12.805 Homework 2

Bryan Kaiser

March 10, 2016

1 Underdetermined Least Squares

For systems in which the number of constraints N is less than the number of solution variables M, the Lagrange multiplier method can be utilized to find the solution vector $\mathbf{x}$ with a minimum square solution norm $\mathbf{J} = \mathbf{x}^\dagger\mathbf{x}$. For the given underdetermined system

$$\mathbf{E} = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

where

$$\mathbf{E}\mathbf{x} = \mathbf{y}. \tag{1}$$

To find a teneable solution such that $\mathbf{x}^\dagger\mathbf{x} = \min(\mathbf{x}^\dagger\mathbf{x})$ the Lagrange multiplier $\boldsymbol{\mu}$ is added to the minimum square solution norm so that the gradient of the norm can be used to find $\min(\mathbf{x}^\dagger\mathbf{x})$

$$\mathbf{J}' = \mathbf{x}^\dagger\mathbf{x} - 2\boldsymbol{\mu}^\dagger(\mathbf{E}\mathbf{x} - \mathbf{y}),$$

therefore the gradients of $\mathbf{J}'$ with respect to the solution and the multiplier are

$$\mathbf{J}' = \mathbf{x}^\dagger\mathbf{x} - 2\boldsymbol{\mu}^\dagger(\mathbf{E}\mathbf{x} - \mathbf{y}),$$

$$\frac{\partial \mathbf{J}'}{\partial \mathbf{x}} = 2\mathbf{x} - 2\mathbf{E}^\dagger\boldsymbol{\mu}, \quad \frac{\partial \mathbf{J}'}{\partial \boldsymbol{\mu}} = -2(\mathbf{E}\mathbf{x} - \mathbf{y}), \tag{2}$$

The stationary point of $\mathbf{J}'$ with respect to the solution and the multiplier by setting both of Equations (2) to zero and solving

$$\mathbf{x} = \mathbf{E}^\dagger\boldsymbol{\mu}, \quad \mathbf{y} = \mathbf{E}\mathbf{x}, \quad \Rightarrow \quad \boldsymbol{\mu} = (\mathbf{E}\mathbf{E}^\dagger)^{-1}\mathbf{y}, \quad \mathbf{x} = \mathbf{E}^\dagger(\mathbf{E}\mathbf{E}^\dagger)^{-1}\mathbf{y}, \tag{3}$$

therefore the solution for $\mathbf{x}$ in Equations (3) is the solution corresponding to $\min(\mathbf{J}')$. Note that the size of the vector $\boldsymbol{\mu}$ must always be $N \times 1$; its length is the number of constraints. For the given system (**Ans. 1a**)

$$\boldsymbol{\mu} = \begin{bmatrix} 5 \\ 7 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 0.750 \\ 0.750 \\ 0.500 \end{bmatrix}, \quad \min(\mathbf{x}^\dagger\mathbf{x}) = 11/8 = 1.375.$$

However, this solution does not permit any contraints to be placed on the noise norm for the system. To minimize the square noise norm for the system as well as the square solution norm, noise is introduced into Equation (1)

$$\mathbf{E}\mathbf{x} + \boldsymbol{\eta} = \mathbf{y}, \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}. \tag{4}$$

Weighted and tapered least squares can be utilized to find a solution vector for an underdetermined system with noise. Here, the goal is to minimize both the square solution norm *and* the square noise norm. Therefore, define the minimization scalar $\mathbf{J}$ as

$$\mathbf{J} = \mathbf{x}^{\dagger}\mathbf{S}^{-1}\mathbf{x} + \boldsymbol{\eta}^{\dagger}\mathbf{W}^{-1}\boldsymbol{\eta},$$

where the matrix $\mathbf{W}$ is a $N \times N$ matrix of chosen noise weights (that could be used to account for variable accuracy and/or correlations in data) and the matrix $\mathbf{S}$ is a $M \times M$ matrix for unit scaling and/or weighting the solutions. Substituting Equation (4) into $\boldsymbol{\eta}$ allows for the minimization of $\mathbf{J}$ with respect to $\mathbf{x}$ and subsequent solution for $\mathbf{x}$

$$\frac{\partial \mathbf{J}}{\partial \mathbf{x}} = 2\mathbf{E}^{\dagger}\mathbf{W}^{-1}(\mathbf{E}\mathbf{x} - \mathbf{y}) + \mathbf{x}^{\dagger}\mathbf{S}^{-1}\mathbf{x} = 0, \quad \Rightarrow \quad \mathbf{x} = (\mathbf{E}^{\dagger}\mathbf{W}^{-1}\mathbf{E} + \mathbf{S}^{-1})^{-1}\mathbf{E}^{\dagger}\mathbf{W}^{-1}\mathbf{y}. \tag{5}$$

With no other information regarding the problem, $\mathbf{W}$ and $\mathbf{S}$ are chosen as identity matrices in order assign each solution to have equal scaling and each noise element to have equal weight. For this choice of scalings and weights, (**Ans. 1b**)

$$\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Rightarrow \quad \mathbf{x} = \begin{bmatrix} 0.600 \\ 0.600 \\ 0.333 \end{bmatrix}.$$

Alternatively, one can minimize the square noise norm for an underdetermined system by placing a constraint on the solution behavior in order to construct a scaling matrix. In this case, the solution should be smooth, where "smoothness" is defined as minimized $(x_1 - x_2)^2 + (x_2 - x_3)^2$. Therefore the weighted and tapered cost function can be written using the "smoothness" parameter

$$\mathbf{D} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}, \quad \Rightarrow \quad \mathbf{x}^{\dagger}\mathbf{D}\mathbf{D}^{\dagger}\mathbf{x} = (x_1 - x_2)^2 + (x_2 - x_3)^2,$$

$$\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{S}^{-1} = \mathbf{D}\mathbf{D}^{\dagger} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix},$$

and solved by using Equation (5) once more (**Ans. 1c**)

$$\mathbf{x} = (\mathbf{E}^{\dagger}\mathbf{W}^{-1}\mathbf{E} + \mathbf{S}^{-1})^{-1}\mathbf{E}^{\dagger}\mathbf{W}^{-1}\mathbf{y} = \begin{bmatrix} 0.7667 \\ 0.7000 \\ 0.5667 \end{bmatrix}.$$

Note that the matrix operator for the smoothness constraint $\mathbf{D}$ could be written for any underdetermined system and will have the size $M \times N$.

Solution **1a.** minimizes the square solution norm, solution **1b.** minimizes the combined square solution norm and square noise norm (assuming uniform noise element weights and uniform solution scalings), and the solution **1c.** minimizes the combined square noise norm and a scaling derived from a constraint placed on the solution behavior (the "smoothness"). A scenario in which one might use any one of the three options could arise when working with real data, therefore they all could be correct when applied to the appropriate system. In other words, the correct solution will be the one that takes full advantage of what is known about the system (**Ans. 1d.**).

2 Underdetermined versus Overdetermined Least Squares

Given the well-determined system

$$\mathbf{E} = \begin{bmatrix} 1 & -3 \\ -3 & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

where

$$\mathbf{E}\mathbf{x} = \mathbf{y},$$

the solution for $\mathbf{x}$ can be obtained by using the ordinary (overdetermined) least squares method (the method for minimizing the square noise norm as shown in homework 1 and in the class notes, **Ans. 2a**)

$$\mathbf{E}\mathbf{x} + \boldsymbol{\eta} = \mathbf{y}, \quad \frac{d}{d\mathbf{x}}\boldsymbol{\eta}^\dagger\boldsymbol{\eta} = 0, \quad \Rightarrow \quad \mathbf{x} = (\mathbf{E}^\dagger\mathbf{E})^{-1}\mathbf{E}^\dagger\mathbf{y} = \mathbf{E}^{-1}\mathbf{y} = \begin{bmatrix} 1/4 \\ -1/4 \end{bmatrix}, \quad \boldsymbol{\eta}^\dagger\boldsymbol{\eta} = 0, \tag{6}$$

and by the method of Lagrange multipliers shown in problem 1 (using the Moore-Penrose inverse **Ans. 2b**)

$$\boldsymbol{\mu} = (\mathbf{E}\mathbf{E}^\dagger)^{-1}\mathbf{y} = \begin{bmatrix} 16 \\ -16 \end{bmatrix}, \quad \mathbf{x} = \mathbf{E}^\dagger(\mathbf{E}\mathbf{E}^\dagger)^{-1}\mathbf{y} = \begin{bmatrix} 1/4 \\ -1/4 \end{bmatrix}, \quad \min(\mathbf{x}^\dagger\mathbf{x}) = \mathbf{x}^\dagger\mathbf{x} = 1/8. \tag{7}$$

The solutions from the overdetermined least squares ($M < N$) method and underdetermined least squares ($M > N$) method collapse to the solution for the well-determined system (**Ans. 2c**) if the matrix $\mathbf{E}$ is square and invertible

$$(\mathbf{E}^\dagger\mathbf{E})^{-1}\mathbf{E}^\dagger = \mathbf{E}^\dagger(\mathbf{E}\mathbf{E}^\dagger)^{-1} = \mathbf{E}^{-1} \quad \forall \quad M = N,$$

because the squared solution norm is unique and squared noise norm is zero for a well-determined system.