = \sum_{j=1}^q\alpha_{j-1}A^{j-1}v$

$A^{-1}b = \sum_{j=1}^q \alpha_{j-1}A^{j-1}b$

$A^{-1}b = [b\ Ab\ \cdots\ A^{q-1}b]\begin{bmatrix}\alpha_0\\ \alpha_1\\ \vdots\\ \alpha_{q-1}\end{bmatrix}$

(1) More material (Nonlinear iterative solvers)

(2) Office Hours

Arnoldi's Method

Given matrix A and vector v

$v_0 = \frac{v}{\|v\|_2}$

for $i=1,\dots,m$

$v_i = \frac{Av_{i-1} - \sum_{j=0}^{i-1}v_j'(Av_{i-1})v_j}{\text{normalize}}$

return $\{v_0,v_1,\dots,v_m\}$

Lemma: Let $A\in\mathbb{R}^{n\times n}$ and $v_0\in\mathbb{R}^n$ be a unit vector. Let $\{v_0,v_1,\dots,v_{k-1}\}$ be the iterates generated by Arnoldi's method by iteration k. Let q = grade(A,v_0). Then

1. If $k \le q$ then $\{v_0,\dots,v_{k-1}\}$ is an orthonormal basis of $K_k(A,v_0)$
2. If $k \ge q$ then $\{v_0,\dots,v_{q-1}\}$ is an orthonormal basis of $K_q(A,v_0)$

Proof $K_q(A,v_0) = K_k(A,v_0)$ for any $k \ge q$ (second)

(1) $\dim(K_k(A,v_0)) = k$

Claim: $\{v_0,v_1,\dots,v_j,Av_j\}$ is lin. ind $j \le q-2$

span$\{v_0,\dots,v_j,Av_j\} = K_{j+1}(A,v_0)$

Base Case: $j=0$

$\{v_0, Av_0\}$ are linearly ind ✓

span$\{v_0,Av_0\} = K_2(A,v_0)$ ✓

Induction hypothesis:

$\{v_0,\dots,v_l,Av_l\}$ is lin ind $\forall l\le j\le q-3$

span$\{v_0,\dots,v_l,Av_l\} = K_{l+2}(A,v_0)$

$v_{l+1}\ne 0$

span$\{v_0,\dots,v_{l+1}\} = K_{l+2}(A,v_0)$ $\forall l \le j \le q-3$

Generalize: $j = q-3$, $j+1$

$v_{j+1} = \frac{Av_j - \sum_{k=0}^{j}v_k'Av_jv_k}{\text{normalize}} \ne 0$

$\{v_0,\dots,v_{j+1},Av_{j+1}\}$ is lin ind.

span$\{v_0,\dots,v_{j+1}\} = K_{j+2}(A,v_0)$

span$\{v_0,\dots,v_{j+1},Av_{j+1}\} = K_{j+3}(A,v_0)$

Av_{j+1} is lin ind of $v_0,\dots,v_{j+1}$

$A^{j+1}v_0 = \sum_{k=0}^{j+1}\alpha_k v_k$ $\quad A^{j+2}v_0 = \sum_{k=0}^{j+1}\alpha_k Av_k$ ($K_{j+3}(A,v_0)$)

$A^{j+2}v_0 = \sum_{k=0}^{j}\alpha_k Av_k + \alpha_{j+1}Av_{j+1}$ ($K_{j+3}(A,v_0)$, $K_{j+3}(A,v_0)$)

$v_1 = \frac{Av_0 - h_{0,0}v_0}{h_{1,0}}$ $\quad Av_0 = h_{0,0}v_0 + h_{1,0}v_1$

$Av_j = \sum_{i=0}^{j}h_{ij}v_i + h_{j+1,j}v_{j+1}$

$v_{j+1} = \frac{Av_j - \sum_{i=0}^j h_{i,j}v_i}{h_{j+1,j}}$ ← $h_{i,j} = v_i'Av_j$, ← $\|Av_j - \sum_{i=0}^j h_{i,j}v_i\|_2$

$[Av_0\ Av_1\ \cdots\ Av_{q-1}] =$

$= [h_{0,0}v_0 + h_{1,0}v_1 \quad h_{0,1}v_0 + h_{1,1}v_1 + h_{2,1}v_2 \quad \cdots \quad \sum_{j=0}^{q-1}h_{j,q-1}v_j + 0]$

$\underbrace{A[v_0 \cdots v_{q-1}]}_{V_q} = \underbrace{[v_0 \cdots v_{q-1}]}_{V_q}\begin{bmatrix}h_{0,0} & h_{0,1} & \cdots & h_{0,q-2} & h_{0,q-1}\\ h_{1,0} & h_{1,1} & \cdots & h_{1,q-2} & h_{1,q-1}\\ 0 & h_{2,1} & \ddots & \vdots & \vdots\\ \vdots & 0 & \ddots & & \\ 0 & 0 & \cdots & h_{q-1,q-2} & h_{q-1,q-1}\end{bmatrix}$ ($q\times q$)

$V_q'V_q = I_q$ Upper Hessenberg Matrix

$A^{-1}r_0 = \sum_{j=0}^{q-1}\alpha_j A^j r_0$

$A^{-1}r_0 = \sum_{j=0}^{q-1}\alpha_j v_j$ $\quad \alpha_j = v_j'A^{-1}r_0$

$r_0 = \sum_{j=0}^{q-1}\alpha_j A^{j+1}r_0$

$r_0 = \sum_{j=0}^{q-1}\alpha_j Av_j = A[v_0 \cdots v_{q-1}]\alpha$

$r_0 = [v_0 \cdots v_{q-1}]H_q\alpha$

$V_q'r_0 = H_q\alpha$ $\quad v_0 = \frac{r_0}{\|r_0\|_2}$

$\|r_0\|_2 e_1 = H_q\alpha$

$\begin{bmatrix}h_{0,0} & h_{0,1} & \cdots\\ h_{1,0} & h_{1,1} & h_{1,2} & \cdots\\ & h_{2,1} & h_{2,2} & \cdots\\ & & h_{3,2} & \ddots\\ & & & \ddots\end{bmatrix}$

See Added Question 21 — Hessenberg matrices

Seed Added Question 22 — Solve a system

1. Given x_0 and $r_0 = Ax_0 - b$
2. Generate an orthogonal basis Krylov space (up to q) using Arnoldi's method
3. Solve the resulting system using Given's rotations + back substitution
4. Recover x^*

3.5. GMRES

$\|Ay - b\|_2^2 = \|A(x_0 - V_q\alpha_q) - b\|_2^2$

$y = x_0 - V_q\alpha_q$

$= \|r_0 - AV_q\alpha_q\|_2^2$

$= \|r_0 - V_qH_q\alpha_q\|_2^2$

$= \|\|r_0\|_2 e_1 - H_q\alpha_q\|_2^2$

4. Row Action Methods

a. Kaczmarz Method

b. Randomized Kaczmarz Method (Strohmer-Vershynin)

c. Generalized Randomized Kaczmarz (GRK)

d. Orthogonalizing GRK

e. Random Sketching

A. Kaczmarz Method

$Ax = b$, $A\in\mathbb{R}^{n\times d}$

$a_i\in\mathbb{R}^d$, $b_i\in\mathbb{R}$ s.t. a_i corresponds to the i^{th} row of A; b_i corresponds to the i^{th} entry of b

Chooses $x_0\in\mathbb{R}^d$ arbitrarily

$x_{k+1} = x_k - \frac{a_{i_k}}{\|a_{i_k}\|_2^2}(a_{i_k}'x_k - b_{i_k})$

where $i_k = \text{rem}(k,n)+1$

B. (Strohmer-Vershynin) Randomized Kaczmarz

i_k are going to be iid r.v.

$P[i_k = j] = \begin{cases}\frac{\|a_j\|_2^2}{\|A\|_F^2} & j=1,\dots,n\\ 0 & \text{otherwise}\end{cases}$

$x_{k+1} = x_k - \frac{a_{i_k}}{\|a_{i_k}\|_2^2}(a_{i_k}'x_k - b_{i_k})$

Theorem: Suppose A is square, invertible matrix and let $x^* = A^{-1}b$. If $\{i_k\}$ are iid according to the above distribution then $E[\|x_k - x^*\|_2^2] \le \left(1 - \frac{\sigma_{min}(A)^2}{\|A\|_F^2}\right)^k\|x_0 - x^*\|_2^2$

Proof

- Observation 1: $x_{k+1} - x_k$ and $x_{k+1} - x^*$ are orthogonal. Problem 32 (a)
- Observation 2: Let i have the same distribution as i_k

$E\left[\frac{(a_i'y)^2}{\|a_i\|_2^2}\right] = \sum_{j=1}^n \frac{\|a_j\|_2^2}{\|A\|_F^2}\cdot\frac{(a_j'y)^2}{\|a_j\|_2^2} = \frac{1}{\|A\|_F^2}\sum_{j=1}^n (a_j'y)^2$

$= \frac{1}{\|A\|_F^2}\|Ay\|_2^2 \ge \frac{\sigma_{min}(A)^2}{\|A\|_F^2}\|y\|_2^2$

First observation

$\|x_k - x^*\|_2^2 = \|x_k - x_{k+1} + x_{k+1} - x^*\|_2^2$

$= \|x_k - x_{k+1}\|_2^2 + \|x_{k+1} - x^*\|_2^2$

$\|x_{k+1} - x^*\|_2^2 = \|x_k - x^*\|_2^2 - \|x_k - x_{k+1}\|_2^2$

$x_{k+1} = x_k - \frac{a_{i_k}}{\|a_{i_k}\|_2^2}(a_{i_k}'x_k - b_{i_k})$ $\quad x^* = A^{-1}b$ $\quad a_{i_k}'x^* = b_{i_k}$

$= x_k - \frac{a_{i_k}}{\|a_{i_k}\|_2^2}(a_{i_k}'x_k - a_{i_k}'x^*)$

$= x_k - \frac{a_{i_k}a_{i_k}'}{\|a_{i_k}\|_2^2}(x_k - x^*)$

$\|x_{k+1} - x^*\|_2^2 = \|x_k - x^*\|_2^2 - \left\|\frac{a_{i_k}a_{i_k}'(x_k-x^*)}{\|a_{i_k}\|_2^2}\right\|_2^2$

$= \|x_k - x^*\|_2^2 - \frac{(x_k-x^*)'a_{i_k}a_{i_k}'a_{i_k}a_{i_k}'(x_k-x^*)}{\|a_{i_k}\|_2^4}$

$= \|x_k - x^*\|_2^2 - \frac{[a_{i_k}'(x_k-x^*)]^2}{\|a_{i_k}\|_2^2}$

$E[\|x_{k+1}-x^*\|_2^2] = E[\|x_k-x^*\|_2^2] - E\left[\frac{[a_{i_k}'(x_k-x^*)]^2}{\|a_{i_k}\|_2^2}\right]$

$\le E[\|x_k-x^*\|_2^2] - \frac{\sigma_{min}(A)^2}{\|A\|_F^2}E[\|x_k-x^*\|_2^2]$

$\le \left(1 - \frac{\sigma_{min}(A)^2}{\|A\|_F^2}\right)E[\|x_k - x^*\|_2^2]$

C. Generalized Random Kaczmarz

(1) Let $\{w_k\}$ be a collection of random vectors in $\mathbb{R}^n$

$x_{k+1} = x_k - \frac{A'w_k}{\|A'w_k\|_2^2}(w_k'Ax_k - w_k'b)$

(2) Let $\{w_k\}$ have arbitrary relationships

Lemma (Meany 1969) Let $f_1,\dots,f_k\in\mathbb{R}^n$ with $k \le n$.

Write $f_k = f^S + f^N$ where f^S belongs to the space $S = \text{span}\{f_1,\dots,f_{k-1}\}$ and f^N is orthogonal to S.

Let $\bar F = [f_1 \cdots f_{k-1}]$

$F = [f_1 \cdots f_{k-1}\ f_k]$

Then $\det(F'F) = \|f^N\|_2^2\det(\bar F'\bar F)$

Proof: Added Question 28

Theorem Let $v_1,\dots,v_k$ be unit vectors in $\mathbb{R}^n$ for some $k\in\mathbb{N}$

Let $S = \text{span}\{v_1,\dots,v_k\}$

Let $\mathcal{F}$ be all matrices F whose columns $\{f_1,\dots,f_r\}\subseteq\{v_1,\dots,v_k\}$ that are maximally linearly independent.

Then $\sup_{y\in S,\|y\|_2=1}\|Qy\|_2^2 \le 1 - \min_{F\in\mathcal{F}}\det(F'F)$

where $Q = (I - v_kv_k')(I - v_{k-1}v_{k-1}')\cdots(I - v_1v_1')$

Proof

k=1. Both sides are zero.

k=j-1. Induction

k=j. Generalize

Notation: $\bar S = \text{span}\{v_1,\dots,v_{j-1}\}$, $\{f_1,\dots,f_{\bar r}\}$ be a maximally lin. ind subset of $\{v_1,\dots,v_{j-1}\}$. $\bar F = [f_1 \cdots f_{\bar r}]$ and

$\bar Q = (I - v_{j-1}v_{j-1}')(I - v_{j-2}v_{j-2}')\cdots(I - v_1v_1')$

For any unit vector $y\in S$, $y = y^{\bar S} + y^N$ where $y^{\bar S}$ is the component of $y^{\bar S}\in\bar S$, $y^N\perp\bar S$

$\|z\|_2 = \|\bar Q y^{\bar S}\|_2 \le \|y^{\bar S}\|_2\sqrt{1 - \det(\bar F'\bar F)}$

$v_j = v^{\bar S} + v^N$

Case A: Suppose $S = \bar S$

- $v_1,\dots,v_j$ are colinear $\quad Q = (I - v_1v_1')$ ✓
- $\dim(\bar S) > 1$ $\quad \{v_1,v_2,v_3\}$ $\quad v_4\in\text{span}\{v_1,\dots,v_3\}$; $\{v_1,v_4,v_3\}$ $\{v_2\}$ lin ind

$\|I - v_jv_j'\|_2 \le 1$ $\quad \|Qy\|_2 \le \|\bar Qy\|_2$

Case B: Suppose $S \supsetneq \bar S$

$\|Qy\|_2^2 = \|(I - v_jv_j')\bar Qy\|_2^2 = \|(I - v_jv_j')(\bar Qy^{\bar S} + \bar Qy^N)\|_2^2$

$= \|(I - v_jv_j')(z + y^N)\|_2^2$

$= (z + y^N)'(I - v_jv_j')(z + y^N)$

$= (z+y^N)'(I - (v^{\bar S}+v^N)(v^{\bar S}+v^N)')(z+y^N)$

$= \|z\|_2^2 + \|y^N\|_2^2 + 2z'y^N$ $(=0)$ $- (z'v^{\bar S})^2 - 2z'v^{\bar S}(v^N)'y^N - ((v^N)'y^N)^2$

$\le \|z\|_2^2 + \|y^N\|_2^2 - (z'v^{\bar S})^2 - 2z'v^{\bar S}\|v^N\|_2\|y^N\|_2 - \|v^N\|_2^2\|y^N\|_2^2$

$\|Qy\|_2^2 \le \|z\|_2^2 + \|y^N\|_2^2 - (|z'v^{\bar S}| - \|v^N\|_2\|y^N\|_2)^2$

Situation 1: $\|v^N\|_2 \le \|y^{\bar S}\|_2$

$\|Qy\|_2^2 \le \|z\|_2^2 + \|y^N\|_2^2$

$\le \|y^{\bar S}\|_2^2(1 - \det(\bar F'\bar F)) + \|y^N\|_2^2$

$= \|y\|_2^2 - \|y^{\bar S}\|_2^2\det(\bar F'\bar F)$

$\le 1 - \|v^N\|_2^2\det(\bar F'\bar F)$

$= 1 - \det(F'F)$

Situation 2: $\|v^N\|_2 > \|y^{\bar S}\|_2$. $\|v_j\|_2 = \|y\|_2 = 1$, $\|v^{\bar S}\|_2 \le \|y^N\|_2$

$\|y^N\|_2\|v^N\|_2 \ge \|v^{\bar S}\|_2\|y^{\bar S}\|_2 \ge \|v^{\bar S}\|_2\|z\|_2 \ge |z'v^{\bar S}|$

$\|y^N\|_2\|v^N\|_2 - |z'v^{\bar S}| \ge \|y^N\|_2\|v^N\|_2 - \|z\|_2\|v^{\bar S}\|_2 \ge 0$

$\|Qy\|_2^2 \le \|z\|_2^2 + \|y^N\|_2^2 - (\|y^N\|_2\|v^N\|_2 - \|z\|_2\|v^{\bar S}\|_2)^2$

$\le \|z\|_2^2 + \|y^N\|_2^2 - \|y^N\|_2^2\|v^N\|_2^2 + 2\|z\|_2\|v^{\bar S}\|_2\|y^N\|_2\|v^N\|_2 - \|z\|_2^2\|v^{\bar S}\|_2^2$

$\le \|z\|_2^2\|v^N\|_2^2 + \|y^N\|_2^2\|v^{\bar S}\|_2^2 + 2\|z\|_2\|v^{\bar S}\|_2\|y^N\|_2\|v^N\|_2$

$= (\|z\|_2\|v^N\|_2 + \|y^N\|_2\|v^{\bar S}\|_2)^2$

$\le (\sqrt{1 - \det(\bar F'\bar F)}\|v^N\|_2\|y^{\bar S}\|_2 + \|y^N\|_2\|v^{\bar S}\|_2)^2$

$\le (\|y^{\bar S}\|_2^2 + \|y^N\|_2^2)(\|v^N\|_2^2(1 - \det(\bar F'\bar F)) + \|v^{\bar S}\|_2^2)$

$= 1 - \|v^N\|_2^2\det(\bar F'\bar F)$

$= 1 - \det(F'F)$

$\begin{bmatrix}\|y^{\bar S}\|_2\\ \|y^N\|_2\end{bmatrix}\begin{bmatrix}\|v^N\|_2\sqrt{1-\det(\bar F'\bar F)}\\ \|v^{\bar S}\|_2\end{bmatrix}$

If $Ax=b$ admits a solution x^*. Then GRK iteration

$x_{k+1} - x^* = x_k - x^* - \frac{A'w_k}{\|A'w_k\|_2^2}(w_k'Ax_k - w_k'b)$ $\quad (b = Ax^*)$

$= \left(I - \frac{A'w_kw_k'A}{\|A'w_k\|_2^2}\right)(x_k - x^*)$

$= \left(I - \frac{A'w_kw_k'A}{\|A'w_k\|_2^2}\right)\cdots\left(I - \frac{A'w_0w_0'A}{\|A'w_0\|_2^2}\right)(x_0 - x^*)$

Let $w\in\mathbb{R}^d$ be a r.v.

1. A "null space induced by w"

$N_A(w) = \{z\in\mathbb{R}^d : P[z'A'w = 0] = 1\}$

$\ker(A)\subseteq N_A(w)$

2. $V_A(w)\perp\ker(A)$ and $N_A(w) = V_A(w) + \ker(A)$
3. $R_A(w)\perp N_A(w)$

$R_A(w)$ is the smallest subspace of $\mathbb{R}^d$ s.t. $P[A'w\in R_A(w)] = 1$

$P[A'w_k\in R_A(w)] = 1$

(1) $\tau_0 = -1$

$\tau_{k+1} = \min\{k > \tau_k : \text{span}\{A'w[\tau_k+1],\dots,A'w[k]\} = R_A(w)\}$

(2) Let $\mathcal{F}_k$ be all matrices F whose columns are maximal linearly ind. subsets of

$\left\{\frac{A'w[\tau_{k-1}+1]}{\|\ \|_2},\dots,\frac{A'w[\tau_k]}{\|\ \|_2}\right\}$

(3) $\gamma_k = 1 - \min_{F\in\mathcal{F}_k}\det(F'F)$

Theorem: Suppose $\{w_k\}$ are defined s.t. $P[A'w_k\in R_A(w)] = 1$.

For any k let $L(k) = \max\{l : k > \tau_l\}$

$\tau_1 \quad \tau_2 \quad \tau_3 \quad \tau_4 \quad k \quad \tau_5$

Then the iterates generated by GRK satisfy

$\|x_k - x^* - P_{N_A(w)}(x_0 - x^*)\|_2^2 \le \left(\prod_{j=1}^{L(k)}\gamma_j\right)\|P_{R_A(w)}(x_0 - x^*)\|_2^2$

Moreover, on the event $\left(\bigcap_{l=0}^{\infty}\{\tau_l < \infty\}\right)\cap\{\prod_{j=1}^{l}\gamma_j\to0\}$

$\lim_{k\to\infty}Ax_k = b$ iff $P_{V_A(w)}(x_0 - x^*) = 0$

Proof Whenever $\tau_l < \infty$

$Q_l = \left(I - \frac{A'w_{\tau_l}w_{\tau_l}'A}{\|A'w_{\tau_l}\|_2^2}\right)\cdots\left(I - \frac{A'w_{\tau_{l-1}+1}w_{\tau_{l-1}+1}'A}{\|A'w_{\tau_{l-1}+1}\|_2^2}\right)$

$x_{\tau_l+1} - x^* = Q_l(x_{\tau_{l-1}+1} - x^*)$

$P_{R_A(w)}$ projection onto $R_A(w)$

$P_{N_A(w)}$ projection onto $N_A(w)$

$P_{V_A(w)}$ projection onto $V_A(w)$

$x_{\tau_l+1} - x^* = Q_l(P_{R_A}(x_{\tau_{l-1}+1} - x^*) + P_{N_A(w)}(x_{\tau_{l-1}+1} - x^*))$

$= Q_lP_{R_A}(x_{\tau_{l-1}+1} - x^*) + P_{N_A(w)}(x_{\tau_{l-1}+1} - x^*)$

✓ $P_{R_A(w)}(x_{\tau_l+1} - x^*) = Q_lP_{R_A(w)}(x_{\tau_{l-1}+1} - x^*)$

✓ $P_{N_A(w)}(x_{\tau_l+1} - x^*) = P_{N_A(w)}(x_{\tau_{l-1}+1} - x^*) = P_{N_A(w)}(x_0 - x^*)$

$\{\tau_1 < \infty\}$

$P_{R_A(w)}(x_0 - x^*)\in R_A(w)$

$\|P_{R_A(w)}(x_{\tau_1+1} - x^*)\|_2^2 = \|Q_1P_{R_A(w)}(x_0 - x^*)\|_2^2 \le \gamma_1\|P_{R_A(w)}(x_0 - x^*)\|_2^2$

$\{\tau_l < \infty\}$

$\|P_{R_A(w)}(x_{\tau_l+1} - x^*)\|_2^2 \le \left(\prod_{j=1}^{l}\gamma_j\right)\|P_{R_A(w)}(x_0 - x^*)\|_2^2$

$\|x_{\tau_l+1} - x^* - P_{N_A(w)}(x_0 - x^*)\|_2^2 \le \left(\prod_{j=1}^{l}\gamma_j\right)\|P_{R_A(w)}(x_0 - x^*)\|_2^2$

$L(k) = l$, $\quad \|I - vv'\|_2 \le 1$

$\lim_{k\to\infty} x_k - x^* = P_{N_A(w)}(x_0 - x^*) = P_{\ker(A)}(x_0 - x^*) + P_{V_A(w)}(x_0 - x^*)$

$\lim_{k\to\infty} Ax_k - b = 0$ iff $AP_{\ker(A)}(x_0 - x^*) + AP_{V_A(w)}(x_0 - x^*) = AP_{V_A(w)}(x_0 - x^*) = 0$

$AP_{V_A(w)}(x_0 - x^*) = 0$ iff $P_{V_A(w)}(x_0 - x^*)\in\ker(A)$ ⟺ $P_{V_A(w)}(x_0 - x^*) = 0$

$\lim Ax_k = b \iff P_{V_A(w)}(x_0 - x^*) = 0$

4.4 Orthogonalize GRK

1. $x_1 = x_0 - \frac{A'w_0}{\|A'w_0\|_2^2}(w_0'Ax_0 - w_0'b)$, $u_0 = \frac{A'w_0}{\|A'w_0\|_2}$ and $J=0$
2. $k \ge 1$

(a) Compute $A'w_k$ and then

$v_k = A'w_k - \sum_{j=0}^{J}(w_k'Au_j)u_j = (I - u_Ju_J')\cdots(I - u_0u_0')A'w_k$

(b) If $v_k\ne0$ then $u_{J+1} = v_k/\|v_k\|_2$

$x_{k+1} = x_k - \frac{v_k}{v_k'A'w_k}(w_k'Ax_k - w_k'b)$

(c) If $v_k \ne 0$, $J \leftarrow J+1$

$x_1 - x^* = x_0 - x^* - \frac{A'w_0}{\|A'w_0\|_2^2}(w_0'Ax_0 - w_0'b)$

$= x_0 - x^* - \frac{A'w_0w_0'A}{\|A'w_0\|_2^2}(x_0 - x^*)$

$= \left(I - \frac{A'w_0w_0'A}{\|A'w_0\|_2^2}\right)(x_0 - x^*)$

$= (I - u_0u_0')(x_0 - x^*)$

→ $A'w_1$ is colinear with u_0

$v_1 = A'w_1 - (u_0'A'w_1)u_0 = 0$

$x_2 = x_1$

→ $A'w_1$ is not colinear with u_0

$v_1 = A'w_1 - (u_0'A'w_1)u_0 \neq 0$

$u_1 = \frac{v_1}{\|v_1\|_2}$, $u_1\perp u_0$

$x_2 - x^* = x_1 - x^* - \frac{v_1}{v_1'A'w_1}(w_1'Ax_1 - w_1'b)$

$= x_1 - x^* - \frac{v_1}{v_1'A'w_1}(w_1'A x_1 - w_1'Ax^*)$

$= \left(I - \frac{v_1w_1'A}{v_1'A'w_1}\right)(x_1 - x^*)$

$= \left(I - \frac{v_1w_1'A}{v_1'A'w_1}\right)(I - u_0u_0')(x_0 - x^*)$

$= \left(I - \frac{(I-u_0u_0')A'w_1w_1'A}{w_1'A(I-u_0u_0')A'w_1}\right)(I - u_0u_0')(x_0 - x^*)$

$= \left((I - u_0u_0') - \frac{(I-u_0u_0')A'w_1w_1'A(I-u_0u_0')}{w_1'A(I-u_0u_0')(I-u_0u_0')A'w_1}\right)(x_0 - x^*)$

$= (I - u_0u_0' - u_1u_1')(x_0 - x^*)$

Suppose at iteration k we have J_k lin ind search directions

$x_{k+1} - x^* = \left(I - \sum_{j=0}^{J_k}u_ju_j'\right)(x_0 - x^*)$

$= (I - u_Ju_J')\cdots(I - u_0u_0')(x_0 - x^*)$

$k = \tau_1$ $\quad J_k = \dim(R_A(w))$

$x_{k+1} - x^* = (I - u_Ju_J')\cdots(I - u_0u_0')(P_{N_A(w)}(x_0 - x^*) + P_{R_A(w)}(x_0 - x^*))$

$= (I - u_Ju_J')\cdots(I - u_0u_0')P_{R_A(w)}(x_0 - x^*) + P_{N_A(w)}(x_0 - x^*)$

$= \underbrace{P_{N_A(w)}P_{R_A(w)}(x_0 - x^*)}_{0} + P_{N_A(w)}(x_0 - x^*)$

$= P_{N_A(w)}(x_0 - x^*)$

$0 = Ax_{k+1} - b = AP_{N_A(w)}(x_0 - x^*) \iff P_{V_A(w)}(x_0 - x^*) = 0$

$\begin{bmatrix}w_0'A\\ w_1'A\\ w_2'A\\ \vdots\\ w_k'A\end{bmatrix}x = \begin{bmatrix}w_0'b\\ w_1'b\\ \vdots\\ w_k'b\end{bmatrix}$ $\quad (WA)x = (Wb)$