Each question is worth 2 points, the maximum possible score is 50 points. When entering the answers, please use the same rounding rules as before (specified at the top of module 1 problem set), if applicable.

To be eligible for partial credit, detailed work has to be submitted with the Submit All Solutions form (as a file attached to the last question). However, partial credit will be applied only if your total score for the exam falls below 40 points.

If you're sure you won't need partial credit, submitting detailed work is optional.

The Least-Squares Problem

Reading: Boyd, Convex Optimization, 2004, Chapter 1, Sections 1.1-1.2
http://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf

Reading: Rao, Engineering Optimization: Theory and Practice, 2009, pages 779 – 783.
https://onlinelibrary-wiley-com.proxy1.library.jhu.edu/doi/book/10.1002/9780470549124

$$\min f(x) = \|Ax - b\|_2^2$$

is a linear least-squares problem. Let's check whether the objective function is convex,

$$f(x) = \|Ax - b\|_2^2 = (Ax - b)^T (Ax - b) = (Ax)^T (Ax) - b^T Ax - (Ax)^T b + b^T b = x^T A^T Ax - 2b^T Ax + b^T b$$

$A^T A$ is 1? because it is symmetric and, for every z, $z^T A^T Az = (Az)^T Az = \|Az\|_2^2 \geq 0$. Therefore, f(x) is 2?.

Question 1: The 1? is
- positive definite
- negative definite
- positive semidefinite
- negative semidefinite
- none of the above

Question 2: The 2? is
- convex
- not convex
- concave
- not concave
- none of the above

This optimization problem can be solved analytically:

$$f(x) = \|Ax - b\|_2^2 = (Ax - b)^T (Ax - b) = x^T A^T Ax - 2b^T Ax + b^T b$$

The gradient is zero at the minimum.

In module 4, we derived the gradient of a quadratic function,

$$g(x) = x^T Qx + b^T x$$

$$\nabla g(x) = \left(Q + Q^T\right)x + b \qquad \text{or} \quad \nabla g(x) = 2Qx + b \text{ if Q is symmetric.}$$

Therefore,

$$\nabla f(x) = 2A^T Ax - ? = 0$$

Question 3: Replace the question mark with the correct expression.

- $2\left(b^T A\right)^T$
- $2b^T A$
- both $2\left(b^T A\right)^T$ and $2b^T A$ are correct
- none of the above

If $A^T A$ is invertible, we get

$$x = \left(A^T A\right)^{-1} A^T b$$

Next, let's obtain the same result using Newton's method.

The gradient of the objective function is

$$\nabla f(x) = 2\left(A^T Ax - A^T b\right)$$

The Hessian of the objective function is (The Hessian of a quadratic function was derived in module 4)

$$\nabla^2 f(x) = ?$$

Question 4: Replace the question mark above with the correct expression.

- $2A^T A$
- $2AA^T$
- both $2A^T A$ and $2AA^T$ are correct
- none of the above

Applying Newton's method,

$$\nabla^2 f(x^{(k+1)})(x^{(k+1)} - x^{(k)}) = -\nabla f(x^{(k)})$$

we have

$$2A^T A(x^{(k+1)} - x^{(k)}) = -2\left(A^T A x^{(k)} - A^T b\right)$$

$$A^T A x^{(k+1)} = A^T b$$

The first iteration of Newton's method gives the final solution.

Question 5: Use the analytical solution to the least-squares problem to fit a line with y-intercept equal 0 to points (-0.9, -3.9), (-0.6, -3.8), (-0.4, -2.9), (-0.4, -1.1), (-0.1, -0.9), (0, 0.7), (0.4, 1.2), (0.5, 1.8), (0.6, 2.3), (0.6, 3.9). The slope of this line is 4.7833

Question 6: (continued from the previous question) Plot the points and the line in MATLAB to verify the answer. Please attach a screenshot of your plot here.

Question 7: Use the analytical solution to the least-squares problem to fit a line to points (-1.1, -4), (-1, -3), (-0.9, -4), (-0.8, -2), (-0.6, -2), (-0.6, -0.7), (-0.2, 0.5), (-0.2, 3.1), (0.1, 2.9), (0.2, 4.8). The y-intercept is no longer necessarily 0. The slope of this line is

A hint: $\min f(x,y) = \|Ax + y - b\|_2^2$ becomes $\min f(x) = \|Ax - b\|_2^2$ if y is viewed as a component of x and a column of 1s is added to matrix A. 6.5511

Question 8: (continued from the previous question) and the y-intercept is 2.9010

Question 9: (continued from the previous question) Plot the points and the line in MATLAB to verify the answer. Please attach a screenshot of your plot here.

Nonlinear Least-Squares

Reading:

Chong, An Introduction to Optimization, 2011, Chapter 9 – Sections 9.3-9.4
https://onlinelibrary-wiley-com.proxy1.library.jhu.edu/doi/book/10.1002/9781118033340

In a non-linear case, we want to minimize the norm of a non-linear vector valued function g(x).

$$\min f(x) = \|g(x)\|_2^2$$

In the linear case, g(x) was

$$g(x) = Ax - b$$

and now it is some other function, for example, sine.

Note that g(x) is a vector valued function, its value is a vector. In other words, there are n functions g_i,

$g_i(x) = x_1 \sin(a_i x_2) - b_i \;\; 1 \le i \le n$

and each of them returns a scalar value and these n scalars, concatenated together into a vector, is the value of g(x). With a linear g(x), we could write the expression for g(x) in a matrix form, but with sine, we need indices.

f(x) is

$$f(x) = \|g(x)\|_2^2 = \sum_{i=1}^{n} \left(g_i(x)\right)^2 = \sum_{i=1}^{n} \left(x_1 \sin(a_i x_2) - b_i\right)^2$$

As with the linear case, we need to find the gradient and the Hessian of f(x). The gradient is

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 1? \\ 2? \end{bmatrix}$$

Question 10: Replace the 1? with the correct expression.

- $2\sum_{i=1}^{n} \left(a_i x_1^2 \sin(a_i x_2) \cos(a_i x_2) - a_i b_i x_1 \cos(a_i x_2)\right)$
- $2\sum_{i=1}^{n} \left(a_i x_1^2 \sin(a_i x_2) \cos(a_i x_2) + a_i b_i x_1 \cos(a_i x_2)\right)$
- $2\sum_{i=1}^{n} \left(x_1 \sin^2(a_i x_2) - b_i \sin(a_i x_2)\right)$
- $2\sum_{i=1}^{n} \left(x_1 \sin^2(a_i x_2) + b_i \sin(a_i x_2)\right)$
- none of the above

Question 11: Replace the 2? with the correct expression.

- $2\sum_{i=1}^{n} \left(a_i x_1^2 \sin(a_i x_2) \cos(a_i x_2) - a_i b_i x_1 \cos(a_i x_2)\right)$
- $2\sum_{i=1}^{n} \left(a_i x_1^2 \sin(a_i x_2) \cos(a_i x_2) + a_i b_i x_1 \cos(a_i x_2)\right)$

- $2\sum_{i=1}^{n}\left(x_1 \sin^2(a_i x_2) - b_i \sin(a_i x_2)\right)$

- $2\sum_{i=1}^{n}\left(x_1 \sin^2(a_i x_2) + b_i \sin(a_i x_2)\right)$

- none of the above

The Hessian is going to be even messier than the gradient, so let's go back and try something different,

$$\nabla f(x) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2\sum_{i=1}^{n} g_i(x)\dfrac{\partial g_i}{\partial x_1} \\ 2\sum_{i=1}^{n} g_i(x)\dfrac{\partial g_i}{\partial x_2} \end{bmatrix}$$

The Hessian is

$$\nabla^2 f(x) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2\sum_{i=1}^{n}\left(\dfrac{\partial g_i}{\partial x_1}\dfrac{\partial g_i}{\partial x_1} + 1?\right) & 2\sum_{i=1}^{n}\left(\dfrac{\partial g_i}{\partial x_1}\dfrac{\partial g_i}{\partial x_2} + 2?\right) \\ 2\sum_{i=1}^{n}\left(\dfrac{\partial g_i}{\partial x_1}\dfrac{\partial g_i}{\partial x_2} + ?\right) & 2\sum_{i=1}^{n}\left(\dfrac{\partial g_i}{\partial x_2}\dfrac{\partial g_i}{\partial x_2} + ?\right) \end{bmatrix}$$

Question 12: Replace the 1? with the correct expression.

- $\dfrac{\partial^2 g_i}{\partial x_1^2}$

- $g_i(x)\dfrac{\partial^2 g_i}{\partial x_1^2}$

- $\left(g_i(x)\right)^2 \dfrac{\partial^2 g_i}{\partial x_1^2}$

- none of the above

Question 13: Replace the 2? with the correct expression.

- $\dfrac{\partial^2 g_i}{\partial x_1 \partial x_2}$

- $g_i(x)\dfrac{\partial^2 g_i}{\partial x_1 \partial x_2}$

- $\left(g_i(x)\right)^2 \dfrac{\partial^2 g_i}{\partial x_1 \partial x_2}$

- none of the above

If the second partial derivatives of $g_i(x)$ are assumed to be negligibly small (the Gauss-Newton approximation) then the Hessian is

$$\nabla^2 f(x) = 2\begin{bmatrix} \sum_{i=1}^{n} \frac{\partial g_i}{\partial x_1}\frac{\partial g_i}{\partial x_1} & \sum_{i=1}^{n} \frac{\partial g_i}{\partial x_1}\frac{\partial g_i}{\partial x_2} \\ \sum_{i=1}^{n} \frac{\partial g_i}{\partial x_1}\frac{\partial g_i}{\partial x_2} & \sum_{i=1}^{n} \frac{\partial g_i}{\partial x_2}\frac{\partial g_i}{\partial x_2} \end{bmatrix}$$

or

$$\nabla^2 f(x) = 2\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_2}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_1} \\ \frac{\partial g_1}{\partial x_2} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_2} \end{bmatrix}\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \\ \cdots & \cdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} \end{bmatrix} = 2J(x)^T J(x)$$

where $J(x)$ is the Jacobian of $g(x)$.

The gradient $\nabla f(x)$ can be expressed using the Jacobian of $g(x)$ as well,

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2\sum_{i=1}^{n} g_i(x)\frac{\partial g_i}{\partial x_1} \\ 2\sum_{i=1}^{n} g_i(x)\frac{\partial g_i}{\partial x_2} \end{bmatrix} = 2\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_2}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_1} \\ \frac{\partial g_1}{\partial x_2} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_2} \end{bmatrix}\begin{bmatrix} g_1(x) \\ g_2(x) \\ \cdots \\ g_n(x) \end{bmatrix} = 2J(x)^T g(x)$$

Applying the Newton method,

$$\nabla^2 f(x^{(k+1)})(x^{(k+1)} - x^{(k)}) = -\nabla f(x^{(k)})$$

we get

$$2J(x)^T J(x)(x^{(k+1)} - x^{(k)}) = -2J(x)^T g(x)$$

or

$$x^{(k+1)} = x^{(k)} - \left(J(x)^T J(x)\right)^{-1} J(x)^T g(x)$$

Finally, $g_i(x) = x_1 \sin(a_i x_2) - b_i$ and so the Jacobian of $g(x)$ in our example is

$$J(x) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \\ \cdots & \cdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 1?(a_1 x_2) & a_1 x_1 2?(a_1 x_2) \\ 1?(a_2 x_2) & a_2 x_1 2?(a_2 x_2) \\ \cdots & \cdots \\ 1?(a_n x_2) & a_n x_1 2?(a_n x_2) \end{bmatrix}$$

Question 14: Replace the 1? with the correct expression.
- sin
- -sin
- cos
- -cos
- none of the above

Question 15: Replace the 2? with the correct expression.
- sin
- -sin
- cos
- -cos
- none of the above

Question 16: Plot the contours f(x)=1, f(x)=5, f(x)=10, f(x)=25 and f(x)=50 in MATLAB where $f(x) = \|g(x)\|_2^2 = \sum_{i=i}^{10} \left(g_i(x)\right)^2$, $x = (x_1, x_2)$ $g_i(x) = x_1 \sin(a_i x_2) - b_i$ $1 \le i \le 10$, a = (-4.7, -2.9, -1.3, -1, -0.3, 1.1, 1.5, 2.8, 3.8, 4.7), b = (1.6, 1.8, 1.6, 0.5, -0.2, -0.8, -1.5, -1.8, -1.7, -1.7). Please attach a screenshot of your plot here.

Question 17: (continued from the previous question) Is f(x) convex?
- yes
- no

Question 18: (continued from the previous question) Based on the contour plot, how many optimal solutions does min f(x) have?

The answer to question 17 indicates that f(x) may have local minimums and we want to find a global minimum. Therefore, Newton method won't necessarily converge to a global minimum. One can use the modifications described in chapters 9.2 and 9.3 (a line search and the Levenberg-Marquardt modification) to improve the chances of arriving at an optimal solution and also, as the number of points in the dataset increases, f(x) becomes smoother. For the purpose of this exercise, to ensure convergence, we will manually select a starting point that is close to the optimal solution.

Question 19: (continued from the previous question) Use the contour plot to select a starting point $(k_1^{(0)}, k_2^{(0)})$ that is located inside the $f(k_1, k_2) = 1$ contour with a positive $k_1^{(0)}$. $k_1^{(0)}$ =?
Note: we renamed x into k here because we will need x (and y) to denote something else further on.
Question 20: (continued from the previous question) $k_2^{(0)}$ =?
Question 21: (continued from the previous question) What is the value of the objective function at the starting point?

Question 22: Use the Gauss-Newton method to fit a $y = k_1 \sin(k_2 x)$ curve to points (-4.7, 1.6), (-2.9, 1.8), (-1.3, 1.6), (-1, 0.5), (-0.3, -0.2), (1.1, -0.8), (1.5, -1.5), (2.8, -1.8), (3.8, -1.7), (4.7, -1.7) using the least squares method. Use the answers to questions 19 and 20 as the starting point. The value of the objective function f(k) at the optimal solution is

A hint: it should be smaller than the answer to the previous question.

Question 23: (continued from the previous question) The fitted parameter k_1^* =

A hint: the optimal solution should be located inside a f(k)=1 contour.

Question 24: (continued from the previous question) k_2^* =

Question 25: (continued from the previous question) Plot the points and the $y = k_1 \sin(k_2 x)$ function with the fitted parameters in MATLAB to verify the answer. Please attach a screenshot of your plot here.

Question 26: If you'd like to be eligible for partial credit, please attach a file with your detailed work (preferably in the pdf format) here.