(1) Let $\gamma \in (0,1)$.
  (a) Show that the sequence $\{\gamma^\nu\}$ converges linearly to zero, but not superlinearly.
  (b) Show that the sequence $\{\gamma^{\nu^2}\}$ converges superlinearly to 0, but not quadratically.
  (c) Finally, show that the sequence $\{\gamma^{2^\nu}\}$ converges quadratically to zero.

(2) Apply Newton's method to minimize the function $f(x) = x^2 + \cos x$ with initial point $x^0 = 2\pi$.

(3) Apply the secant method to minimize the function $f(x) = x^2 + e^x$ with starting points $x_{-1} = 1 - 10^{-3}$ and $x_0 = 1$. How many ierations does it take to obtain a point for which $|f'(x^k)| \le 10^-8$? Compare this performance with that of Newton's method and the method of steepest descent.

(4) Let $H = M^{-1}$ where $M \in \mathbb{R}^{n\times n}$ is symmetric and positive definite. Let $s, y \in \mathbb{R}^n$ be such that $s^T y > 0$. Define

$$\overline{M} = M + \frac{yy^T}{y^T s} - \frac{Mss^T M}{s^T M s}.$$

Show that

$$\overline{H} = H + \frac{(s + Hy)^T y s s^T}{(s^T y)^2} - \frac{Hys^T + sy^T H}{s^T y}$$

satisfies $\overline{M}^{-1} = \overline{H}$.
Hint: Multiply them together and show that you get the identity matrix.

(5) Let $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{k\times k}$, and $U, V \in \mathbb{R}^{n\times k}$. If the matrices A, C and $(C^{-1} + V^T A^{-1} U)$ are all invertible, show that $(A + UCV^T)$ is also invertible by showing that its inverse is given by

$$(A + UCV^T)^{-1} = A^{-1} - A^{-1}U(C^{-1} + V^T A^{-1} U)^{-1} V^T A^{-1} \ .$$

Hint: Just multiply it out and show that you get the identity.

(6) Show that every rank 1 matrix W in $\mathbb{R}^{n\times n}$ can be written in the form $W = uv^T$ for some choice of non-zero vectors $u, v \in \mathbb{R}^n$.

(7) Show that every symmetric $n\times n$ positive semi-definite matrix M can be written in the form $M = VV^T$ for some $n\times n$ matrix V having the same rank as M. Also, show that V can be chosen to be an $n \times k$ matrix where $k = \text{rank}(M)$.

(8) Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable with $F'(x)$ invertible on $\mathbb{R}^n$. The Newton iteration for solving $F(x) = 0$ is given by

$$x^{k+1} = x^k - F'(x^k)^{-1} F(x^k) \ .$$

Although this iteration converges at a quadratic rate locally, it may not converge if the initial point x^0 is too far from a solution. To compensate for this deficiency, the Newton iteration is often replaced by a *damped* Newton iteration of the form

$$x^{k+1} = x^k - t_k F'(x^k)^{-1} F(x^k)$$

for some choice of $t_k > 0$. We consider one approach to choosing t_k.
  (a) Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = \|F(x)\|_2$. Show that if $x \in \mathbb{R}^n$ is such that $F(x) \neq 0$, then f is differentiable at x with

$$\nabla f(x) = F'(x)^T \frac{F(x)}{f(x)} \ .$$

  (b) Note that one can attempt to solve $F(x) = 0$ by minimizing the function $f(x)$. This indicates that a damped Newton step size $t_k > 0$ can be computed by doing a line search on the function f in the Newton direction $d_N = -F'(x)^{-1} F(x)$. As a first step, show that the Newton direction d_N is a direction of descent at points x for which $F(x) \neq 0$ by showing that

$$f'(x; d_N) = -\|F(x)\|_2 \ .$$

  (c) Show that the backtracking line search applies and takes the form

$$\begin{array}{rll} t_k = & \max & \gamma^s \\ & \text{s.t.} & s \in \{0, 1, \dots\} \\ & & \|F(x^k + \gamma^s d_N^k)\|_2 \le (1 - c\gamma^s)\|F(x^k)\|_2 \end{array} ,$$

where $\gamma \in (0,1)$ and $c \in (0,1)$ are the line search parameters.

(9) Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable and consider the problem of finding a point x such that $F(x) = 0$. This is the same problem that led to the development of Newton's method except now we are interested in the case where $n \neq m$. If $m < n$, then there are most likely infinitely many solutions. On the other-hand, if $m > n$, then there are probably no solutions. In this problem we consider the case when $m > n$. Since it is unlikely that there exists an x satisfying $F(x) = 0$, we instead try to find an x that makes $\|F(x)\|$ as small as possible. For this we define $f(x) = \frac{1}{2}\|F(x)\|_2^2$. It is easily shown (try it) that

$$\begin{aligned} \nabla f(x) &= F'(x)^T F(x) \\ \nabla^2 f(x) &= F'(x)^T F'(x) + \sum_{i=1}^{m} F_i(x) \nabla^2 F_i(x) \ , \end{aligned}$$

where the functions $F_i : \mathbb{R}^n \to \mathbb{R}$ are the component functions of F.

(a) In this setting we still try to follow the path suggested by Newton where we successively solve for the search direction using the linearization of F. But in this case the search direction is given as the solution to the linear least squares problem

$$\mathcal{GN} \qquad \min_{d \in \mathbb{R}^n} \frac{1}{2}\|F(x^k) + F'(x^k)d\|_2^2 \ .$$

Methods using the solution to this problem as a search direction are called *Gauss-Newton* methods. Denote a solution by d_{GN}. Under what conditions does this problem have a unique solution?

(b) When a unique solution to the problem $\mathcal{GN}$ exists give a formula for d_{GN}.

(c) Use this formula for d_{GN} to show that it is a descent direction for f giving a formula for $f'(x; d_{GN})$.

(d) Show that d_{GN} is also a descent direction for the function $h(x) = \|F(x)\|_2$.

(e) Assuming that $F'(x)$ is Lipschitz continuous, show that

$$h'(x; d) \leq \|F(x) + F'(x)d\|_2 - h(x) \ .$$