

12. Uncertainty Quantification

We have studied exact and approximate methods for solving the three most important classes of partial differential equations. These are enormously important in modern scientific engineering modeling, but they all take for granted that we have known values for all the material parameters in the problem. In practice, we only know values of these quantities up to some tolerance. How do changes to these values affect the solution?

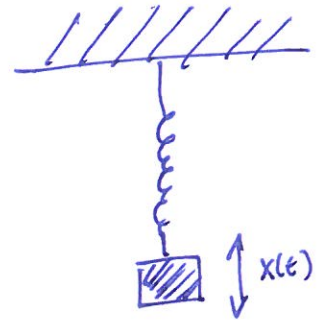
The field of uncertainty quantification provides tools to answer this question - and many more like it - from several different perspectives. For a good overview, I recommend the text: Uncertainty Quantification: Theory, Implementation, and Applications by Ralph C. Smith (SIAM, 2014), from which these notes are drawn.

To begin with, we should address two distinct perspectives on the subject. For a concrete example, consider the damped harmonic oscillator equation

$$x''(t) + c x'(t) + k x(t) = 0.$$

(Notice that this same equation arose in our study ^{36.2} of damped wave equations.)

We might hope to measure the coefficient K accurately; this describes, for example, the stiffness of a Spring, which can be estimated accurately using experiments with various masses and Hooke's Law. The damping coefficient, C , is more difficult to approximate. We would like to do one or more experiments, recording values of $x(t_k)$ at various fixed times t_1, \dots, t_n . From this data, we want to estimate C .



Frequentist Perspective: The equation $x''(t) + Cx'(t) + Kx(t) = 0$ describes the exact behavior of the system. We want to extract from the model an unbiased estimate for C and understand the variance of this estimate.

Bayesian Perspective: The data could be consistent with a range of different C values - but some of these C values are quite unlikely to have produced the data set. Using whatever prior knowledge of the experimental noise that we might have, build a probability distribution for C .

Frequentist Perspective (Smith, sections 7.2-7.3) 36.3

Our first goal is to construct MAXIMUM LIKELIHOOD estimates for a set of parameter values. We begin with a linear model

$$Y = X \theta_0 + \epsilon$$

Observations \nearrow design matrix \nearrow true parameters \nearrow noise

$$\begin{bmatrix} Y \\ \vdots \\ Y \end{bmatrix} = \begin{bmatrix} X \\ \vdots \\ X \end{bmatrix} \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_0 \end{bmatrix} + \begin{bmatrix} \epsilon \\ \vdots \\ \epsilon \end{bmatrix}$$

$X \in \mathbb{R}^{n \times p}$ $\theta_0 \in \mathbb{R}^p$

We assume the noise has zero expected value;

$$\mathbb{E}(\epsilon_i) = 0 \quad \text{for } i=1, \dots, n,$$

Constant variance

$$\text{Var}(\epsilon_i) = \sigma_0^2 = \text{constant for all } i=1, \dots, n$$

and is uncorrelated:

$$\text{Cov}(\epsilon_i, \epsilon_j) = 0 \quad \text{for } i, j=1, \dots, n, i \neq j.$$

The expected value of a random variable, P ,
mean of its probability distribution, denoted $\mathbb{E}(P)$.

The variance is defined by

$$\text{Var}(P) = \mathbb{E}((P - \mathbb{E}(P))^2)$$

$$= \mathbb{E}(P^2 - P\mathbb{E}(P) - \mathbb{E}(P)P + \mathbb{E}(P)^2)$$

$$= \mathbb{E}(P^2) - 2\mathbb{E}(P)\mathbb{E}(P) + \mathbb{E}(P)^2 \quad \left(\text{Using linearity of } \mathbb{E}(\cdot) \right)$$

$$= \mathbb{E}(P^2) - \mathbb{E}(P)^2$$

Covariance is defined in a similar fashion:

36.4

$$\begin{aligned}\text{Cov}(P, Q) &= \mathbb{E}((P - \mathbb{E}(P))(Q - \mathbb{E}(Q))) \\ &= \mathbb{E}(PQ) - \mathbb{E}(P)\mathbb{E}(Q)\end{aligned}$$

So that $\text{Var}(P) = \text{Cov}(P, P)$.

We will not go into more technical definitions; these can be found in any basic probability text.

Given our model

$$Y = Xq_0 + \varepsilon$$

We seek an unbiased estimator for q_0 . We shall find \hat{q} as the least squares solution

$$\|Y - X\hat{q}\| = \min_{q \in \mathbb{R}^n} \|Y - Xq\|$$

and show that it fits the bill.

We know how to find least-squares solution in a more general setting — this is just the best approximation problem from the start of the semester — but it helps to see how this is derived using different notation in Smith's setting. The error functional is given by

$$J(q) = (Y - Xq)^T (Y - Xq) = \|Y - Xq\|^2$$

To minimize the error function, find the 36.5
 $q \in \mathbb{R}^n$ that zeros out the gradient

$$\begin{aligned} \nabla_q J(q) &= \nabla_q (Y - Xq)^T (Y - Xq) \\ &\quad + (Y - Xq)^T \nabla_q (Y - Xq) \quad (\text{product rule}) \\ &= 2(Y - Xq)^T \nabla_q (Y - Xq) \end{aligned}$$

Note: $\nabla_q (Y - Xq) = \nabla_q \left(Y - \begin{bmatrix} \sum_k x_{1k} q_k \\ \vdots \\ \sum_k x_{nk} q_k \end{bmatrix} \right)$

Thus $\frac{\partial}{\partial q_j} (Y - Xq) = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{bmatrix} = j^{\text{th}} \text{ column of } X$

So $\nabla_q (Y - Xq) = X$

Hence, setting the gradient to zero:

$$\begin{aligned} 0 &= \nabla_q J(q) = 2(Y - Xq)^T \nabla_q (Y - Xq) \\ &= 2(Y - Xq)^T X \end{aligned}$$

$$\Rightarrow 0 = Y^T X - q^T X^T X$$

Taking the transpose $\Rightarrow \boxed{X^T X q = X^T Y}$

$$\boxed{X^T} \begin{bmatrix} X \\ q \end{bmatrix} = \boxed{X^T} \begin{bmatrix} Y \end{bmatrix}$$

Note that this is precisely what we would expect from our past experience from the theory of best approximation.

$$\boxed{X^T X} \begin{bmatrix} q \end{bmatrix} = \begin{bmatrix} X^T Y \end{bmatrix}$$

↑
A square $p \times p$ system

Write

$$X = [x_1 | x_2 | \dots | x_p]$$

The problem of approximating Y from the subspace $\mathcal{V} = \text{span}\{x_1, \dots, x_p\}$ gives $\min_c \left\| Y - \sum_{j=1}^p c_j x_j \right\|$,

Requiring us to construct the Gram matrix equation $Gc = b$

$$G = \begin{pmatrix} (x_1, x_1) & \dots & (x_1, x_p) \\ \vdots & & \vdots \\ (x_p, x_1) & \dots & (x_p, x_p) \end{pmatrix}, \quad b = \begin{pmatrix} (x_1, Y) \\ (x_2, Y) \\ \vdots \\ (x_p, Y) \end{pmatrix}.$$

Here the inner product for vectors u is just $(x_j, x_k) = x_k^T x_j$.

Thus we can also write

$$G = X^T X, \quad b = X^T Y$$

So solving $Gc = b$ gives the solution

$$c = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} \text{ such that } Xc = \sum_{j=1}^p c_j x_j$$

is the best approximation to Y .

Notice that what we have called " c " is the same thing as the solution q to

$$X^T X q = X^T Y.$$

We denote this solution as \hat{q} .