1. Suppose the conditional gradient ("Frank-Wolfe") algorithm as described in the notes is applied with the following modification: The subproblem is solved inexactly at each iteration, but is required to attain at least half the potential decrease in the linearized function that would be obtained from an exact solution. Specifically, we require \( \bar{x}_k \) to be feasible, and to satisfy 
\[
\nabla f(x_k)^T(\bar{x}_k - x_k) \leq \frac{1}{2} \min_{z \in \Omega} \nabla f(x_k)^T(z - x_k).
\]

Apart from this modification, the complete algorithm is the same as in “Algorithm 2” from the class notes. By modifying the analysis given in the notes, and using the same notation as there, show that this inexact conditional gradient algorithm converges sublinearly with the following rate expression:
\[
f(x_k) - f(x^*) \leq \frac{8LD^2}{k + 8}.
\]

2. Consider the inequality constrained nonlinear optimization problem
\[
\min_{x} f(x) \text{ s.t. } c_i(x) \geq 0, \quad i = 1, 2, \ldots, m,
\]
where \( f \) and all \( c_i \) are smooth. Suppose that LICQ, KKT, and second-order sufficient conditions are satisfied at the point \( x^* \), with active set \( \mathcal{A}^* = \{i = 1, 2, \ldots, m : c_i(x^*) = 0\} \). Show that LICQ, KKT, and second-order sufficient conditions are satisfied at \( (x^*, s^*) \) for \( s^* = c(x^*) \) for the slack-variable formulation
\[
\min_{(x,s)} f(x) \text{ s.t. } c(x) - s = 0, \quad s \geq 0,
\]
where \( c(x) \) and \( s \) are the vectors in \( \mathbb{R}^m \) whose components are \( c_i(x) \) and \( s_i, \ i = 1, 2, \ldots, m, \) respectively.

3. Consider the equality constrained nonlinear optimization problem:
\[
\min_{x} f(x) \text{ s.t. } c_i(x) = 0, \quad i = 1, 2, \ldots, m.
\]
Suppose that \( x^* \) satisfies the KKT conditions and second-order sufficient conditions, but that the LICQ is not satisfied at \( x^* \).
(a) Can we guarantee that the SQP method will converge to $x^*$ when started from a point $x^0$ that is close to $x^*$?

(b) Can we guarantee that the quadratic penalty function

$$P_\mu(x) := f(x) + \frac{\mu}{2} \sum_{i=1}^{m} c_i^2(x)$$

will have a minimizer close to $x^*$, provided that $\mu$ is sufficiently large?

Explain your answers.

4. Suppose you have an algorithm for nonlinear programming that generates a sequence of primal-dual points $(x^k, \lambda^k)$ that converges to $(x^*, \lambda^*)$.

(a) Suppose that we can establish the following bound on the change in error from one iteration to the next.

$$\left\| \begin{bmatrix} x^{k+1} - x^* \\ \lambda^{k+1} - \lambda^* \end{bmatrix} \right\| \leq \eta \left\| x^k - x^* \right\|^{3/2},$$

for some constant $\eta > 0$.

(i) What can you say about the rate of convergence of $\{x^k\}$ to $x^*$? Is it Q-superlinear? R-superlinear? Linear?

(ii) What can you say about the rate of convergence of $\{\lambda^k\}$ to $\lambda^*$? Is it Q-superlinear? R-superlinear? Linear?

(b) Now suppose that we can establish the following bound on the change in error from one iteration to the next.

$$\|x^{k+1} - x^*\| \leq \eta \left\| \begin{bmatrix} x^k - x^* \\ \lambda^k - \lambda^* \end{bmatrix} \right\|^{3/2},$$

for some $\eta > 0$.

(i) What can you say about the rate of convergence to $\{x^k\}$ to $x^*$ in this case? What about the rate of convergence of $\{\lambda^k\}$ to $\lambda^*$?

(ii) Suppose that in addition to the bound given in part (b), you know that $\|\lambda^k - \lambda^*\| \leq \beta \|x^k - x^*\|$ for some $\beta > 0$. What can you say about the rates of convergence of $\{x^k\}$ to $x^*$ and $\{\lambda^k\}$ to $\lambda^*$ in this case?