Caleb Logemann
MATH 565 Continuous Optimization
Homework 5

1. Page 269: Problem 10.1
   Let J be an $m \times n$ matrix $m \geq n$.

   (a) Show that J has full column rank if and only if $J^T J$ is nonsingular.

   Proof. Suppose that J has full column rank, then J has a full singular value decomposition. That is $J = U\Sigma V^T$, where U is an orthogonal $m \times m$ matrix, V is an orthogonal $n \times n$ matrix and Σ is a diagonal $m \times n$ matrix. Since J is full column rank this implies that the diagonal of Σ is nonzero. Now using this decomposition $J^T J$ can be written as

   $$J^T J = (U\Sigma V^T)^T (U\Sigma V^T) = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T \Sigma V^T.$$

   Note that $\Sigma^T \Sigma$ is a $n \times n$ matrix, and it is diagonal as it is the product of diagonal matrices. Also the diagonal of $\Sigma^T \Sigma$ is nonzero, as the diagonal of Σ is nonzero. Thus this shows that $J^T J$ can be diagonalized as $V\Sigma^T \Sigma V^T$ and that the eigenvalues are nonzero. Therefore $J^T J$ is nonsingular as the eigenvalues are nonzero.
   Now suppose that $J^T J$ is nonsingular. □

   (b) Show that J has full column rank if and only if $J^T J$ is positive definite.

   Proof. Suppose that $J^T J$ is positive definite. This implies that the eigenvalues of $J^T J$ are all strictly greater than zero. Since all of the eigenvalues are strictly positive, this implies that $\det(J^T J) > 0$ or equivalently that $J^T J$ is nonsingular. Now by part (a) this implies that J has full column rank.
   Suppose on the other hand J has full column rank. Now by part (a) this implies that $J^T J$ is nonsingular. Since $J^T J$ is nonsingular, the eigenvalues of $J^T J$ are nonzero. As in part (a) it was shown that $J^T J$ could be diagonalized as $V\Sigma^T \Sigma V^T$. Considering the diagonal of $\Sigma^T \Sigma$, we see that the eigenvalues of $J^T J$ are the squares of the singular values of J. Since squares are all nonnegative, this shows that the eigenvalues of $J^T J$ are nonnegative. This fact along with the fact that the eigenvalues are nonzero implies that $J^T J$ has only positive eigenvalues or equivalently that $J^T J$ is positive definite. □

2. Page 269: Problem 10.5
   Suppose that each residual function r_j and its gradient are Lipschitz continuous with Lipschitz constant L, that is,

   $$|r_j(x) - r_j(x^*)| \leq L\|x - x^*\|, \quad \|\nabla r_j(x) - \nabla r_j(x^*)\| \leq L\|x - x^*\|$$

   for all $j = 1, 2, \ldots, m$ and all $x, x^* \in D$, where D is a compact subset of $\mathbb{R}^n$. Assume also that the r_j are bounded on D, that is, there exists $M > 0$ such that $|r_j(x)| \leq M$ for all $j = 1, 2, \ldots, m$ and all $x \in D$. Find Lipschitz constants for the Jacobian J (10.3) and the gradient ∇f (10.4) over D.

   First I will suppose that all norms are the 2-norm or the matrix norm induced by the vector 2-norm. First I will consider the Jacobian.

   $$\begin{aligned}\|J(x) - J(x^*)\| &= \sup_{\|y\|=1} \left(\|(J(x) - J(x^*))y\|\right) \\ &= \sup_{\|y\|=1} \left(\sqrt{\sum_{j=1}^{m} \left((\nabla r_j(x) - \nabla r_j(x^*))^T y\right)^2}\right)\end{aligned}$$

Using the Cauchy-Schwarz inequality

$$
\begin{aligned}
&\leq \sup_{\|y\|=1}\left(\sqrt{\sum_{j=1}^{m}\left(\|\nabla r_j(x)-\nabla r_j(x^*)\|\,\|y\|\right)^2}\right)\\
&=\sqrt{\sum_{j=1}^{m}\|\nabla r_j(x)-\nabla r_j(x^*)\|^2}
\end{aligned}
$$

Using the Lipschitz continuity of the gradient

$$
\begin{aligned}
&\leq \sqrt{\sum_{j=1}^{m}L^2\|x-x^*\|^2}\\
&=\sqrt{mL^2\|x-x^*\|^2}\\
&=\sqrt{m}L\|x-x^*\|
\end{aligned}
$$

Therefore a Lipschitz constant for the Jacobian is $\sqrt{m}L$.

Now I will consider the gradient ∇f.

3. Consider the underdetermined linear system $Jx = r$, where $J \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$, $r \in \mathbb{R}^m$, and $m < n$ (i.e. there are less equations than unknowns). Assume that the rank of J is m (i.e., it has full rank). There will exist infinitely many solutions. The minimum norm solution of $Jx = r$ is the solution closest to the origin, which may be regarded as the solution of the constrained optimization problem:

$$\min_{x\in\mathbb{R}^n} \|x\|^2 \quad \text{subject to} \quad Jx = r.$$

   (a) Use the Lagrange multiplier method, derive the solution to this optimization problem:

$$x = J^T(JJ^T)^{-1}r$$

Proof. The Lagrange multiplier method states that the solution to this minimization problem satisfies the following equations,

$$
\begin{aligned}
\nabla_x\mathcal{L}(x,\lambda) &= 0,\\
c_i(x) &= 0, \forall i \in \mathcal{E},\\
c_i(x) &\geq 0, \forall i \in \mathcal{I},\\
\lambda_i &\geq, \forall i \in \mathcal{I},\\
\lambda_i c_i(x) &= 0, \forall i \in \mathcal{E}\cup\mathcal{I},
\end{aligned}
$$

where $\mathcal{L}$ is the Lagrangian. In this problem the Lagrangian is

$$\mathcal{L}(x,\lambda) = \|x\|^2 - \lambda^T(Jx-r).$$

There are only equality constraints which are

$$Jx - r = 0$$

The full set of equations for this problem is thus

$$
\begin{aligned}
\nabla_x(\|x\|^2 - \lambda^T(Jx-r)) &= 0,\\
Jx - r &= 0,\\
\lambda^T(Jx-r) &= 0,
\end{aligned}
$$

□

(b) Find the minimum norm solution of the 3×5 system $Jx = r$, where

$$J = \begin{pmatrix} 1 & 2 & 0 & 3 & 2 \\ -1 & -1 & 4 & 2 & 0 \\ 3 & -2 & 2 & 1 & 1 \end{pmatrix}, \quad r = \begin{pmatrix} 4 \\ 1 \\ -7 \end{pmatrix}.$$

4. An important problem in signal processing amounts to finding parameters $c_1, c_2, \ldots, c_n$ and $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that

$$\sum_{k=1}^{n} \left(c_k e^{-\lambda_k t} \right) \approx f(t),$$

for a given signal function $f(t)$. One approach for solving this problem is to formulate a nonlinear least squares problems. Let

$$x := (x_1, x_2, \ldots, x_{2n}) = (c_1, \ldots, c_n, \lambda_1, \ldots, \lambda_n)$$

be the vector of parameters to be determined. Let $s_j = f(t_j)$ be given samples of f for $j = 1, \ldots, 2n$ and set

$$r_j(x) = \sum_{k=1}^{n} \left(c_k e^{-\lambda_k t_j} \right) - s_j = \sum_{k=1}^{n} \left(x_k e^{-x_{n+k} t_j} \right) - s_j.$$

We then obtain the parameters as the solution of the nonlinear least squares problem:

$$\min_{x \in \mathbb{R}^{2n}} \|r(x)\|^2.$$

(a) Find the general expression for $J(x)$.
The i, j entry of the Jacobian is defined as

$$J(x)_{ij} = \frac{\partial r_j}{\partial x_i}$$

This value can be computed using the following definition of r_j for this problem,

$$r_j(x) = \sum_{k=1}^{n} \left(c_k e^{-\lambda_k t_j} \right) - s_j = \sum_{k=1}^{n} \left(x_k e^{-x_{n+k} t_j} \right) - s_j.$$

The entries are

$$J(x)_{ij} = \frac{\partial r_j}{\partial x_i}$$

$$\frac{\partial *}{\partial \sum_{k=1}^{n} \left(x_k e^{-x_{n+k} t_j} \right) - s_j} x_i$$

If $i \leq n$, then there exists $k = i$, so

$$J(x)_{ij} = e^{-x_{n+i} t_j}$$

If $i > n$, then there exists k such that $i = n + k$, so

$$J(x)_{ij} = -t_j x_{i-n} e^{-x_i t_j}$$

Thus the general definition of the Jacobian by entry is

$$J(x)_{ij} = \begin{cases} e^{-x_{n+i} t_j} & i \leq n \\ -t_j x_{i-n} e^{-x_i t_j} & i > n \end{cases}$$

where $i, j = 1, 2, \ldots, 2n$.

(b) Let $n = 2$, and

$$t = (0.0, 0.3, 0.6, 0.9) \quad \text{and} \quad s = (2.700, 1.480, 0.819, 0.458).$$

In MATLAB or PYTHON, program a Gauss-Newton iteration scheme for this problem. Apply the scheme the following initial guess:

$$x_0 = (1, 1, 1, 2)$$

and run until convergence.

5. Page 352: Problem 12.4
If f is convex and the feasible region Ω is convex, show that local solutions of the problem (12.3) are also global solutions. Show that the set of global solutions is convex. (Hint: See Theorem 2.5.)

Proof. Let f be a convex function and let Ω, the feasible region, be convex. Suppose that x is a local minimum of f, but not a global minimum This implies that there exists $x^*\Omega$, a global minimum, such that $f(x^*) < f(x)$. Since Ω is convex this implies that $tx + (1-t)x^* \in \Omega$ for all $t \in [0, 1]$. Also since f is convex,

$$\begin{aligned} f(tx + (1-t)x^*) &\leq tf(x) + (1-t)f(x^*) \\ &\leq tf(x) + (1-t)f(x) \\ &= f(x) \end{aligned}$$

Now for any local neighborhood, N, of x, there exists a $t \in [0, 1]$ such that $tx + (1-t)x^* \in N$ and $f(tx + (1-t)x^*) < f(x)$. This directly contradicts the fact that x is a local minumum, therefore if x is a local minimum it must also be a global minimum. ☐

6. Page 353: Problem 12.13
Show that for the feasible region defined by

$$\begin{aligned} (x_1 - 1)^2 + (x_2 - 1)^2 &\leq 2, \\ (x_1 - 1)^2 + (x_2 + 1)^2 &\leq 2, \\ x_1 &\geq 0, \end{aligned}$$

the MFCG is satisfied at $x^* = (0, 0)^T$ but the LICQ is not satisfied.

Proof. First the MFCQ require the the gradients of the constraints by computed. Let

$$\begin{aligned} c_1(x) &= 2 - (x_1 - 1)^2 - (x_2 - 1)^2 \\ c_2(x) &= 2 - (x_1 - 1)^2 - (x_2 + 1)^2 \\ c_3(x) &= x_1 \end{aligned}$$

where $\mathcal{I} = \{1, 2, 3\}$ and $\mathcal{E} = \varnothing$. Then the gradients of the constraints are

$$\begin{aligned} \nabla c_1(x) &= \begin{bmatrix} -2(x_1 - 1) \\ -2(x_2 - 1) \end{bmatrix} \\ \nabla c_2(x) &= \begin{bmatrix} -2(x_1 - 1) \\ -2(x_2 + 1) \end{bmatrix} \\ \nabla c_3(x) &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \end{aligned}$$

At the point $x^* = (0, 0)$, these gradients are

$$\nabla c_1(x^*) = \begin{bmatrix} -2(-1) \\ -2(-1) \end{bmatrix}$$
$$\nabla c_2(x^*) = \begin{bmatrix} -2(-1) \\ -2(1) \end{bmatrix}$$
$$\nabla c_3(x^*) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

First note that $\{\nabla c_i(x^*), i \in \mathcal{E}\}$ is empty as there are no equality constraints. Therefore this set is trivially linearly independent.

At the point $x^* = (0, 0)$, all of the inequality constraints are active. Consider the vector $w = []^T$.

$\square$

7. Page 354: Problem 12.18
Consider the problem of finding the point on the parabola $y = \frac{1}{5}(x-1)^2$ that is closest to $(x, y) = (1, 2)$, in the Euclidean norm sense. We can formulate this problem as

$$\min f(x, y) = (x-1)^2 + (y-2)^2 \quad \text{subject to } (x-1)^2 = 5y.$$

   (a) Find all the KKT points for this problem. Is the LICQ satisfied?
   (b) Which of these points are the solutions?
   (c) By directly substituting the constraint into the object function and eliminating the variable x, we obtain an unconstrained optimization problem. Show that the solutions of this problem cannot be solutions of the original problem.

8. Page 354: Problem 12.21
Find the maxima of $f(x) = x_1 x_2$ over the unit disk defined by the inequality constraint $1 - x_1^2 - x_2^2 \geq 0$.