

Linear Systems: Least Squares Assignment

1. The least square approximate solution to problem $\mathbf{Ax} = \mathbf{b}$, $\mathbf{A} \in \mathbb{R}^{m \times n}$ is obtained by minimizing the following objective function,

$$O(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2 = \sum_{i=1}^m (\tilde{\mathbf{a}}_i^T \mathbf{x} - \mathbf{b}_i)^2$$

If the the objective function was defined differently, where the different components were given different weights for the different terms,

$$O_w(\mathbf{x}) = \sum_{i=1}^m w_i (\tilde{\mathbf{a}}_i^T \mathbf{x} - \mathbf{b}_i)^2$$

This is the *weighted least squares*. Find the expression for the approximate solution of the weighted least squares problem.

2. Consider a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with independent columns. There is no matrix \mathbf{X} , such that $\mathbf{AX} = \mathbf{I}$. Find an expression for the matrix \mathbf{X} , such that $\|\mathbf{AX} - \mathbf{I}\|^2$ is minimum. Hint: Consider the individual columns of \mathbf{X} and \mathbf{I} .
3. Consider the following polynomial equation,

$$y = \sum_{i=0}^n \beta_i x^i, \quad x, y, \beta_i \in \mathbb{R}$$

This expression is linear in the polynomial coefficients. Fitting a polynomial to data can be done through a linear least square procedure. Consider a set of measurements $\{(x_l, y_l)\}_{l=1}^m$. We are interested in fitting a polynomial that fits this data, such that the difference between the polynomial and data is as low as possible.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \dots & x_m^n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$$

\mathbf{X} is called the *Vandermonde* matrix. To estimate $\boldsymbol{\beta}$ through a least squares procedure, \mathbf{X} must have independent columns. Prove that for a set of different x_i s the *Vandermonde* matrix has independent columns.

You are provided with CSV format data file (polyfit.csv) containing data $\{(x_l, y_l)\}_{l=1}^m$, to which you are required to fit a polynomial function. The choice of the order of the polynomial is up to you. This can be come from *a priori* knowledge of the system from which data is collected, or it can be decided based on eyeballing the scatter plot between x_i and y_i .

- (a) **Data fitting error:** Fit polynomials of orders 0 to 10 to the given data, and for each order determine the data fitting error.

$$e_n = \|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y}\|$$

Plot e_n versus the polynomial model order n . You should observe the fitting error to monotonically decrease as a function of n . Why is this? Can e_n ever be zero?

- (b) **Model validation:** Even though increasing the polynomial order decreases the data fitting error, this is not always desirable. Given that noise is ubiquitous in all measurements, increasing model order will result in a polynomial that not only fits the general trend in the data, but also the observed measurement noise. Thus, the optimal choice for the model order is determined through a validation procedure, where the data is split into two sets - *training* set and a *testing* set. In order to understand this, you are required to:
 - i. Split your data D into two sets of size 80% and 20%, corresponding to the *training* D_{train} and *testing* D_{test} sets; these percentages are arbitrary. Split you data randomly, such that each data entry is randomly assigned to D_{train} and D_{test} .
 - ii. Fit the polynomial model of a particular order to D_{train} . Let the model parameters obtained be $\hat{\boldsymbol{\beta}}$. The validation error for this model is defined as the following.

$$e_n^{val} = \|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y}\|$$

where, \mathbf{X} and \mathbf{y} comes from the test data set D_{test} . Estimate the validation error for different models order 0 to 10, and plot e_n^{val} versus the polynomial model order n . How is this plot different from the plot e_n versus n ? What is the optimal choice for the model order based on the validation procedure?

- (c) **Regularized data fitting:** Instead of minimizing $\|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|^2$ of the data, now fit a model that minimizes, $\|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|^2 + \lambda \boldsymbol{\beta}^T \boldsymbol{\beta}$, where $\lambda \geq 0$. In this particular case fit the model order to a high value (e.g. 10) and the entire data set D . Perform the data fitting procedure for different values of λ . Plot $\|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|$ versus λ . Compare the values of $\hat{\boldsymbol{\beta}}$ for the different values of λ and compare these to your optimal choice of model parameters from the previous question.

4. Consider a time series $\mathbf{x} = \{x_0, x_1, \dots, x_{N-1}\}$ consisting of N data points, where n indicates time index. The time series is corrupted by noise, and we are interested in filtering the time series to obtain a smooth estimate of the the general trend in the time series. This can be posed a problem of estimating a new time series $\hat{\mathbf{x}}$, such that the difference $\mathbf{x} - \hat{\mathbf{x}}$ is minimized and $\hat{\mathbf{x}}$ is

smooth, i.e. the adjacent values of the signal do not change abruptly.

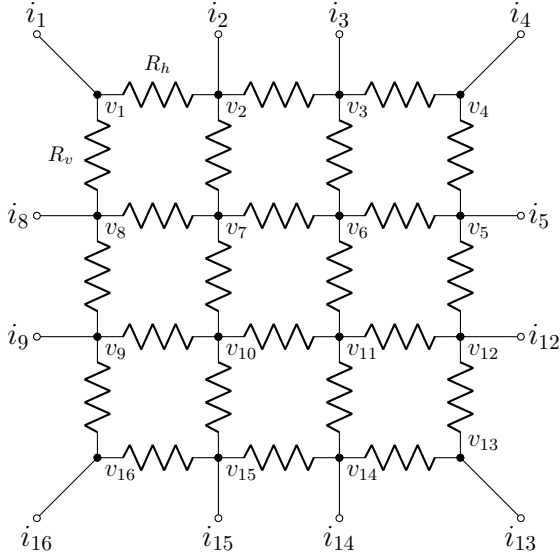
$$O(\hat{x}) = \sum_{i=1}^N (x_i - \hat{x}_i)^2 + \lambda \sum_{i=2}^{N-1} (2\hat{x}_i - \hat{x}_{i-1} - \hat{x}_{i+1})^2$$

If \mathbf{x} and $\hat{\mathbf{x}}$ are considered as N -vectors, then this can be written as

$$O(\hat{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 + \lambda \|\mathbf{D}\hat{\mathbf{x}}\|^2$$

You are provided with a CSV data file (timeseries.csv) consisting of a time series. Filter this time series by minimizing $O(\hat{\mathbf{x}})$ for different values of λ . Plot \mathbf{x} and $\hat{\mathbf{x}}$ for different values of λ . What role does λ play in the minimization problem?

5. Consider the following resistive network, where the horizontal resistors have resistance of $R_h = 1\Omega$ and the vertical resistors have a resistance of $R_v = 2\Omega$. Your goal is to determine a set of currents in the $\mathbf{i} = [i_1 \ i_2 \ \dots \ i_{16}]^T$, so as to achieve a particular distribution of potentials at the different nodes of the network. Note that we have control only over the currents at the edge nodes, and in the internal nodes the current is zero, i.e. $i_6 = i_7 = i_{10} = i_{11} = 0$.



Let $\mathbf{v} = [v_1 \ v_2 \ \dots \ v_{16}]$ represent the vector of potential distributions in the network. Then determine \mathbf{i} such that $\|\mathbf{v}_T - \mathbf{v}\|^2$ is minimized for the following desired potential distribution (Note that the potentials are

arranged in a matrix $\mathbf{V}_{map} = \begin{bmatrix} v_1 & v_2 & v_3 & v_4 \\ v_8 & v_7 & v_6 & v_5 \\ v_9 & v_{10} & v_{11} & v_{12} \\ v_{16} & v_{15} & v_{14} & v_{13} \end{bmatrix}$

(a) $\mathbf{V}_{map} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix}$

(b) $\mathbf{V}_{map} = \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 1 & 1 & 2 \\ 2 & 1 & 1 & 2 \\ 2 & 2 & 2 & 2 \end{bmatrix}$

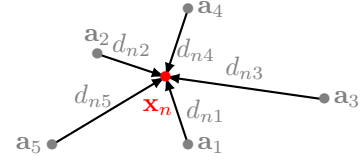
(c) $\mathbf{V}_{map} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

for some desired potential distribution \mathbf{v}_T , subject to the constraint $\sum_{k=1}^{16} i_k = 0$.

6. Trialiteration is a process of determining the position $\mathbf{x} = [x_1 \ x_2]^T$ of a point P given the distance of the point from a N control points of known locations $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N$. At any given point in time n , we have N different measurements corresponding to the distance of the point P from the N control points, i.e.

$$\mathbf{d}_n = [d_{n1} \ d_{n2} \ \dots \ d_{nN}]^T$$

where, $d_{ni}^2 = \|\mathbf{x}_n - \mathbf{a}_i\|_2^2$ for all $1 \leq i \leq N$.



\mathbf{x}_n are nonlinear functions of the measurements and the control point locations. However, taking the difference between two squared distance measurements d_j^2 and d_i^2 results in linear equations in the unknown coordinates,

$$(\mathbf{a}_i - \mathbf{a}_j)^T \mathbf{x} = \frac{(d_j^2 - \|\mathbf{a}_j\|_2^2) - (d_i^2 - \|\mathbf{a}_i\|_2^2)}{2}$$

Consider the point P that moves with time; $\mathbf{x}_n \in \mathbb{R}^2$ represents the location of the point at time instant n . The distance of this point from a set of 50 different control points is given to you in a CSV data file (trilatdist.csv); the data rows correspond to distance measurements from the different control points at a given point in time, while the columns correspond to the distance of the point from a particular control point for all time. The locations of the 50 control points are available in a separate file CSV file (trilatctrlpos.csv). Each of these distance measurements is affected by noise. Your aim is to use these distance measurements to reconstruct the trajectory of the point \mathbf{x}_n as a function of time.

- (a) What is the minimum number of distance measurements you need to estimate the unknown coordinates of the point \mathbf{x} ?
- (b) You are provided the actual trajectory of point P in the file trilatactpos.csv. How does you estimate compare with that of the actual trajectory? If we are informed that the point P does not undergo large changes in position between two consecutive time instants $n-1$ and n , how would you use this information to improve your estimate?