Comments and Reading

This is the last homework assignment. Although it is due on Thursday of the 10th week of the quarter (i.e., on the very last lecture of the quarter), you should start look at these problems immediately as understanding them is critical to performing the second computer assignment, which is due on the very same day.

Carefully review and study the class lectures (and the midterm question) on Maximum Likelihood Estimation (MLE) under the linear gaussian assumption. This material will help in solving problems 1, 2, and 3 given below.

Finally, read the lecture supplement on nonlinear least-squares and generalized gradient descent algorithms located on the class website. This information is critical for understanding the algorithms used in the second Matlab computer assignment.

2nd Matlab Computer Assignment - GPS Positioning - Due Thursday, 12/11/2014

As mentioned above, the second computer assignment on GPS positioning is also due Thursday, December 11, 2014. You should immediately begin this assignment.

Your report will be compared against descriptive websites and other articles on the GPS algorithm available on the internet and in the literature, as well as against other current and past project reports. Any copying from these banned sources into your project report will be considered an Academic Integrity violation and dealt with forcefully.

Homework

1. Maximum Likelihood Estimation for the Gaussian Linear Model. Suppose we obtain two different independent noisy measurements of an unknown quantity \( x \) using two different measurement devices,

\[
y_k = x + n_k , \quad k = 1, 2 ,
\]

where each measurement device \( k \), \( k = 1, 2 \), has been calibrated and is known to have a measurement error, \( n_k \), which is zero–mean and gaussian with a known variance \( \sigma_k^2 \). The standard deviation of measurement device \( k \), \( \sigma_k \), gives its precision and it is assumed that \( \sigma_1 \neq \sigma_2 \). Note, then, that we are measuring the quantity \( x \) using two devices of differing precision.

Let \( y = (y_1, y_2)^T \) denote the vector of observations. The conditional probability of \( y \) for a fixed possible value of the unknown \( x \), \( p(y | x) \), is known as the likelihood of \( x \) given the measurement \( y \), or simply as the likelihood function or likelihood. Under the independent–measurements assumption the likelihood can be written as the product of marginal probabilities as,

\[
p(y | x) = p(y_1 | x) \ p(y_2 | x) ,
\]
where under our noise model,
\[
p(y_k | x) = e^{-\frac{1}{2} \left( \frac{y_k - x}{\sigma_k} \right)^2} \sqrt{\frac{2}{\pi}} \sigma_k, \quad k = 1, 2.
\]

The maximum likelihood estimate (MLE), \( \hat{x}_{\text{MLE}} \), of the unknown quantity \( x \) is determined as,
\[
\hat{x}_{\text{MLE}} = \arg \max_x p(y | x) = \arg \min_x \{- \ln p(y | x) \}.
\]

(a) Show that the MLE, \( \hat{x}_{\text{MLE}} \), is the solution to a weighted least–squares problem,
\[
\hat{x}_{\text{MLE}} = \arg \min_x \| y - Ax \|_W^2.
\]

Give the quantities \( A, W, A^*, A^+ \), and determine the MLE, \( \hat{x}_{\text{MLE}} \).

(b) (i) What is the solution in the limit that \( \sigma_1 \to 0 \) (i.e., when device \( k = 1 \) is perfectly precise and has no measurement error). (iii) What is the solution in the limit that \( \sigma_1 \to \infty \) (i.e., when device \( k = 1 \) has no precision). (iii) Explain why these results make sense.

(c) Suppose that \( \sigma_1 = \sigma_2 = \sigma \). (This case could arise if the two measurements are made by the same measurement device.) Show that in this case the MLE is found as the solution to a regular (i.e., unweighted) least-squares problem and give the solution, \( \hat{x}_{\text{MLE}} \). Does this solution seem reasonable?

2. **Consistent Estimation in the Gaussian Linear Model.** Suppose that
\[
y_i = \alpha t_i + n_i \quad \text{for} \quad i = 1, \cdots, m,
\]
where \( n_i \sim N(0, \sigma^2) \) are iid.

(a) Find the least–squares estimate (which here is equal to the maximum likelihood estimate), \( \hat{\alpha}(m) \), obtained from the measurements \( (y_i, t_i) \), \( i = 1, \cdots, m \).

(b) Show that \( \hat{\alpha}(m) \) is unbiased, \( E \{ \hat{\alpha}(m) \} = \alpha \).

(c) Define the estimation error \( \tilde{\alpha}(m) = \hat{\alpha}(m) - \alpha \) and find the variance, \( \text{Var} \{ \tilde{\alpha}(m) \} \). Assuming that
\[
\sum_{i=1}^{m} t_i^2 \to \infty \quad \text{as} \quad m \to \infty,
\]
show that
\[
\text{Var} \{ \tilde{\alpha}(m) \} \to 0 \quad \text{as} \quad m \to \infty.
\]

This shows that
\[
\lim_{m \to \infty} \hat{\alpha}(m) = \alpha
\]
in the mean–square sense. An estimator that asymptotically converges to the unknown parameter is said to be consistent.
3. Dealing with Bias in the Gaussian Linear Model. In class we discuss the statistical model,
\[ y = Ax + n \quad n \sim N(0, \sigma^2 I), \]
where \( A \) is \( m \times n \) and has full column rank.\(^1\)

Now suppose that all measurements, \( y_i \), have the same systematic error (bias),
\[ \mathbb{E}\{n_i\} = \beta \neq 0, \quad \text{for} \quad i = 1, \ldots, m. \]

(a) Show that the least-squares estimate, \( \hat{x} \), is biased, \( \mathbb{E}\{\hat{x}\} \neq x \).

(b) Show that we can modify the problem to obtain an unbiased least-squares estimate of the augmented vector,
\[ \mathbf{x} = \begin{pmatrix} x \\ \beta \end{pmatrix}. \]

Do this by finding the least-squares estimate, \( \hat{\mathbf{x}} \), of the augmented vector and directly showing that it is unbiased, \( \mathbb{E}\{\hat{\mathbf{x}}\} = \mathbf{x} \). (Assume that the associated augmented \( A \)-matrix is one-to-one.)

4. Vocabulary.
Define the following: Regularized Least Squares; Maximum Likelihood Estimator; Multivariate Taylor Series Expansion; Gradient Descent Algorithm; Gauss-Newton Algorithm; Newton Algorithm; Generalized Gradient Descent Algorithm; Method of Lagrange Multipliers. (It is enough to paraphrase the definitions given in the lecture supplements.)

Using nonlinear least-squares, we wish to solve the nonlinear scalar inverse problem,
\[ y = h(x) \]
where both \( x \) and \( y \) are real and one-dimensional (scalar). Derive the following algorithms from first principles and show that each of them is a special case of generalized gradient descent (expressly show \( Q_j \) for each of the algorithms).

(a) Gradient Descent Algorithm.

(b) Gauss-Newton Algorithm (from iterative re-linearization of \( y = h(x) \))

(c) Newton Algorithm (from iterative quadratic approximation of the least-squares loss function).

Now use each of the three algorithms derived in the previous problem to find the cube root of an arbitrary real number. Code your algorithms in Matlab and use the algorithms to find the cube root of 1,000\(\pi\) using step sizes of 1.0; 0.5; and 0.1, and initial conditions of 3,000; 100, and 0.

---

\(^1\)As discussed in lecture, \( n \sim N(0, \sigma^2 I) \) is equivalent to the assumption that the noise components, \( n_i, i = 1, \ldots, m \), are iid with \( n_i \sim N(0, \sigma^2) \).