| Name: | ID#: |
|---|---|

# CSE 4345 / CSE 5315 - *Computational Methods*

Practice Exam 2- Fall 2011

Note: This problem set is somewhat larger and more difficult than one exam

Closed book, 2 pages of notes

Date: Nov. 29 2011, 3:00 pm - 8:20 pm

## Nonlinear Least Squares

1. Consider the following nonlinear least squares data fitting problem.

$$f(\alpha, x) = \alpha_1 + (x + \alpha_2)^2 \quad , \quad Data\ (x, y): \{((0, 1), (1, 3), (2, 4)\}$$

a) Derive the Gauss-Newton formulation for this problem.

For the Gauss-Newton method we need to derive the residual function and the corresponding Jacobian.

$$r(\alpha) = \begin{pmatrix} y_1 - f(\alpha, x_1) \\ y_2 - f(\alpha, x_2) \\ y_3 - f(\alpha, x_3) \end{pmatrix} = \begin{pmatrix} 1 - \alpha_1 - \alpha_2^2 \\ 3 - \alpha_1 - (1 + \alpha_2)^2 \\ 4 - \alpha_1 - (2 + \alpha_2)^2 \end{pmatrix}$$

$$J_r(\alpha) = \begin{pmatrix} -1 & -2\alpha_2 \\ -1 & -2(1 + \alpha_2) \\ -1 & -2(2 + \alpha_2) \end{pmatrix}$$

The Gauss-Newton formulation then provides an iterative step:

$$J_r^T(\alpha_k) J_r(\alpha_k) s = -J_r^T r(\alpha_k)$$
$$\alpha_{k+1} = \alpha_k + s$$

b) Show the first 2 steps of Gauss-Newton on the resulting system. (You can use normal equations to solve the linear least squares step).

Using normal equations and Gaussian Elimination to solve the linear step:

Initial conditions: $\alpha_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$r(\alpha_0) = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

$$J_r(\alpha_0) = \begin{pmatrix} -1 & 0 \\ -1 & -2 \\ -1 & -4 \end{pmatrix}$$

Step 1:

$$J_r^T(\alpha_0)J_r(\alpha_0) = \begin{pmatrix} -1 & -1 & -1 \\ 0 & -2 & -4 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ -1 & -2 \\ -1 & -4 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 20 \end{pmatrix}$$

$$-J_r^T(\alpha_0)r(\alpha_k) = -\begin{pmatrix} -1 & -1 & -1 \\ 0 & -2 & -4 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = -\begin{pmatrix} -3 \\ -4 \end{pmatrix}$$

Solving using Gaussian elimination:

$$\begin{pmatrix} 3 & 6 \\ 6 & 20 \end{pmatrix} s = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 6 \\ 0 & 8 \end{pmatrix} s = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$

$$s_2 = \frac{-2}{8} = -\frac{1}{4}$$

$$s_1 = \frac{3-6*s_2}{3} = \frac{3+\frac{3}{2}}{3} = \frac{3}{2}$$

$$\alpha_1 = \alpha_0 + s = \begin{pmatrix} \frac{3}{2} \\ -\frac{1}{4} \end{pmatrix}$$

$$r(\alpha_1) = \begin{pmatrix} 1 - \frac{3}{2} - \frac{1}{16} \\ 3 - \frac{3}{2} - \frac{9}{16} \\ 4 - \frac{3}{2} - \frac{49}{16} \end{pmatrix} = \begin{pmatrix} -\frac{9}{16} \\ \frac{15}{16} \\ -\frac{9}{16} \end{pmatrix}$$

$$J_r(\alpha_1) = \begin{pmatrix} -1 & \frac{1}{2} \\ -1 & -\frac{3}{2} \\ -1 & -\frac{7}{2} \end{pmatrix}$$

Step 2:

$$J_r^T(\alpha_1)J_r(\alpha_1) = \begin{pmatrix} -1 & -1 & -1 \\ \frac{1}{2} & -\frac{3}{2} & -\frac{7}{2} \end{pmatrix} \begin{pmatrix} -1 & \frac{1}{2} \\ -1 & -\frac{3}{2} \\ -1 & -\frac{7}{2} \end{pmatrix} = \begin{pmatrix} 3 & \frac{9}{2} \\ \frac{9}{2} & \frac{59}{4} \end{pmatrix}$$

$$-J_r^T(\alpha_1)r(\alpha_1) = -\begin{pmatrix} -1 & -1 & -1 \\ \frac{1}{2} & -\frac{3}{2} & -\frac{7}{2} \end{pmatrix} \begin{pmatrix} -\frac{9}{16} \\ \frac{15}{16} \\ -\frac{9}{16} \end{pmatrix} = -\begin{pmatrix} \frac{3}{16} \\ \frac{9}{32} \end{pmatrix}$$

Solving using Gaussian elimination:

$$\begin{pmatrix} 3 & \frac{9}{2} \\ \frac{9}{2} & \frac{59}{4} \end{pmatrix} s = \begin{pmatrix} -\frac{3}{16} \\ -\frac{9}{32} \end{pmatrix}$$
$$\begin{pmatrix} 3 & \frac{9}{2} \\ 0 & \frac{32}{4} \end{pmatrix} s = \begin{pmatrix} -\frac{3}{16} \\ 0 \end{pmatrix}$$
$$s_2 = 0$$
$$s_1 = \frac{-\frac{3}{16} - 0}{3} = -\frac{1}{16}$$
$$\alpha_2 = \alpha_1 + s = \begin{pmatrix} \frac{23}{16} \\ -\frac{1}{4} \end{pmatrix}$$

# Unconstrained Optimization

2. Optimization is concerned with finding an optimum of a function.

   a) What is the difference between a local and a global optimum ?

      A local optimum is any point that is an optimum (minimal or maximal) within an (arbitrarily small) local neighborhood of feasible points around the point. A global optimum is a point that is an optimum (i.e. minimal or maximal) for all feasible points within the entire domain of the function.

   b) Name conditions under which all local minima of a function are also global minima.

      If a function $f(x)$ is convex on the set of feasible points then every local minimum is also a global minimum.

## One-Dimensional Optimization

3. In the Bisection method for root finding, the concept of a bracket was based on the observation that if the two endpoints of an interval have opposite signs and the function is continuous, then the function has to take on a value of 0 somewhere in the interval. Using this, computing the

value of the midpoint allowed to select which half of the interval formed a new bracket. Golden Section Search uses a similar concept but uses three points to characterize the active interval. Using this and picking one new point and computing its value here again allows the algorithm to deterministically remove one of the endpoints, effectively making the interval narrower.

a) Discuss how the data point is chosen and how the endpoint that is removed is selected.

In golden section search, the additional interior point in the already existing interval $[a, b]$ is chosen either as $a + \frac{3-\sqrt{5}}{2}(b-a)$ or as $a + \frac{\sqrt{5}-1}{2}(b-a)$ depending wether the right or the left interval end point has been removed in the last iteration.

Given the four points, $a$, $x_1$, $x_2$, $b$, defining the interval $[a, b]$ and the two interior points ($x_1$ being the left and $x_2$ being the right), the interval end point that is closer to the larger interior point is removed and that interior point becomes the new interval endpoint.

b) Describe the rationale and assumptions behind this choice and how they ensure that there will always remain a minimum in the remaining interval.

The assumption underlying the interval shrinking criterion in Golden Section Search is that the function is convex over the given interval (it is actually sufficient that it has a convex section since it will eventually shrinking down to the point where it contains only one of the convex segments). If the function is convex over the interval (which requires that the initial interior point has a lower function value than the interval endpoints), then a local minimum of the function has to lie between the interior point with the lower function value and either its neighboring interval endpoint or the other interior point. This is the case since these three points are (given that they fulfill this condition initially) always such that the interior point's function value is lower than the function values of the new interval bounds and thus any convex function through them has to have a minimum within this interval.

4. Consider the following nonlinear optimization problem to find the value for $x$ that (locally) minimizes the function $f(x) = x^4 + 3x^2 + 4x - 6$ .

a) Formulate Newton's method for one-dimensional optimization for this problem (i.e. derive all the terms you need to apply Newton's method).

To use Newton's method for one-dimensional optimization we need the first and second derivative of the function:

$$f'(x) = 4x^3 + 6x + 4$$
$$f''(x) = 12x^2 + 6$$

This gives us Newton's method as:

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} = x_k - \frac{4x_k^3 + 6x_k + 4}{12x_k^2 + 6} = x_k - \frac{2x_k^3 + 3x_k + 2}{6x_k^2 + 3}$$

b) Show the first three iterations of Newton's method on this problem starting from $x_0 = 0$.

Step 1:

$x_1 = x_0 - \frac{2x_0^3+3x_0+2}{6x_0^2+3} = -\frac{2}{3}$

Step 2:

$x_2 = x_1 - \frac{2x_1^3+3x_1+2}{6x_1^2+3} = -\frac{2}{3} - \frac{2(-\frac{2}{3})^3+3(-\frac{2}{3})+2}{6(-\frac{2}{3})^2+3} = -\frac{2}{3} - \frac{-\frac{16}{27}-2+2}{\frac{24}{9}+3} = -\frac{2}{3} - \frac{-\frac{16}{27}}{\frac{51}{9}} = -\frac{2}{3} + \frac{16}{153} = -\frac{86}{153}$

Step 3:

$x_3 = x_2 - \frac{2x_2^3+3x_2+2}{6x_2^2+3} = -\frac{86}{153} - \frac{2(-\frac{86}{153})^3+3(-\frac{86}{153})+2}{6(-\frac{86}{153})^2+3} = -\frac{86}{153} - \frac{-\frac{1272112}{3581577}-\frac{258}{153}+2}{\frac{44376}{23409}+3} = -0.56209 - \frac{-0.0414566}{4.895681} - 0.5536235$

## Multi-Dimensional Optimization

5. Present the basic idea behind the Nelder-Mead method for direct search-based multi-dimensional optimization. In particular, how does it address the shortcomings of Golden Section search in multi-dimensional optimization problems and how do the different operations address finding an area containing a minimum.

   The main problem with extending Golden Section Search to multi-dimensional optimization is that if the interval shrinking is applied independently to different dimensions, then it is possible that the complete (multi-dimensional) interval during shrinking will drop the interval, effectively not being able to converge around the interval. This could partially be addressed by explicitly modeling the interval as a multi-dimensional interval. This, however would increase the number of points required exponentially, making it intractable. Nelder-Mead addresses this starting with a simplex and allowing it to not only shrink but also to move and (temporarily) expand in oder to allow it to search for a convex region that contains a local minimum and then converging around it.

   Reflection is the main operation used to allow the area to move if the values of the simplex points suggest that it is more likely that the minimum is on the opposite side of the hyperplane described by all the points except the worst one. Expansion is used to (temporarily) increase the size of the area in order to make sure the minimum is contained within the region. Contraction and Reduction are used to shrink the area if it looks as if the minimum lies within the current simplex.

6. Consider the following unconstrained optimization problem to find the values for $x$ and $y$ that minimize the value for the function $f(x, y) = x^2 + y^4 + x + y$.

   a) Derive the terms needed for an optimization using Newton's method for multi-dimensional optimization.

      For Newton's method for multi-dimensional optimization we need the derivative of the function and its Hessian (matrix of second derivatives):

$$\nabla f(x, y) = \begin{pmatrix} 2x + 1 \\ 4y^3 + 1 \end{pmatrix}$$

$$H_f(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & 12y^2 \end{pmatrix}$$

From this we get Newton's method, $H_f(x_k, y_k)s = -\nabla f(x_k, y_k)$, as:

$$\begin{pmatrix} 2 & 0 \\ 0 & 12y_k^2 \end{pmatrix} s = -\begin{pmatrix} 2x_k + 1 \\ 4y_k^3 + 1 \end{pmatrix}$$
$$x_{k+1} = x_k + s_1$$
$$y_{k+1} = y_k + s_2$$

b) Show the first 2 steps of Newton's method on this problem starting with $x_0 = y_0 = 0$.

Initial conditions:

$$\nabla f(x_0, y_0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$
$$H_f(x_0, y_0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

From this we can see that the rank of the Hessian is less than 2. This implies that we can not compute its inverse and thus not compute the Newton step (which would also not converge).

Note: this was an oversight on my part, so tho show the actual iteration, we will slightly change the initial conditions to $y_0 = -0.1$. Then we have:

$$\nabla f(x_0, y_0) = \begin{pmatrix} 1 \\ 0.8 \end{pmatrix}$$
$$H_f(x_0, y_0) = \begin{pmatrix} 2 & 0 \\ 0 & 0.12 \end{pmatrix}$$

Step 1:

$$\begin{pmatrix} 2 & 0 \\ 0 & 0.12 \end{pmatrix} s = \begin{pmatrix} -1 \\ -0.8 \end{pmatrix}$$
$$s_1 = -0.5$$
$$s_2 = -\tfrac{2}{3}$$
$$x_1 = 0 - 0.5 = -0.5$$
$$y_1 = -0.1 - \tfrac{2}{3} = -\tfrac{23}{30}$$

Step 2:

$$\nabla f(x_1, y_1) = \begin{pmatrix} 0 \\ 0.54937 \end{pmatrix}$$
$$H_f(x_0, y_0) = \begin{pmatrix} 2 & 0 \\ 0 & 7.05333 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 0 \\ 0 & 7.05333 \end{pmatrix} s = \begin{pmatrix} 0 \\ -0.54937 \end{pmatrix}$$
$$s_1 = 0$$
$$s_2 = -0.077888$$
$$x_2 = -0.5 + 0 = -0.5$$
$$y_2 = -0.54937 - 0.077888 = -0.627258$$

# Constrained Optimization

7. Constrained optimization problems allow for two types of constraints: equality constraints and inequality constraints.

   a) Discuss some of the differences between optimization with only equality constraints and with inequality constraints. In particular focus on the applicability of different methods to the solution.

      Equality constraints basically tightly constrain the variables. This allows the application of simpler methods than with inequality constraints. In particular, it allows to use the Lagrange function to build a system of equations that includes the objective function as well as the equality constraints. With inequality constraints, this method can only be used if it is possible to identify the constraints that are active (i.e. for which the equality part holds at the solution). Similarly, the use of a merit function is easier with equality constraints since they can be modeled using a quadratic (and thus symmetric) penalty function. Inequality constraints, on the other hand, have to be modeled using asymmetric penalty functions which are not defined for parts of the space and thus the used unconstrained optimization approach has to verify explicitly that it did not violate a constraints.

   b) Why is optimization with inequality constraints more difficult than with equality constraints ?

      Inequality are harder to deal with since they not always influence the choice of variables. In particular, they are only ”active” if the constrained minimum lies on the constraint (i.e. fulfills the equal relation). Since it is not possible a-priori to determine which constraints will be active at the solution, the solution methods have to explicitly deal with the asymmetric influence these constraints have on the solution. This makes converting the constrained optimization problem into a modified unconstrained optimization problem significantly more difficult.

8. Derive the extended form with slack variables for the following linear programming problem.

$$\begin{aligned} Objective\ function\ :&\quad f(\vec{x}) = x_1 + 5x_2 - 7x_3 \\ Constraints\ :&\quad h_1(\vec{x}) = 2x_1 - 3x_2 - 2 \leq 0 \\ &\quad h_2(\vec{x}) = x_1 + 2x_3 - 4 \leq 0 \\ &\quad h_3(\vec{x}) = 3x_2 - 6x_3 \leq 0 \\ &\quad x_1 \geq 0\ ,\ x_2 \geq 0\ ,\ x_3 \geq 0 \end{aligned}$$

Since there are 3 constraints we need 3 slack variables $xs_i$. Also, to include the objective function, we introduce an additional variable $F$, leading to the following system:

$$\begin{aligned} Objective\ function\ :&\quad F - x_1 - 5x_2 + 7x_3 = 0 \\ Constraints\ :&\quad 2x_1 - 3x_2 + xs_1 = 2 \\ &\quad x_1 + 2x_3 + xs_2 = 4 \\ &\quad 3x_2 - 6x_3 + xs_3 = 0 \\ &\quad x_1 \geq 0\ ,\ x_2 \geq 0\ ,\ x_3 \geq 0\ ,\ xs_1 \geq 0\ ,\ xs_2 \geq 0\ ,\ xs_3 \geq 0 \end{aligned}$$

This results in the extended form system:

$$\begin{pmatrix} 1 & -1 & -5 & 7 & 0 & 0 & 0 \\ 0 & 2 & -3 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 0 & 3 & -6 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} F \\ x_1 \\ x_2 \\ x_3 \\ xs_1 \\ xs_2 \\ xs_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 4 \\ 0 \end{pmatrix}$$

9. One way to address constrained optimization with equality constraints is to convert the original objective function into a merit function by adding a weighted penalty function for the constraints. The most common one of these is the square of the constraints, $g(x)^T g(x)$.

   Consider the following optimization problem with equality constraints:

$$\begin{aligned} Objective\ function\ :&\quad f(\vec{x}) = x_1 + x_2^2 \\ Constraints\ :&\quad g_1(\vec{x}) = x_1 - x_2 = 0 \end{aligned}$$

a) Show the merit function for this problem.

Using the penalty function described above, the Merit function would be
$\phi(\vec{x}) = f(\vec{x}) + \frac{1}{2}\rho g(\vec{x})^T g(\vec{x}) = x_1 + x_2^2 + \frac{1}{2}\rho(x_1 - x_2)^2$

b) Provide a description of the basic algorithm to solve this problem using the merit function.

The basic algorithm operates as follows:

i. Pick an initial value of $\rho_0$ (e.g. $\rho_0 = 1$) and a start point $x_0$
ii. Run an unconstrained optimization algorithm on the resulting Merit function
iii. Increase $\rho$ (e.g. $\rho_{k+1} = \rho_k * 4$)
iv. if $\rho_{k+1}$ is smaller than a convergence threshold, goto step 2. using the previous solution of the unconstrained optimization as the start point
v. Return the solution to the last unconstrained optimization

c) Show the first 2 iterations of the solution using the unconstrained optimization scheme of choice in each iteration. Note: an iteration here refers to a choice of mixing parameter between objective function and penalty term, not an iteration of the unconstrained optimization algorithm you choose (e.g. Steepest descent, Newton's method, etc). You can be very lax about the termination condition for your optimization approach in each iteration (i.e. you can let it converge after a very small number of iterations - e.g. 3). You should start at $x_1 = x_2 = 0$ in the first step of the first iteration.

Using steepest descent and 2 steps for each optimization

Initial Conditions:

$$\rho_0 = 1$$
$$\vec{x}_0 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

First Iteration:

Perform steepest descent with line search:

Step 1:

$$\nabla\phi(\vec{x}_0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
$$Line\ :\ \vec{x}_1 = \vec{x}_0 - \alpha_0 \nabla\phi(\vec{x}_0)$$
$$Find\ \alpha\ that\ minimizes\ value\ along\ line:$$
$$\alpha_0 = argmin_\alpha \phi(-\alpha, 0)$$
$$\alpha_0 = 0.5$$
$$\vec{x}_1 = \begin{pmatrix} -0.5 \\ 0 \end{pmatrix}$$

Step 2:

$$\nabla\phi(\vec{x}_1) = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$
$$Line\ :\ \vec{x}_2 = \vec{x}_1 - \alpha_1 \nabla\phi(\vec{x}_1)$$
$$Find\ \alpha\ that\ minimizes\ value\ along\ line:$$
$$\alpha_1 = argmin_\alpha \phi(-0.5 - 0.5 * \alpha, 0 - 0.5 * \alpha)$$
$$\alpha_1 = 1$$
$$\vec{x}_2 = \begin{pmatrix} -1 \\ -0.5 \end{pmatrix}$$

Second Iteration: Change $\rho$ to 4

Perform steepest descent with line search:

Step 1:

$$\nabla\phi(\vec{x}_2) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$
$$Line\ :\ \vec{x}_3 = \vec{x}_2 - \alpha_2 \nabla\phi(\vec{x}_2)$$
$$Find\ \alpha\ that\ minimizes\ value\ along\ line:$$
$$\alpha_2 = argmin_\alpha \phi(-1 + \alpha, -0.5 - \alpha)$$
$$\alpha_2 = -\frac{3}{8}$$
$$\vec{x}_3 = \begin{pmatrix} -\frac{5}{8} \\ -\frac{1}{8} \end{pmatrix}$$

Step 2:

$$\nabla\phi(\vec{x}_3) = \begin{pmatrix} -1 \\ \frac{7}{4} \end{pmatrix}$$
$$Line\ :\ \vec{x}_4 = \vec{x}_3 - \alpha_3 \nabla\phi(\vec{x}_3)$$
$$Find\ \alpha\ that\ minimizes\ value\ along\ line:$$
$$\alpha_3 = argmin_\alpha \phi(-\frac{5}{8} + \alpha, -\frac{1}{8} - \frac{7}{4} * \alpha)$$
$$\alpha_3 = 0.111684$$
$$\vec{x}_4 = \begin{pmatrix} -0.736684 \\ -0.236684 \end{pmatrix}$$