Math 433/833 Spring 2009 **Project #3** Due Mon Apr 27

This project can be done using Sage or any other computer software of your choice. However, you cannot use any built-in functions that make the problem trivial. Basic linear algebra commands (such as taking the inverse of a matrix) and calculus commands (such as computing the gradient or Hessian) are fine to use.
Submit (on paper) your code with explanation—comments in the code are helpful.
You may work individually or in groups of two for this project. Each group should submit just one project report.

1. Gauss used least squares minimization to predict the orbit of Ceres in 1801. This project will examine aspects of fitting data with the simpler figures of circles. The equation of a circle in $\mathbb{R}^2$ centered at (x_0, y_0) and with radius a is
$$\frac{(x-x_0)^2}{r^2} + \frac{(y-y_0)^2}{r^2} = 1. \tag{1}$$
An equivalent parametric form is $(x, y) = (r\cos(t), r\sin(t)) + (x_0, y_0)$, for $0 \le t < 2\pi$.
2. Let $f(x, y)$ denote the left side of equation (1). Given data $(x_1, y_1), \ldots, (x_n, y_n)$, use least squares optimization to find the center and radius of the circle that minimizes $\sum_{i=1}^{n} \left(f(x_i, y_i) - 1\right)^2$. Since $f(x, y)$ is not linear in x_0, y_0, a, it's not immediately clear how to use least squares optimization. Write $f(x, y) = c_0(x^2+y^2)+c_1x+c_2y$, and use least squares optimization to find the best c_0, c_1, c_2. Then solve for x_0, y_0, a in terms of c_0, c_1, c_2. For each of the data sets given on the course webpage, apply this method.
3. One problem with the above method is that it's not clear what the function $\sum_{i=1}^{n} \left(f(x_i, y_i) - 1\right)^2$ is minimizing. We would like to minimize the sum of the squares of the distances of the data points to the circle. Given a circle with center (x_0, y_0) and radius r, determine a formula for the distance from (x_i, y_i) to the closest point on the circle.
4. The Gauss-Newton Method: Suppose that we wish to minimize the function $f(\mathbf{x}) = \sum_{i=1}^{n} r_i(\mathbf{x})^2$ over $\mathbb{R}^m$, where $r_i : \mathbb{R}^m \to \mathbb{R}$ for $i = 1, \ldots, n$. Using Newton's Method for minimization, we would iteratively update $\mathbf{x}^{(k)}$ using the rule
$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \left[Hf(\mathbf{x}^{(k)})\right]^{-1} \nabla f(\mathbf{x}^{(k)}).$$
Note that the jth component of $\nabla f(\mathbf{x})$ is $\sum_{i=1}^{n} 2r_i(\mathbf{x})\frac{\partial r_i(\mathbf{x})}{\partial x_j}$, and the j, ℓ-th component of $Hf(\mathbf{x})$ is
$$2\sum_{i=1}^{n}\left(\frac{\partial r_i(\mathbf{x})}{\partial x_j}\frac{\partial r_i(\mathbf{x})}{\partial x_\ell} + r_i(\mathbf{x})\frac{\partial^2 r_i(\mathbf{x})}{\partial x_j}\right).$$
If we neglect the second-order derivatives, then
$$Hf(\mathbf{x}) \approx 2\sum_{i=1}^{n}\left(\frac{\partial r_i(\mathbf{x})}{\partial x_j}\frac{\partial r_i(\mathbf{x})}{\partial x_\ell}\right) = 2\sum_{i=1}^{n} J_{ij}J_{i\ell},$$
where $J_{ij} = \frac{\partial r_i(\mathbf{x})}{\partial x_j}$ is the i, j-th entry of the Jacobian matrix J for $\mathbf{r}(\mathbf{x}) = (r_1(\mathbf{x}), \ldots, r_n(\mathbf{x}))$. Thus, the gradient $\nabla f(\mathbf{x})$ can be rewritten as $2J^{\mathrm{T}}\mathbf{r}$ and the approximate Hessian as $Hf(\mathbf{x}) \approx 2J^{\mathrm{T}}J$. The update rule for the Gauss-Newton Method then becomes
$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \left[2J^{\mathrm{T}}J\right]^{-1} 2J^{\mathrm{T}}\mathbf{r}(\mathbf{x}^{(k)}) = \mathbf{x}^{(k)} - (J^{\mathrm{T}}J)^{-1}J^{\mathrm{T}}\mathbf{r}(\mathbf{x}^{(k)}), \quad \text{where } J = J(\mathbf{x}^{(k)}).$$

Note that this is equivalent to solving the equation $(J^{\mathrm{T}}J)\left[\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right] = -J^{\mathrm{T}}\mathbf{r}(\mathbf{x}^{(k)})$, which is the normal equation giving the best least squares fit to the linear system $J\left[\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right] = -\mathbf{r}(\mathbf{x}^{(k)})$. The expression

$$\mathbf{r}(\mathbf{x}) \approx \mathbf{r}(\mathbf{x}^{(k)}) + J\left[\mathbf{x} - \mathbf{x}^{(k)}\right]$$

is the first-order Taylor's series approximation of $\mathbf{r}(\mathbf{x})$ centered at $\mathbf{x}^{(k)}$. The next value $\mathbf{x}^{(k+1)}$is given by where $\mathbf{r}(\mathbf{x}^{(k+1)}) = 0$. In general, the linear system $\mathbf{r}(\mathbf{x}^{(k)}) + J\left[\mathbf{x} - \mathbf{x}^{(k)}\right] = \mathbf{0}$ is inconsistent, and so the best least squares fit is used to find $\mathbf{x}^{(k+1)}$.

5. For each of the given data sets, find the circle with center (x_0, y_0) and radius a that minimizes the sum of the squares of the distances from each data point to the circle. That is, minimize

$$\sum_{i=1}^{n} \left(\|(x_i, y_i) - (x_0, y_0)\| - a\right)^2 .$$

Note that this is a nonlinear least squares problem. Use the Gauss-Newton Method to minimize the function, starting from the solution found above.

6. Another slightly different approach is to minimize the function

$$\sum_{i=1}^{n} \left(\|(x_i, y_i) - (x_0, y_0)\|^2 - a^2\right)^2 .$$

Let $\mathbf{z}^{(0)} = (x_0, y_0)$ and $\mathbf{z}^{(i)} = (x_i, y_i)$. Then

$$\begin{aligned} a^2 - \|(x_0, y_0) - (x_i, y_i)\|^2 &= a^2 - \mathbf{z}^{(0)} \cdot \mathbf{z}^{(0)} + 2\mathbf{z}^{(0)} \cdot \mathbf{z}^{(i)} - \mathbf{z}^{(i)} \cdot \mathbf{z}^{(i)} \\ &= \mathbf{b}^{(i)} \cdot \mathbf{u} - \mathbf{z}^{(i)} \cdot \mathbf{z}^{(i)}, \end{aligned}$$

where $\mathbf{u} = (2x_0, 2y_0, a^2 - \mathbf{z}^{(0)} \cdot \mathbf{z}^{(0)})$ and $\mathbf{b}^{(i)} = (x_i, y_i, 1)$. Thus, we wish to minimize

$$\sum_{i=1}^{n} \left(\mathbf{b}^{(i)} \cdot \mathbf{u} - \mathbf{z}^{(i)} \cdot \mathbf{z}^{(i)}\right)^2 = \|B\mathbf{u} - \mathbf{w}\|^2 ,$$

where the rows of B are $\mathbf{b}^{(i)}$ and the components of $\mathbf{w}$ are $\|(x_i, y_i)\|^2$. This can be solved using our linear algebra approach to least squares optimization. Apply this method to each of the data sets. How do the fitted circles compare to the previous methods?

7. For each data set, plot the data points and the ellipses obtained from each method. Are the ellipses good fits to the data?