1. The Rosenbrock function is

\[ f(x, y) = 100(y - x^2)^2 + (1 - x)^2 \]

(a) Compute the gradient and Hessian of \( f(x, y) \).

(b) Show that that \( f(x, y) \) has zero gradient at the point \((1, 1)\).

(c) By considering the Hessian matrix at \((x, y) = (1, 1)\), show that this point is a minimum.

(a) Gradient and Hessian

\[ \nabla f = \begin{pmatrix} 400x^3 - 400xy + 2x - 2 \\ 200(y - x^2) \end{pmatrix} \]

\[ H = \begin{bmatrix} 1200x^2 - 400y + 2 & -400x \\ -400x & 200 \end{bmatrix} \]

(b) gradient at the point \((1, 1)\)

\[ \nabla f = \begin{pmatrix} 400x^3 - 400xy + 2x - 2 \\ 200(y - x^2) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \]

(b) Hessian at the point \((1, 1)\)

\[ H = \begin{bmatrix} 1200x^2 - 400y + 2 & -400x \\ -400x & 200 \end{bmatrix} = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix} \]

Examine eigenvalues:
- det is positive, so eigenvalues have same sign (thus not saddle point)
- trace is positive, so eigenvalues are positive
- Thus a minimum
- \( \lambda_1 = 1001.6006, \lambda_2 = 0.39936077 \)
2. In Newton type minimization schemes the update step is of the form
\[ \delta x = -H^{-1}g \]
where \( g = \nabla f \). By considering \( g.\delta x \) compare convergence of:
(a) Newton, to
(b) Gauss Newton
for a general function \( f(x) \) (i.e. where \( H \) may not be positive definite).

A note on positive definite matrices

An \( n \times n \) symmetric matrix \( M \) is **positive definite** if
- \( x^\top M x > 0 \) for all non-zero vectors \( x \)
- All the eigen-values of \( M \) are positive

In each case consider \( df = g.\delta x \). This should be negative for convergence.
(a) Newton
\[ g.\delta x = -g^\top H^{-1}g \]
Can be positive if \( H \) **not** positive definite.
(b) Gauss Newton
\[ g.\delta x = -g^\top (2J^\top J)^{-1}g \]
Non positive, since \( J^\top J \) is positive definite.
3. Explain how you could use the Gauss Newton method to solve a set of simultaneous non-linear equations.

Square the non-linear equations and add them – the resulting cost is then a sum of squared residuals, and so has a structure suitable for the Gauss Newton method.

For example, the set of equations:

\[ g_1(x, y) = 0 \]
\[ g_2(x, y) = 0 \]

can be solved for \( x = (x, y) \) by the following optimization problem which has the required sum of squares form

\[ \min_x f(x) = g_1(x)^2 + g_2(x)^2 \]
4. Sketch the feasible regions defined by the following inequalities and comment on the possible optimal values.

(a)

\[-x_1 + x_2 \geq 2\]
\[x_1 + x_2 \leq 1\]
\[x_1 \geq 0\]
\[x_2 \geq 0\]

(b)

\[2x_1 - x_2 \geq 2\]
\[x_1 \leq 4\]
\[x_1 \geq 0\]
\[x_2 \geq 0\]

There is no feasible region, and therefore no possible solutions.
The feasible region is unbounded, but the optimum can still be bounded for some cost functions.
5. More on linear programming.

(a) Show that the optimization
\[ \min_{x} \sum_{i} |a_i^\top x - b_i| \]
where the vectors \( a_i \) and scalars \( b_i \) are given, can be formulated as a linear programming problem.

(b) Solve the following linear programming problem using Matlab:

\[
\begin{align*}
\max_{x_1, x_2} & \quad 40x_1 + 88x_2 \\
\text{subject to} & \quad 2x_1 + 8x_2 \leq 60 \\
& \quad 5x_1 + 2x_2 \leq 60 \\
& \quad x_1 \geq 0 \\
& \quad x_2 \geq 0
\end{align*}
\]

(a) The absolute value operator \(| . |\) is not linear, so at first sight this does not look like a linear programming problem. However, it can be transformed into one by adding extra variables and constraints. Introduce additional variables \( \alpha_i \) with the constraints for each \( i \) that \( |a_i^\top x - b_i| \leq \alpha_i \). This can be written as the two linear constraints:

\[
\begin{align*}
a_i^\top x - b_i & \leq \alpha_i \\
a_i^\top x - b_i & \geq -\alpha_i
\end{align*}
\]
or equivalently

\[
\begin{align*}
a_i^\top x - b_i & \leq \alpha_i \\
b_i - a_i^\top x & \leq \alpha_i
\end{align*}
\]

Then the linear programming problem

\[ \min_{x, \alpha_i} \sum_{i} \alpha_i \]

subject to these constraints, optimizes the original problem.
(b)

\[
\begin{align*}
\max_{x_1, x_2} & \quad 40x_1 + 88x_2 \\
\text{subject to} & \quad 2x_1 + 8x_2 \leq 60 \\
& \quad 5x_1 + 2x_2 \leq 60 \\
& \quad x_1 \geq 0 \\
& \quad x_2 \geq 0
\end{align*}
\]

Matlab code

```matlab
f = [-40; -88];
A = [2 8
    5 2];
b = [60; 60];
lb = zeros(2,1);
x = linprog(f,A,b,[],[],lb);

solution is x_1 = 10.0000, x_2 = 5.0000.
```
6. Interior point method using a barrier function. Show that the following 1D problem

\[
\begin{align*}
\text{minimize} & \quad f(x) = x^2, x \in \mathbb{R} \\
\text{subject to} & \quad x - 1 \geq 0 
\end{align*}
\]

can be reformulated using a logarithmic barrier method as

\[
\begin{align*}
\text{minimize} & \quad x^2 - r \log(x - 1)
\end{align*}
\]

Determine the solution (as a function of \( r \)), and show that the global optimum is obtained as \( r \to 0 \).

We need to find the optimum of

\[
\min_x B(x, r) = x^2 - r \log(x - 1)
\]

Differentiating wrt \( x \) gives

\[
2x - \frac{r}{x - 1} = 0
\]

and rearranging gives

\[
2x^2 - 2x - r = 0
\]

Use the standard formula for a quadratic equation

\[
x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}
\]

to obtain

\[
x = \frac{2 \pm \sqrt{4 + 8r}}{4} = \frac{1}{2} \pm \frac{1}{2} \sqrt{1 + 2r}
\]

and since only \( x > 1 \) is admissible (due to the log)

\[
x^*(r) = \frac{1}{2} + \frac{1}{2} \sqrt{1 + 2r}
\]

and as \( r \to 0, x \to 1 \), which is the global optimum.
7. Mean and median estimates. For a set of measurements \( \{a_i\} \), show that

(a) \[ \min_x \sum_i (x - a_i)^2 \]

is the mean of \( \{a_i\} \).

(b) \[ \min_x \sum_i |x - a_i| \]

is the median of \( \{a_i\} \).

(a)

\[ \min_x \sum_i \frac{N}{i} (x - a_i)^2 \]

To find the minimum, differentiate \( f(x) \) wrt \( x \), and set to zero:

\[ \frac{df(x)}{dx} = \sum_i 2(x - a_i) = 0 \]

and rearranging

\[ \sum_i x = \sum_i a_i \]

and so

\[ x = \frac{1}{N} \sum_i a_i \]

i.e. the mean of \( \{a_i\} \).

(b)

\[ \min_x f(x) = \sum_i |x - a_i| \]

- The derivative of \( |x - a_i| \) wrt \( x \) is +1 when \( x > a_i \) and \(-1 \) when \( x < a_i \).
- The derivative of \( f(x) \) is zero when there are as many values of \( a_i \) less than \( x \) as there are greater than \( x \).
- Thus \( f(x) \) minimized at the median of values of \( \{a_i\} \).

Note, the median is immune to changes in \( a_i \) that lie far from the median – the value of the cost function changes, but not the position of the median.
8. Determine in each case if the following functions are convex:

(a) The sum of quadratic functions \( f(x) = a_1(x - b_1)^2 + a_2(x - b_2)^2 \), for \( a_i > 0 \)
(b) The piecewise linear function \( f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i) \)
(c) \( f(x) = \max\{x, 1/x\} \)
(d) \( f(x) = ||Ax - b||^2 \)

(a) The sum of quadratic functions \( f(x) = a_1(x - b_1)^2 + a_2(x - b_2)^2 \), for \( a_i > 0 \)
Consider expanding the two quadratics, then the coefficient of \( x^2 \) is \( a_1 + a_2 \). Using the second derivative test for convexity:
\[
\frac{d^2 f}{dx^2} \geq 0
\]
then the sum is convex provided that \( a_1 + a_2 \geq 0 \). So the function is convex since \( a_i > 0 \).

(b) The piecewise linear function \( f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i) \)
This is convex. Most easily seen by drawing and applying the geometric chord test for convexity.
(c) \( f(x) = \max\{x, 1/x\} \) for \( x > 0 \)

This is convex. Most easily seen by drawing and applying the geometric chord test for convexity.

(d) \( f(x) = \|Ax - b\|^2 \)

Compute the Hessian

\[
\nabla f(x) = 2A^\top(Ax - b) = 2A^\top Ax - 2A^\top b
\]

\[
H = 2A^\top A
\]

Hence \( H \) is positive definite (because \( A^\top A \) is positive definite), and so the function is convex.