13 An introduction to nonlinear problems.

13.1 The standard form.

The standard form for a nonlinear problem is in these notes given by

\[
\begin{align*}
\max \ f_0(x) & \text{ when } \ f_f(x) \leq 0 \\
& \vdots \\
& f_p(x) \leq 0
\end{align*}
\]

and \( f_r : \mathbb{R}^n \to \mathbb{R} \) for all \( r \in \{0, \ldots, P\} \). The standard form covers equality constraints by splitting each equality into two inequalities. In what follows it is also assumed that \( \nabla f_r : \mathbb{R}^n \to \mathbb{R}^n \) exists for all \( r \in \{0, \ldots, P\} \), where

\[
\nabla f_r(x) = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1}(x) \\
\vdots \\
\frac{\partial f_n}{\partial x_n}(x)
\end{bmatrix}.
\]

A function is said to be smooth if all the partial derivatives of all orders exist. It is assumed that all functions are smooth.

Example:

The standard primal problem in the standard form is given by

\[
P = n + m
\]
\[
f_0(x) = c^T x = c_1 x_1 + \cdots + c_n x_n
\]
\[
f_j(x) = a_1^j x_1 + \cdots + a_n^j x_n - b_j \quad \text{for } j \in \{1, \ldots, m\}
\]
\[
f_{m+i}(x) = -x_i \quad \text{for } i \in \{1, \ldots, n\}
\]

13.2 The Kuhn-Tucker conditions.

The Kuhn-Tucker conditions characterize a ‘multiplier’ \( \hat{\lambda} \in \mathbb{R}^P \) at a feasible global maximum \( \hat{x} \in \mathbb{R}^n \) by:

1. \( \hat{\lambda} \geq 0 \)
2. \( \hat{\lambda}_r f_r(\hat{x}) = 0 \) for all \( r \in \{1, \ldots, P\} \)
3. \( \nabla f_0(\hat{x}) = \sum_{r=1}^{P} \hat{\lambda}_r \nabla f_r(\hat{x}) \)

Example:

In the case of the standard primal problem let \( \hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_m, \hat{\mu}_1, \ldots, \hat{\mu}_n) \), where \( \hat{\mu}_i = \hat{\lambda}_{m+i} \) for all \( i \in \{1, \ldots, n\} \).
Condition (3) becomes
\[
\begin{bmatrix}
    c_1 \\
    \vdots \\
    c_n
\end{bmatrix}
= \begin{bmatrix}
    a_1^1 \\
    \vdots \\
    a_m^1 \\
\end{bmatrix} \lambda_1 + \cdots + \begin{bmatrix}
    a_1^m \\
    \vdots \\
    a_m^m \\
\end{bmatrix} \lambda_m 
- \begin{bmatrix}
    \hat{\mu}_1 \\
    \vdots \\
    \hat{\mu}_n
\end{bmatrix}
\]
Since \( f_r(\hat{x}) \leq 0 \) for all \( r \in \{1, \ldots, n\} \), it follows that \( \hat{x} \) is feasible in the primal problem. When the matrix on the left is compared with the matrix on the right, it is seen that \( \hat{\mu}_j \) is equal to the slack in the \( j \)th constraint in the dual problem. When this is combined with the fact that \( \hat{\lambda} \geq 0 \) it follows that \( (\hat{\lambda}_1, \ldots, \hat{\lambda}_m) \) is feasible in the dual problem. The second Kuhn-Tucker condition corresponds to the complementary slackness conditions of the primal and dual linear programming problems. Note that
\[
c^T \hat{x} = \hat{x}^T c = \begin{bmatrix}
    c_1 \\
    \vdots \\
    c_n
\end{bmatrix}
= \begin{bmatrix}
    a_1^1 \\
    \vdots \\
    a_m^1 \\
\end{bmatrix} \hat{\lambda}_1 \hat{x} + \cdots + \begin{bmatrix}
    a_1^m \\
    \vdots \\
    a_m^m \\
\end{bmatrix} \hat{\lambda}_m \hat{x} - \hat{x}^T \begin{bmatrix}
    \hat{\mu}_1 \\
    \vdots \\
    \hat{\mu}_n
\end{bmatrix}
\]
and
\[
\hat{x}^T \begin{bmatrix}
    \hat{\mu}_1 \\
    \vdots \\
    \hat{\mu}_n
\end{bmatrix}
= 0 \iff \hat{\lambda}_j \hat{x} = \hat{\lambda}_j b_j
\]
follow from the second Kuhn-Tucker condition. The conclusion is that
\[
c^T \hat{x} = \begin{bmatrix}
    \hat{\lambda}_1 \\
    \vdots \\
    \hat{\lambda}_m
\end{bmatrix}
\]
and hence both \( \hat{x} \) and \( (\hat{\lambda}_1, \ldots, \hat{\lambda}_m) \) are optimal.

**Example:**
Let \( C = \{ x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1 \} \) be the standard unit disk. Consider the function \( f_0 : C \rightarrow \mathbb{R} \) given by \( f_0(x_1, x_2) = x_1 + 2x_2 \). To find the global maximum, let \( f_1(x_1, x_2) = x_1^2 + x_2^2 - 1 \) and analyze the Kuhn-Tucker conditions. Search for \( \hat{\lambda} \in \mathbb{R} \) and \( \hat{x} = (\hat{x}_1, \hat{x}_2) \in \mathbb{R}^2 \) such that
\[
\begin{align*}
(1) \quad & \hat{\lambda} \geq 0 \\
(2) \quad & \hat{\lambda}(\hat{x}_1^2 + \hat{x}_2^2 - 1) = 0 \\
(3) \quad & \begin{bmatrix}
1 \\
2
\end{bmatrix}
= \begin{bmatrix}
\frac{2\hat{x}_1}{\sqrt{\lambda}} \\
\frac{2\hat{x}_2}{\sqrt{\lambda}}
\end{bmatrix}
\end{align*}
\]
From (1) and (3) it follows that \( \hat{\lambda} > 0 \). Now (2) implies that \( \hat{x}_1^2 + \hat{x}_2^2 = 1 \), and (3) implies \( \frac{1}{\lambda} = 2\hat{x}_1 \) and \( \frac{1}{\lambda} = 2\hat{x}_2 \). Hence \( \hat{x}_2 = 2\hat{x}_1 \), and \( \hat{x}_1^2 + (2\hat{x}_1)^2 = 1 \). From this only two possibilities remain: \( \hat{x}_a = \left( \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right) \) and \( \hat{x}_b = \left( -\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}} \right) \). The value of the function is given by \( f_0(\hat{x}_a) = \sqrt{5} \). 

13-2
and \( f_0(\hat{x}) = -\sqrt{5} \). Using the level curves of \( f_0 : \mathbb{R}^2 \to \mathbb{R} \), it is seen that \( \hat{x} \) is a global maximum.

### 13.3 The Lagrange multiplier rule.

In the case with exactly \( m \) equality constraints the Kuhn-Tucker conditions simplify. Each equality constraint \( f_r(x) = 0 \) is split as \( f_r(x) \leq 0 \) and \( -f_r(x) \leq 0 \). The corresponding multipliers should be combined as \( \hat{\eta}_r = \hat{\lambda}_{2r-1} - \hat{\lambda}_{2r} \). The third Kuhn-Tucker condition has the form \( \nabla f_0(\hat{x}) = \sum_{r=1}^{m} \hat{\eta}_r \nabla f_r(\hat{x}) \). The second condition is automatically satisfied. The first condition is satisfied by choosing \( \hat{\lambda}_{2r-1} = \hat{\eta}_r, \hat{\lambda}_{2r} = 0 \) if \( \hat{\eta}_r \geq 0 \) and \( \hat{\lambda}_{2r-1} = 0, \hat{\lambda}_{2r} = -\hat{\eta}_r \) otherwise. Observe that the gradient rule is the familiar Langrange multiplier rule.

### 13.4 Failing Kuhn-Tucker conditions.

It is always possible to write down the Kuhn-Tucker conditions provided the gradients \( \nabla f_r \) exist for all \( r \in \{0, \ldots, P\} \). It is not always true that the Kuhn-Tucker conditions are satisfied at a global maximum.

**Example:**

Let \( C_c = \{ x \in \mathbb{R}^2 \mid x_2^2 + c \leq x_1^3 \} \) where \( c \in \mathbb{R} \) is some given fixed constant. Consider the function \( f_0 : C_c \to \mathbb{R} \) given by \( f_0(x_1, x_2) = -x_1 \). To find the global maximum, let \( f_0(x_1, x_2) = -x_1^3 + x_2^2 + c \) and analyze the Kuhn-Tucker conditions. Note the resemblance between this example and the previous example. Search for \( \hat{\lambda} \in \mathbb{R} \) and \( \hat{x} = (\hat{x}_1, \hat{x}_2) \in \mathbb{R}^2 \) such that

\[
\begin{align*}
(1) \quad & \hat{\lambda} \geq 0 \\
(2) \quad & \hat{\lambda}(-\hat{x}_1^3 + \hat{x}_2^2 + c) = 0 \\
(3) \quad & \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -3\hat{x}_1^2 \\ 2\hat{x}_2 \end{bmatrix}
\end{align*}
\]

From (1) and (3) it follows that \( \hat{\lambda} > 0 \). Now (2) implies that \( \hat{x}_2^2 + c = \hat{x}_1^3 \), and (3) implies \( \hat{x}_2 = 0 \). If \( c \neq 0 \), then \( \hat{x}_1 = c^{1/3} \) and \( \hat{\lambda} = \frac{1}{3c^{2/3}} > 0 \). Note that \( c \leq x_1^3 \) for all feasible \( (x_1, x_2) \), so it follows that \( \hat{x} = (c^{1/3}, 0) \) is a global maximum when \( c \neq 0 \).

The case \( c = 0 \) is remarkably different. This time \( \hat{x}_1 = c^{1/3} = 0 \), which contradicts (3). Hence there is no pair \( \hat{\lambda} \in \mathbb{R} \) and \( \hat{x} = (\hat{x}_1, \hat{x}_2) \in \mathbb{R}^2 \) such that all three Kuhn-Tucker conditions are satisfied. Since \( 0 \leq x_2^2 \leq x_1^3 \) for all feasible \( (x_1, x_2) \), it is nonetheless true that \( \hat{x} = (c^{1/3}, 0) = (0, 0) \) is a global maximum.
13.5 Admissible directions.

Suppose the problem is to minimize a smooth function \( f : C \to \mathbb{R} \) where \( C \subset \mathbb{R}^n \). Given a point \( \hat{x} \in C \) define the set of admissible directions \( v \in \mathbb{R}^n \), \( v \neq 0 \), at \( \hat{x} \) by requiring the existence of a smooth curve \( \gamma_v^\hat{x} : [0, +\infty) \to C \) such that \( \gamma_v^\hat{x}(0) = \hat{x} \), and the right-hand derivative satisfies \( (\gamma_v^\hat{x})'(0) = v \). Denote the set of admissible directions at \( \hat{x} \in C \) by \( A_\hat{x} \). If \( \hat{x} \in C \) is a global minimum, and \( v \in A_\hat{x} \), then \( f(\gamma_v^\hat{x}(t)) - f(\hat{x}) \geq 0 \) for all \( t \in [0, \infty) \). For \( t > 0 \) it follows that \( \frac{f(\gamma_v^\hat{x}(t)) - f(\hat{x})}{t} \geq 0 \).

Since both \( f : C \to \mathbb{R} \) and \( \gamma_v^\hat{x} : [0, \infty) \to C \) are assumed smooth, an application of the chain rule shows that their composition is smooth. Take the right-hand limit, \( t \to 0^+ \), of the difference quotient and conclude that \( \nabla f(\hat{x}) \cdot v \geq 0 \). Note that if both \( v \in A_\hat{x} \) and \( -v \in A_{\hat{x}} \), then \( \nabla f(\hat{x}) = 0 \).

Example:

Suppose \( \hat{x} = (1, 0) \in \mathbb{R}^2 \) and \( C = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\} \). In this case \( A_\hat{x} = \{(v_1, v_2) \in \mathbb{R}^2 \mid v_1 = 0, v_2 = 0\} \).

Given an element in \( A_\hat{x} \), one possible curve is given by \( \gamma_v^\hat{x}(t) = (\cos(v_2 t), \sin(v_2 t)) \).

13.6 Admissible directions in the linearized problem.

Consider again the standard nonlinear problem. Suppose \( \hat{x} \in C \) so that \( f_r(\hat{x}) \leq 0 \) for all \( r \in \{1, \ldots, P\} \). If \( f_r(\hat{x}) = 0 \), then the constraint \( f_r(\hat{x}) \leq 0 \) is said to be active. A direction \( v \in \mathbb{R}^n \), \( v \neq 0 \) is said to be admissible in the linearized problem if \( \nabla f_r(\hat{x}) \cdot v \leq 0 \) for each active constraint \( f_r(\hat{x}) \leq 0 \). Let \( B_\hat{x} \) denote the set of admissible directions in the linearized problem at \( \hat{x} \in C \).

Example:

Suppose \( \hat{x} = (1, 0) \in \mathbb{R}^2 \) and \( C = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1\} \). In this case \( B_\hat{x} = \{(v_1, v_2) \in \mathbb{R}^2 \mid v_1 \leq 0, v_1^2 + v_2^2 > 0\} \).

because \( \nabla f(\hat{x}) = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \).

13.7 Necessary Kuhn-Tucker conditions.

It is now possible to state, without proofs, two circumstances when the Kuhn-Tucker conditions must be satisfied at a global maximum. The two circumstances guarantee that the Kuhn-Tucker conditions are satisfied. The converse is not in general true. It is also possible that the Kuhn-Tucker conditions hold despite that neither of the two circumstances are satisfied.
13.7.1 Regularity.

Suppose \( \hat{x} \in C \) is a global maximum of standard nonlinear problem where all functions are assumed to be smooth. Let

\[
\Gamma = \left\{ r \in \{1, \ldots, P\} \mid f_r(\hat{x}) = 0, \text{ the first index of the functions associated with a splitting of an equality constraint} \right\}
\]

If the collection of gradients \( \{\nabla f_r(\hat{x}) \mid f_r(\hat{x}) = 0, r \in \Gamma\} \) is linearly independent, then the Kuhn-Tucker conditions must be satisfied. Note that the linear independence is only required for the gradients of active constraints.

Example:
The example with failing Kuhn-Tucker conditions has one constraint \( f_1(x, y) = y^2 - x^3 \). The gradient is given by

\[
\nabla f_1(x, y) = \begin{bmatrix} -3x^2 \\ 2y \end{bmatrix}.
\]

With \((\hat{x}, \hat{y}) = (0, 0)\) the gradient is zero. A collection containing a zero vector is not linearly independent. When \( c \neq 0 \), regularity is restored since the gradient no longer is zero.

13.7.2 Constraint Qualification.

Suppose \( \hat{x} \in C \) is a global maximum of standard nonlinear problem where all functions are assumed to be smooth. If \( A_{\hat{x}} = B_{\hat{x}} \), then the Kuhn-Tucker conditions must be satisfied.

Example:
The example with failing Kuhn-Tucker conditions has one constraint \( f_1(x, y) = y^2 - x^3 \). The gradient is given by

\[
\nabla f_1(x, y) = \begin{bmatrix} -3x^2 \\ 2y \end{bmatrix}.
\]

With \((\hat{x}, \hat{y}) = (0, 0)\) the gradient is zero. If follows that \( B_{\hat{x}} \) is the set of all nonzero vectors in \( \mathbb{R}^2 \).

A curve in \( C \), which starts at \((\hat{x}, \hat{y}) = (0, 0)\), has an initial tangent vector aiming into the right-hand half-plane, hence \( A_{\hat{x}} = B_{\hat{x}} \).

Finally, the case \( c \neq 0 \) is handled using regularity.