Second-order optimality conditions for equality-constrained optimization

Mark S. Gockenbach

1 The second-order necessary condition

I now derive second-order optimality conditions for the equality-constrained nonlinear program

$$\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad g(x) = 0,
\end{align*}$$

under the assumption that $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable. I will assume that $x^* \in \mathbb{R}^n$ is a local minimizer of (1) and that $x : [-a,a] \to \mathbb{R}^n$ is a feasible path with $x(0) = x^*$. Defining $\phi : [-a,a] \to \mathbb{R}$ by

$$\phi(t) = f(x(t)),$$

it follows that $\phi$ has a local minimizer at $t = 0$, which implies that $\phi''(0) \geq 0$. Using the chain rule, I can compute the derivatives of $\phi$ as follows:

$$\begin{align*}
\phi(t) = f(x(t)) & \Rightarrow \phi'(t) = \nabla f(x(t)) \cdot \dot{x}(t) \\
& \Rightarrow \phi''(t) = \dot{x}(t) \cdot \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) \cdot \ddot{x}(t) \\
& \Rightarrow \phi''(0) = \dot{x}(0) \cdot \nabla^2 f(x^*) \dot{x}(0) + \nabla f(x^*) \cdot \ddot{x}(0).
\end{align*}$$

I can eliminate the reference to $\ddot{x}(0)$ by using the Lagrange multipliers and the fact that $x$ is a feasible path. Assuming $\lambda^*$ is a Lagrange multiplier associated with $x^*$, then differentiating twice via the chain rule yields

$$\begin{align*}
\lambda^* \cdot g(x(t)) = 0 & \Rightarrow \sum_{i=1}^m \lambda_i^* g_i(x(t)) = 0 \\
& \Rightarrow \sum_{i=1}^m \sum_{j=1}^n \lambda_i^* \frac{\partial g_i}{\partial x_j}(x(t)) \dot{x}_j(t) = 0 \\
& \Rightarrow \sum_{i=1}^m \sum_{j=1}^n \left\{ \lambda_i^* \frac{\partial g_i}{\partial x_j}(x(t)) \ddot{x}_j(t) + \sum_{k=1}^n \lambda_k^* \frac{\partial^2 g_i}{\partial x_k \partial x_j}(x(t)) \dot{x}_j(t) \dot{x}_k(t) \right\} = 0 \\
& \Rightarrow (\nabla g(x(t)) \lambda^*) \cdot \ddot{x}(t) + \sum_{i=1}^m \dot{x}(t) \cdot (\lambda_i^* \nabla^2 g_i(x(t))) \dot{x}(t) = 0 \\
& \Rightarrow (\nabla g(x(t)) \lambda^*) \cdot \ddot{x}(t) + \dot{x}(t) \cdot \left( \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x(t)) \right) \dot{x}(t) = 0.
\end{align*}$$

Therefore,

$$\begin{align*}
(\nabla g(x^*) \lambda^*) \cdot \ddot{x}(0) + \dot{x}(0) \cdot \left( \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^*) \right) \dot{x}(0) = 0.
\end{align*}$$
But
\[ \nabla g(x^*) \lambda^* = \nabla f(x^*), \]
so
\[ \nabla f(x^*) \cdot \dot{x}(0) = -\dot{x}(0) \cdot \left( \sum_{i=1}^{m} \lambda_i^* \nabla^2 g_i(x^*) \right) \dot{x}(0). \]
Substituting this expression for \( \nabla f(x^*) \cdot \dot{x}(0) \) into the formula for \( \phi''(0) \) yields
\[ \phi''(0) = \dot{x}(0) \cdot \left( \nabla^2 f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla^2 g_i(x^*) \right) \dot{x}(0). \]
The reader will recall that, under the constraint qualifications presented earlier, \( \dot{x}(0) \) is an arbitrary member of \( \mathcal{N}(\nabla g(x^*)^T) \). It follows that \( \phi''(0) \geq 0 \) if and only if the matrix
\[ \nabla^2 f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla^2 g_i(x^*) \]
is positive semidefinite on the subspace \( \mathcal{N}(\nabla g(x^*)^T) \). To be precise, the second-order necessary condition for \( x^* \) to be a local minimizer of (1) is
\[ z \cdot \left( \nabla^2 f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla^2 g_i(x^*) \right) z \geq 0 \text{ for all } z \in \mathcal{N}(\nabla g(x^*)^T). \]  

2 The Lagrangian

The first- and second-order necessary conditions can be conveniently expressed in terms of the so-called Lagrangian function.

**Definition 2.1** The Lagrangian of the nonlinear program
\[
\min_{x} \quad f(x) \\
\text{s.t.} \quad g(x) = 0
\]
is the function \( \ell : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \) defined by
\[
\ell(x; \lambda) = f(x) - \lambda \cdot g(x) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x).
\]
The gradient of \( \ell \) (with respect to \( x^1 \)) is given by
\[ \nabla \ell(x; \lambda) = \nabla f(x) - \nabla g(x) \lambda. \]
Therefore, the first-order necessary condition for \( x^* \) to be a local minimizer (or maximizer) of (1) is
\[ \text{There exists } \lambda^* \in \mathbb{R}^m \text{ such that } \nabla \ell(x^*; \lambda^*) = 0. \]

\footnote{All gradients and Hessians will be taken with respect to \( x \) only, not \( \lambda \), unless I specifically indicate otherwise.}
The Hessian of $\ell$ is
$$\nabla^2 \ell(x; \lambda) = \nabla^2 f(x) - \sum_{i=1}^{m} \lambda_i \nabla^2 g_i(x),$$
which is the matrix appearing in the second-order necessary condition (2). Therefore (2) can be equivalently expressed as
$$z \cdot \nabla^2 \ell(x^*; \lambda^*) z \geq 0 \text{ for all } z \in \mathcal{N}(\nabla g(x^*)^T). \quad (3)$$
It should be noted that the Hessian of the Lagrangian need not be positive semidefinite, that is, (3) need not hold for all $z$, but rather only for $z \in \mathcal{N}(\nabla g(x^*)^T)$. The Lagrangian need not have a minimizer at $x^*$, but rather only a stationary point. In many cases, it is possible to decrease $f$ by moving away from $x^*$ in an infeasible direction.

3 The second-order sufficient condition

If $x^*$ is a feasible point of (1) and there exists $\lambda^* \in \mathbb{R}^m$ such that $\nabla \ell(x^*; \lambda^*) = 0$, and if $\nabla^2 \ell(x^*; \lambda^*)$ is actually positive definite on the subspace $\mathcal{N}(\nabla g(x^*)^T)$, then $x^*$ is a strict local minimizer of (1). To be precise, the following conditions imply that $x^*$ is a strict local minimizer:

$$\begin{align*}
g(x^*) &= 0, \\
\nabla \ell(x^*; \lambda^*) &= 0, \\
z \cdot \nabla^2 \ell(x^*; \lambda^*) z &> 0 \text{ for all } z \in \mathcal{N}(\nabla g(x^*)^T), \ z \neq 0.
\end{align*} \quad (4)$$

I will not prove this result here, as the analysis is somewhat subtle.²

4 Convex programs

The reader will notice that there are no conditions that are both necessary and sufficient for $x^*$ to be a local minimizer of the general equality-constrained nonlinear program. As in the case of unconstrained minimization, necessary and sufficient conditions exist for the special case in which the problem is convex.

I will begin by proving the following general theorem, which applies also to inequality-constrained convex programs and, in fact, to unconstrained problems with a convex objective function.

**Theorem 4.1** Suppose $f : C \to \mathbb{R}$ is convex, where $C$ is a convex set. If $x^* \in C$ is a local minimizer of $f$ over the set $C$, that is, a local solution to the problem

$$\min_{x \in C} f(x) \ \text{s.t.} \ x \in C,$$

then $x^*$ is in fact a global minimizer of $f$ over $C$.

**Proof:** Assume $x^*$ is a local minimizer of $f$ over $C$ and suppose, by way of contradiction, that $y \in C$ satisfies $f(y) < f(x^*)$. By the convexity of $C$,

$$(1 - \alpha)x^* + \alpha y \in C \text{ for all } \alpha \in [0, 1],$$

and, by the convexity of $f$,

$$f((1 - \alpha)x^* + \alpha y) \leq (1 - \alpha)f(x^*) + \alpha f(y) \text{ for all } \alpha \in [0, 1] \Rightarrow f(x^*) + \alpha(f(y) - f(x^*)) < f(x^*) \text{ for all } \alpha \in (0, 1].$$

²For the proof, see Tapia [1].
But this last inequality contradicts the assumption that \( x^* \) is a local minimizer of \( f \) over \( C \). QED

Since the feasible set defined by the linear constraints \( Ax = b \), where \( A \in \mathbb{R}^{m \times n} \) and \( b \in \mathbb{R}^m \), is a convex set, the preceding theorem applies to the linearly-constrained NLP

\[
\begin{align*}
\min \quad & f(x) \\
\text{s.t.} \quad & Ax = b,
\end{align*}
\]

assuming that \( f \) is convex. Such an NLP is referred to as a (equality-constrained) convex program.

I can prove directly that the first-order optimality conditions are sufficient for \( x^* \) to be a global solution of a convex program. (In proving the following result, I do not use the previous theorem, which I presented for its own sake.)

**Theorem 4.2** Suppose \( f : \mathbb{R}^n \to \mathbb{R} \) is convex and continuously differentiable and let \( A \in \mathbb{R}^{m \times n} \), \( b \in \mathbb{R}^m \) be given. If \( x^* \in \mathbb{R}^n \) and \( \lambda^* \in \mathbb{R}^m \) satisfy

\[
\begin{align*}
Ax^* &= b, \\
\nabla f(x^*) &= A^T \lambda^*,
\end{align*}
\]

then \( x^* \) is a local (and hence global) minimizer of the convex program

\[
\begin{align*}
\min \quad & f(x) \\
\text{s.t.} \quad & Ax = b.
\end{align*}
\]

**Proof:** It is easy to show that the Lagrangian

\[
\ell(x; \lambda^*) = f(x) - \lambda^* \cdot (Ax - b)
\]

is a convex function of \( x \). By assumption, \( \nabla \ell(x^*; \lambda^*) = 0 \), which implies by an earlier theorem that \( \ell(\cdot; \lambda^*) \) has its global minimizer at \( x = x^* \). Since \( \ell(x; \lambda^*) = f(x) \) for all \( x \) in the feasible set, this shows in particular that \( x^* \) is the global minimizer of \( f \) subject to the constraint \( Ax = b \). QED

The proof of the previous theorem shows that, if the Lagrange multiplier were known, the convex program

\[
\begin{align*}
\min \quad & f(x) \\
\text{s.t.} \quad & Ax = b
\end{align*}
\]

could be replaced by the unconstrained problem

\[
\min_x \ell(x; \lambda^*).
\]

I want to emphasize that this is not true for a nonconvex problem, even if only local solutions are considered.

**Example 4.3** I define \( f : \mathbb{R}^2 \to \mathbb{R} \) and \( g : \mathbb{R}^2 \to \mathbb{R} \) by

\[
\begin{align*}
f(x) &= x_1^2 - x_2^2 + x_2, \\
g(x) &= x_2,
\end{align*}
\]

and analyze the NLP

\[
\begin{align*}
\min \quad & f(x) \\
\text{s.t.} \quad & g(x) = 0.
\end{align*}
\]
The reader will notice that $f$ is not convex. The first-order optimality conditions result in the system of equations

$$
\begin{bmatrix}
2x_1 \\
-2x_2 + 1
\end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \lambda, \\
x_2 = 0.
$$

It is easy to see that the solution is $x^* = 0$, $\lambda^* = 1$ and therefore

$$
\ell(x; \lambda^*) = f(x) - g(x) = x_1^2 - x_2^2,
$$

which has a saddle point (not a minimizer) at $x^*$. However, $x^*$ is a local (in fact, global) minimizer of $f$ subject to the constraint, since, for all feasible $x$, $f(x) = x_1^2$.

References