Homework Numerical Method in Data Science

Homework

Lemma 1: Let $A \in R^{mxn}$ matrix, then:

1) if $x \in N(A^TA)$ then $Ax \in R(A)$ and $Ax \in N(A^T)$

2) $N(A^TA) = N(A)$

3) $rank(A) = rank(A^TA)$

4) if A has linear independent columns then A^TA is non singular.

Proof

1. As $R(A) = \{Ax|x \in R^n\}$. So $Ax \in R(A)$

On the other hand, we have:

$$\begin{aligned} & x \in N(A^TA) \\ \Rightarrow \quad & A^Ax = 0 \\ \Rightarrow \quad & Ax \in N(A^T) \end{aligned}$$

2. By Definition, $N(A) \subseteq N(A^TA)$. Let's suppose: $x \in N(A^TA)$.

Then,

$$\begin{aligned} & x^TA^TAx = 0 \\ \Rightarrow \quad & (Ax)^T(Ax) = 0 \\ = \quad & Ax = 0 \end{aligned}$$

That means, $N(A) = N(A^TA)$ 3. By rank - nullity theorem, if $B \in R^{mxn}$, then $n = dimR(B) + dimN(B)$.

Apply it to A and A^TA.

$$n = dimR(A) + dimN(A)$$

$$n = dimR(A^TA) + dimN(A^TA)$$

As $N(A) = N(A^TA)$, then $dimN(A) = dimN(A^TA)$. So, $dimR(A) = dimR(A^TA)$.

Thus, $rank(A) = rank(A^TA)$

Theorem 1:

If $A \in R^{mxn}$, then $R(A)$ is a subspace of R^m.

- Denote $R(A)$ is the range of A
- Definition: $R(A) = \{Y : \exists X \in R^n | AX = Y\}$

Proof

Notice that if Y is in $R(A)$, then $Y = AX$ for some $X \in R^n$. Y will be in R^m. This shows that the set $R(A) \subseteq R^m$.

Let $W = R(A)$. a) Let 0_m denote the zero vectorr in R^m and 0_n denote the zero vector in R^n. Notice that $A.0_n = 0_m$. Hence, $AX = 0_m$ is satisfied by at least one $X \in R^n$. Thus O_m is indeed in W. Hence a) is valid for W.

b) Suppose that Y_1 and Y_2 are in W. Each of the matrix equations $AX = Y_1$ and $AX = Y_2$ has at least one solution. Suppose that $X = X_1 \in R^n$ satisfying first equation. That is $AX_1 = Y_1$. As thee same for $X = X_2 \in R^n$. Consider the matrix equation, $AX = Y_1 + Y_2$. Let $X = X_1 + X_2 \in R^n$. We obtain,

$$AX = A(X_1 + X_2) = AX_1 + AX_2 = Y_1 + Y_2 \in R^m$$

This shows that W is closed under addition and b) is valid for W.

c) Suppose that $Y_1 \in W$. Let c by any scalar. Since $Y_1 \in W$, $\exists X \in R^m$ such that $AX_1 = Y_1$. Now consider the matrix equation $AX = cY_1$. Let $X = cX_1$, a vector in R^n, we have:

$$AX = A(cX_1) = c(AX) = cY_1$$

Therefore, $cY_1 \in W$. Therefore, $Y_1 + Y_2 \in W$. This shows that W is closed under scalar multiplication and c) is valid or W.

WE have proved that $W = R(A) \subseteq R^m$ satisfying the 3 subspace requirements. Hence, $R(A) \subseteq R^m$.

Theorem 2

If $A \in R^{mxn}$. Then $N(A)$ is a subspace of R^n.

- Denote $N(A)$ is the null space of A
- Def: $N(A) = \{X : AX = 0_m\}$

Proof

If $X \in N(A)$. Then $AX = 0_m$. Since A is m x n and AX is m x 1, it follows that X must be n x1. That is, X is in R^n. Therefore, $N(A)$ is a subset of R^n. Let $W = N(A)$.

a) Notice that $A0_n = 0_m$. Hence, $AX = 0_m$ is satisfied by $X = 0_n$. It means that 0_n is indeed in W.

b) Suppose that X_1, X_2 are in W. Let $X = X_1 + X_2$. Then,

$$AX = A(X_1 + X_2) = AX_1 + AX_2 = 0_m + 0_m = 0_m$$

Therefore, $X = X_1 + X_2$ is in W. This shows that W is closed under addition and b) is valid for W.

c) Suppose that X_1 is in W. Let c be any scalar. Since $X_1 \in W$, we have: $AX_1 = 0_m$. Let $X = cX_1$. Then,

$$AX = A(cX_1) = c(AX_1) = c0_m = 0_m$$

Therefore, $X_c X_1 \in W$. This shows that W is closed under scalar multiplication and c) is valid for W.

We have proved that $W = N(A)$ is a subset of R^n satisfying the 3 subspace requirements. Hence $N(A) \subseteq R^n$.

Theorem: $Ker(A)^{\perp} = Im(A^T)$.

Take arbitrary $x \in Ker(A)$, $\Rightarrow Ax = 0$.

$y \in im(A^T)$, $\Rightarrow \exists z \in V$ such that $A^T z = y$.

We have:

$$\begin{aligned} yox &= x^T y \\ &= x^T (A^T z) \\ &= (Ax)^T z \\ &= 0 \end{aligned}$$

$\Rightarrow y \perp x$.

Thus, $y \in ker(A)^{\perp}$.

Therefore, $Im(A^T) \subseteq ker(A)^{\perp}$.

On the other hand, Take $x \in Ker(A), \Rightarrow Ax = 0$.

Thus, $0 = yoAx = (Ax)^T y = x^T A^T y = (A^T y)ox$.

Therefore, $x \in Im(A^T)^{\perp}$.

Then, $ker(A) \subseteq Im(A^T)^{\perp}$.

So, $ker(A)^{\perp} \subseteq Im(A^T)^{\perp\perp} = Im(A^T)$

In conclusion, $Ker(A)^{\perp} = Im(A^T)$

Application of Least Square Method

The method of least squares is a standard approach in regression analysis to approximate the solution of overdetermined systems (sets of equations in which there are more equations than unknowns) by minimizing the sum of the squares of the residuals made in the results of every single equation.

Least square problems fall into two categories: linear or ordinary least squares and non-linear least squares, depending on whether or not the residuals are linear in all unknowns. The linear least-squares problem occurs in statistical regression analysis, it has a closed-form solution. The nonlinear problem is usually solved by iterative refinement, at each iteration the system is approximated by a linear one and thus the core calculation is similar in both cases.

There are a variety of applications that least squares can be applied to. Let's discuss some of them.

Least Square Method for Time Series Analysis

Time series analysis is one of the most important analytical tools in the experimental sciences. As such, many procedures for finding periodicities in data have been developed.

The Linear Least Squares Spectral Analysis (LLSSA) method is an extension of the curve-fitting approach developed in Bloomfield.

For environmental science, where signals generally have more than one periodicity but relatively few data points, this method may be the best available. Many procedures exist for determining periodicities in these kinds of data, but this method has several compelling features worth noting:

- Unevenly spaced data are analyzed as easily as evenly spaced data.
- Any frequency may be considered.
- The statistical significance of every peak is easily computed.

In addition to these benefits, all output is easily understood and validated.

The Vostok ice core is a 2,083-meter-long core drilled in East Antarctica. Air bubbles from this core have been analyzed to recreate the temperature, CO2, and CH4 levels over the past 160,000 years.

We have applied the LLSSA method to the methane data set. The data and the least squares spectrum fit are plotted in Figure 1. The least squares model of the data are composed of the 10 most important periodicities.

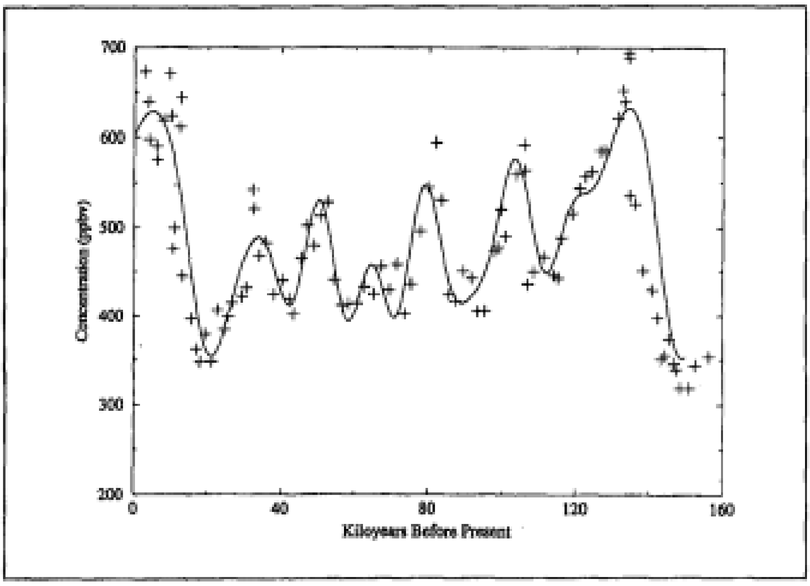

Figure 1. The Vostok ice core methane database (plus signs) with the top 10 frequency fit based upon analysis with the LLSSA method (solid line). The frequencies are 0.001, 0.008, 0.016, 0.024, 0.032, 0.040, 0.047, 0.055, 0.062, and 0.071 (all in inverse kiloyears).

Least Square Method to forecast sales for a company

Some companies and organization need to forecast their future sales using the previously obtained data so as to have a proper plan that can lead them in accomplishing their set goals and target. Also, lack of forecast data has led to the collapse of several businesses in the past.

Table 1: Pharmaceutical Sales

Year	Sales ($)
2014	10000
2015	21000
2016	50000
2017	70000
2018	71000
2019	?
2020	?
2021	?

Years	Sales	X	X^2	XY	Trend value
2014	10000	-2	4	-20000	**10200**
2015	21000	-1	1	-21000	**27300**
2016	50000	0	0	0	**44400**
2017	70000	1	1	70000	**61500**
2018	71000	2	4	142000	**78600**
2019	?	3	9		**95700**
2020	?	4	16		**112800**
2021	?	5	25		**129900**

N= 5

$\sum X = 0$

$\sum Y = 222000 \ i.e \ (sum \ of \ sales)$

$\sum X^2 = 10$

$\sum XY = 171000$

For y = a + bx

$\sum Y = Na + b \times \sum x$

When 222000 = 5a+ b $\times 0$ we found a = 44400

Also, $\sum XY = a\sum X + b\sum X^2$

When 171000 = (222000$\times 0$) + (b$\times 10$)

171000= 10b we found b = 17100

Therefore, for trend value, Y = a + bx

Y = 44400 + 17100 X

Trends values can be calculated for each year.

For Y 2014, where X = -2. We have:

Y 2014 = 44400 + 17100 (-2)

Y 2014 = 10200

For Y 2015, where X = -1. We have:

Y 2014 = 44400 + 17100 (-1)

Y 2014 = 27300

For Y 2016, where X = 0. We have:

Y 2016 = 44400 + 17100 (0)

Y 2016 = 44400

For Y 2017, where X = 1. We have:

Y 2017 = 44400 + 17100 (1)

Y 2017 = 61500

For Y 2018, where X = 2. We have:

Y 2018 = 44400 + 17100 (2)

Y 2018 = 78600

For Y 2019, where X = 3. We have:

Y 2019 = 44400 + 17100 (3)

Y 2019 = 95700

For Y 2020, where X = 4. We have:

Y 2020 = 44400 + 17100 (4)

Y 2020 = 112800

For Y 2021, where X = 5. We have:

Y 2021 = 44400 + 17100 (5)

Y 2021 = 129900

Least Square Method for the interpretation of data from seismic surveys

During large-scale seismic surveys it is often impossible to arrange shot points and seismometers in a simple pattern, so that the data cannot be treated as simply as those of small-scale prospecting arrays. It is shown that the problem of reducing seismic observations from m shot points and n seismometers (where there is no simple pattern of arranging these) is equivalent to solving (m + n) normal equations with (m + n) unknowns. These normal equations are linear, the matrix of their coefficients is symmetric. When all the shots have been observed at all the seismometers, the solution can even be given generally. Otherwise, a certain amount of computation is necessary.

Least Square Monte Carlo Method in American Put Option Pricing

American options, compared to European counterparts, have an additional flexibility of exercising at anytime before maturity, which adds to complexity in pricing. Monte Carlo simulation is a rather flexible valuation approach which is applicable to almost any feature a financial product can exhibit: American and Bermudan exercise, Asian and lookback features (i.e. path dependency). For pricing American options, nested Monte Carlo can be used, but is very computationally expensive. In 2001, Longstaff-Schwartz proposed least-squares method (LSM) in Monte Carlo which uses least squares to estimate the conditional expected payoff to the option holder from continuation. Although this reduced the computational time drastically, it introduces approximation error from the least squares regression.

Least Squares Methods for Elliptic Systems

For various reasons, it is of interest to extend the theory of least squares methods to include elliptic systems. First, if a second-order elliptic equation is written as a first-order system, it would seem (and this is borne out by our analysis) that the smoothness requirements for the spaces of approximating functions would be reduced, thus eliminating one of the disadvantages of the method. A second motivation for extending the least squares method to elliptic systems is that elliptic systems occur frequently in applications. An example of an elliptic system is the system of equations for Stokes flow. For this system, the least squares method does not require the space of approximating vector fields to be incompressible. Instead, the incompressibility condition is considered as one of the equations in the system, and the analysis provides, in a natural way, weights to put on the residual in the incompressibility equation. The difficulties associated with finding approximating spaces of incompressible vector fields are well-known; the least squares method provides an alternate way of treating these difficulties. Finally, it is desirable to extend the least squares method to elliptic systems to close the gap in the theory of the method.

Signal Denoising using Least Square Methods

Any unwanted signal is defined as noise. In real world, signals are often affected by device-specific noise. Therefore, signal denoising is a challenging task for researchers. The noise is caused due to several reasons such as electrical fluctuations in devices and electromagnetic interference. Noise adds unwanted information to the signal which leads to distorted signal. To overcome this, we use different denoising techniques in signal processing. The main objective in signal denoising is to remove maximum noise to get a clean signal. Denoising is the kernel of signal processing. Removing the noise and retaining details of a signal is the key goal of signal denoising techniques.

II. MATHEMATICAL BACKGROUND

The mathematics behind the concept of least square based approach for signal deconvolution and denoising is discussed in this section.

A. Deconvolution

The problem of determining the input to a Linear Time Invariant (LTI) system when the output signal is known, is termed as Deconvolution. Let $y\ (n)$ be the output signal which is given by,

$$y(n) = h(0)x(n) + h(1)x(n-1) + \ldots\ldots + h(N)x(n-N)$$

where $x(n)$ is the input signal and $h(n)$ is the impulse response. $y(n)$ can be written in terms of $y = Hx$, where H is given by the matrix,

$$\mathbf{H}=\begin{bmatrix} h(0) & & & \\ h(1) & h(0) & & \\ h(2) & h(1) & h(0) & \\ \vdots & & & \ddots \end{bmatrix}$$

The H matrices are constant valued along the diagonal and are called Toeplitz matrices. The input signal x should atleast approximately satisfy $y = Hx$. The problem formulation for the same is given by

$$\min_x \|y - \mathbf{H}x\|_2^2 + \lambda \|\mathbf{D}x\|_2^2 \qquad (1)$$

where D is represented by,

$$D = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & & \ddots & \\ & & & 1 & -2 & 1 \end{bmatrix}$$

The solution for signal denoising by minimizing Eq.(1) is given by,

$$x = (H^T H + \lambda D^T D)^{-1} H^T y \qquad (2)$$

B. Smoothing

In this method, the idea is to obtain a denoised signal closer to noisy input by least squares weighted regularization approach [10]. As smoothness of the signal increases, the energy of its derivative becomes smaller. In other terms, it is interpreted as if x is smooth, $\|Dx\|_2^2$ becomes small.

Let $y(n)$ be the noisy input signal and $x(n)$ be the output that approximates $y(n)$, then the problem formulation is given by,

$$\min_x \|y - x\|_2^2 + \lambda \|\mathbf{D}x\|_2^2 \qquad (3)$$

where, Dx is the second order differentiation of $x(n)$. λ is a parameter on which x depends, for smoothening of a noisy signal. The signal x get smoother as the value of λ increases. The mathematical equation for signal denoising using least squares weighted regularization in the method of smoothing is given by,

$$x = (I + \lambda D^T D)^{-1} y \qquad (4)$$

where I is the identity matrix of same size as that of D.

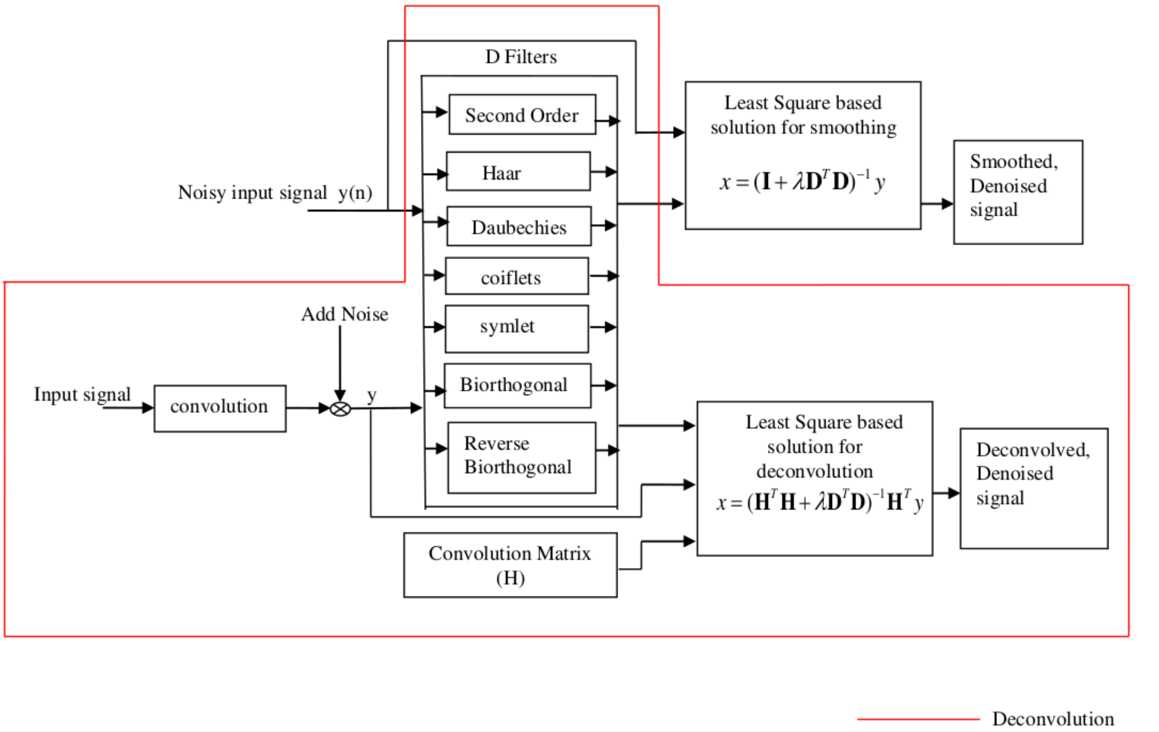

Image Deblurring using Least Square Method

Image deblurring

$$y = Ax + v$$

- x is unknown image, y is observed image
- A is (known) blurring matrix, v is (unknown) noise
- images are $M \times N$, stored as MN-vectors

blurred, noisy image y

deblurred image $\hat{x}$

Least squares deblurring

$$\text{minimize} \quad \|Ax - y\|^2 + \lambda(\|D_{\mathrm{v}}x\|^2 + \|D_{\mathrm{h}}x\|^2)$$

- 1st term is '*data fidelity*' term: ensures $A\hat{x} \approx y$
- 2nd term penalizes differences between values at neighboring pixels

$$\begin{aligned} &\|D_{\mathrm{h}}x\|^2 + \|D_{\mathrm{v}}x\|^2 \\ &\quad = \sum_{i=1}^{M}\sum_{j=1}^{N-1}(X_{ij} - X_{i,j+1})^2 + \sum_{i=1}^{M-1}\sum_{j=1}^{N}(X_{ij} - X_{i+1,j})^2 \end{aligned}$$

if X is the $M \times N$ image stored in the MN-vector x