Notes for Part 1: Optimality conditions and why they are important

Nick Gould, CSED, RAL, Chilton, OX11 0QX, England (n.gould@rl.ac.uk)

January 11, 2006

1 Sketches of proofs for Part 1

Theorems 1.1—1.3 can be found in any good book on analysis. Theorems 1.1 and 1.2 follow directly by considering the remainders of truncated Taylor expansions of the univariate function $f(x + \alpha s)$ with $\alpha \in [0, 1]$, while Theorem 1.3 uses the Newton formula

$$F(x+s) = F(x) + \int_0^1 \nabla_x F(x + \alpha s) s d\alpha.$$

Proof of Farkas' lemma

The result is trivial if $\mathcal{C} = 0$. So otherwise, suppose that $g \in \mathcal{C}$ and that $s^T a_i \geq 0$ for $i \in \mathcal{A}$. Then

$$s^T g = \sum_{i \in \mathcal{A}} y_i s^T a_i \geq 0.$$

Hence $\mathcal{S}$ is empty, since $s^T g$ is non-negative.

Conversely, suppose that $g \notin \mathcal{C}$, and consider

$$\min_{c \in \mathcal{C}} \|g - c\|_2 = \min_{c \in \bar{\mathcal{C}}} \|g - c\|_2,$$

where

$$\bar{\mathcal{C}} = \mathcal{C} \bigcap \{c \mid \|g - c\|_2 \leq \|g - \bar{c}\|_2\}$$

and $\bar{c}$ is any point in $\mathcal{C}$. Since $\mathcal{C}$ is closed, and $\{c \mid \|g - c\|_2 \leq \|g - \bar{c}\|_2\}$ is compact, $\bar{C}$ is non-empty and compact, and it follows from Weierstrass' Theorem (namely, that the minimizer of continuous function within a compact set is achieved) that

$$c_* = \arg \min_{c \in \mathcal{C}} \|g - c\|_2$$

exists. As $\mathcal{C}$ is convex with 0, $c_* \in \mathcal{C}$, $\alpha c_* \in \mathcal{C}$ for all $\alpha \geq 0$, and hence $\phi(\alpha) = \|g - \alpha c_*\|_2^2$ is minimized at $\alpha = 1$. Hence $\phi'(1) = 0$ and thus

$$c_*^T (c_* - g) = 0. \tag{1.1}$$

By convexity, if $c \in \mathcal{C}$, so is $c_* + \theta(c - c_*)$ for all $\theta \in [0, 1]$, and hence by optimality of c_*

$$\|g - c_*\|_2^2 \leq \|g - c_* + \theta(c_* - c)\|_2^2.$$

Expanding and taking the limit as θ approaches zero, we deduce that

$$0 \leq (g - c_*)^T (c_* - c) = (c_* - g)^T c$$

using (1.1). Thus, defining $s = c_* - g$, $s^T c \geq 0$ for all $c \in \mathcal{C}$, and in particular $s^T a_i \geq 0$ for all $i \in \mathcal{A}$. But as $s \neq 0$, as $c_* \in \mathcal{C}$ and $g \notin \mathcal{C}$, and $s^T g = -s^T s < 0$, using (1.1), we have exhibited the separating hyperplane $s^T v = 0$ as required when $g \notin \mathcal{C}$.

1.1 Proof of Theorem 1.4

Suppose otherwise, that $g(x_*) \neq 0$. A Taylor expansion in the direction $-g(x_*)$ gives

$$f(x_* - \alpha g(x_*)) = f(x_*) - \alpha \|g(x_*)\|^2 + O(\alpha^2).$$

For sufficiently small α, $\frac{1}{2}\alpha \|g(x_*)\|^2 \geq O(\alpha^2)$, and thus

$$f(x_* - \alpha g(x_*)) \leq f(x_*) - \tfrac{1}{2}\alpha \|g(x_*)\|^2 < f(x_*).$$

This contradicts the hypothesis that x_* is a local minimizer.

1.2 Proof of Theorem 1.5

Again, suppose otherwise that $s^T H(x_*) s < 0$. A Taylor expansion in the direction s gives

$$f(x_* + \alpha s) = f(x_*) + \tfrac{1}{2}\alpha^2 s^T H(x_*) s + O(\alpha^3),$$

since $g(x_*) = 0$. For sufficiently small α, $-\frac{1}{4}\alpha^2 s^T H(x_*) s \geq O(\alpha^3)$, and thus

$$f(x_* + \alpha s) \leq f(x_*) + \tfrac{1}{4}\alpha^2 s^T H(x_*) s < f(x_*).$$

Once again, this contradicts the hypothesis that x_* is a local minimizer.

1.3 Proof of Theorem 1.6

By continuity $H(x)$ is positive definite for all x in a open ball $\mathcal{N}$ around x_*. The generalized mean value theorem then says that if $x_* + s \in \mathcal{N}$, there is a value z between the points x_* and $x_* + s$ for which

$$f(x_* + s) = f(x_*) + g(x_*)^T s + \tfrac{1}{2} s^T H(z) s = f(x_*) + \tfrac{1}{2} s^T H(z) s > f(x_*)$$

for all nonzero s, and thus x_* is an isolated local minimizer.

1.4 Proof of Theorem 1.7

We consider feasible perturbations about x_*. Consider a vector valued C^2 (C^3 for Theorem 1.8) function $x(\alpha)$ of the scalar α for which $x(0) = x_*$ and $c(x(\alpha)) = 0$. (The constraint qualification is that all such feasible perturbations are of this form). We may then write

$$x(\alpha) = x_* + \alpha s + \tfrac{1}{2}\alpha^2 p + O(\alpha^3) \tag{1.2}$$

and we require that

$$
\begin{aligned}
0 &= c_i(x(\alpha)) = c(x_* + \alpha s + \tfrac{1}{2}\alpha^2 p + O(\alpha^3)) \\
&= c_i(x_*) + a_i^T(x_*)\,(\alpha s + \tfrac{1}{2}\alpha^2 p) + \tfrac{1}{2}\alpha^2 s^T H_i(x_*) s + O(\alpha^3) \\
&= \alpha a_i^T(x_*) s + \tfrac{1}{2}\alpha^2 \left(a_i^T(x_*) p + s^T H_i(x_*) s \right) + O(\alpha^3)
\end{aligned}
$$

using Taylor's theorem. Matching similar asymptotic terms, this implies that for such a feasible perturbation

$$A(x_*) s = 0 \tag{1.3}$$

and

$$a_i^T(x_*) p + s^T H_i(x_*) s = 0 \tag{1.4}$$

for all $i = 1, \ldots, m$. Now consider the objective function

$$
\begin{aligned}
f(x(\alpha)) &= f(x_* + \alpha s + \tfrac{1}{2}\alpha^2 p + O(\alpha^3)) \\
&= f(x_*) + g(x_*)^T\,(\alpha s + \tfrac{1}{2}\alpha^2 p) + \tfrac{1}{2}\alpha^2 s^T H(x_*) s + O(\alpha^3) \\
&= f(x_*) + \alpha g(x_*)^T s + \tfrac{1}{2}\alpha^2 \left(g(x_*)^T p + s^T H(x_*) s \right) + O(\alpha^3)
\end{aligned}
\tag{1.5}
$$

This function is unconstrained along $x(\alpha)$, so we may deduce, as in Theorem 1.4, that

$$g(x_*)^T s = 0 \text{ for all } s \text{ such that } A(x_*) s = 0. \tag{1.6}$$

If we let S be a basis for the null-space of $A(x_*)$, we may write

$$g(x_*) = A^T(x_*) y_* + S z_* \tag{1.7}$$

for some y_* and z_*. Since, by definition, $A(x_*) S = 0$, and as it then follows from (1.6) that $g^T(x_*) S = 0$, we have that

$$0 = S^T g(x_*) = S^T A^T(x_*) y_* + S^T S z_* = S^T S z_*.$$

Hence $S^T S z_* = 0$ and thus $z_* = 0$ since S is of full rank. Thus (1.7) gives

$$g(x_*) - A^T(x_*) y_* = 0. \tag{1.8}$$

Proof of Theorem 1.8

We have shown that

$$f(x(\alpha)) = f(x_*) + \tfrac{1}{2}\alpha^2 \left(p^T g(x_*) + s^T H(x_*) s \right) + O(\alpha^3) \tag{1.9}$$

for all s satisfying $A(x_*) s = 0$, and that (1.8) holds. Hence, necessarily,

$$p^T g(x_*) + s^T H(x_*) s \geq 0 \tag{1.10}$$

for all s and p satisfying (1.3) and (1.4). But (1.8) and (1.4) combine to give

$$p^T g(x_*) = \sum_{i=1}^{m} (y_*)_i p^T a_i(x_*) = -\sum_{i=1}^{m} (y_*)_i s^T H_i(x_*) s$$

and thus (1.10) is equivalent to

$$s^T \left(H(x_*) - \sum_{i=1}^{m} (y_*)_i H_i(x_*) \right) s \equiv s^T H(x_*, y_*) s \geq 0$$

for all s satisfying (1.3).

Proof of Theorem 1.9

As in the proof of Theorem 1.6, we consider feasible perturbations about x_*. Since any constraint that is inactive at x_* (i.e., $c_i(x_*) > 0$) will remain inactive for small perturbations, we need only consider perturbations that are constrained by the constraints active at x_*, (i.e., $c_i(x_*) = 0$). Let $\mathcal{A}$ denote the indices of the active constraints. We then consider a vector valued C^2 (C^3 for Theorem 1.10) function $x(\alpha)$ of the scalar α for which $x(0) = x_*$ and $c_i(x(\alpha)) \geq 0$ for $i \in \mathcal{A}$. In this case, assuming that $x(\alpha)$ may be expressed as (1.2), we require that

$$\begin{aligned} 0 \;\leq\; & c_i(x(\alpha)) = c(x_* + \alpha s + \tfrac{1}{2}\alpha^2 p + O(\alpha^3)) \\ = \;& c_i(x_*) + a_i(x_*)^T \alpha s + \tfrac{1}{2}\alpha^2 p + \tfrac{1}{2}\alpha^2 s^T H_i(x_*) s + O(\alpha^3) \\ = \;& \alpha a_i(x_*)^T s + \tfrac{1}{2}\alpha^2 \left(a_i(x_*)^T p + s^T H_i(x_*) s \right) + O(\alpha^3) \end{aligned}$$

for all $i \in \mathcal{A}$. Thus

$$s^T a_i(x_*) \geq 0 \tag{1.11}$$

and

$$p^T a_i(x_*) + s^T H_i(x_*) s \geq 0 \;\text{ when }\; s^T a_i(x_*) = 0 \tag{1.12}$$

for all $i \in \mathcal{A}$. The expansion of $f(x(\alpha))$ (1.5) then implies that x_* can only be a local minimizer if

$$\mathcal{S} = \{ s \mid s^T g(x_*) < 0 \;\text{ and }\; s^T a_i(x_*) \geq 0 \;\text{ for }\; i \in \mathcal{A} \} = \emptyset.$$

But then the result follows directly from Farkas' Lemma—a proof of this famous result is given, for example, as Lemma 9.2.4 in

R. Fletcher "Practical Methods of Optimization", Wiley (1987, 2nd edition).

Farkas' Lemma. Given any vectors g and a_i, $i \in \mathcal{A}$, the set

$$\mathcal{S} = \{ s \mid s^T g < 0 \;\text{ and }\; s^T a_i \geq 0 \;\text{ for }\; i \in \mathcal{A} \}$$

is empty if and only if

$$g = \sum_{i \in \mathcal{A}} y_i a_i$$

for some $y_i \geq 0$, $i \in \mathcal{A}$

Proof of Theorem 1.10

The expansion (1.5) for the change in the objective function will be dominated by the first-order term $\alpha s^T g(x_*)$ for feasible perturbations unless $s^T g(x_*) = 0$, in which case the expansion (1.9) is relevant. Thus we must have that (1.10) holds for all feasible s for which $s^T g(x_*) = 0$. The latter requirement gives that

$$0 = s^T g(x_*) = \sum_{i \in \mathcal{A}} y_i s^T a_i(x_*),$$

and hence that either $y_i = 0$ or $s^T a_i(x_*) = 0$ (or both).

We now focus on the *subset* of all feasible arcs that ensure $c_i(x(\alpha)) = 0$ if $y_i > 0$ and $c_i(x(\alpha)) \geq 0$ if $y_i = 0$ for $i \in \mathcal{A}$. For those constraints for which $c_i(x(\alpha)) = 0$, we have that (1.3) and (1.4) hold, and thus for such perturbations $s \in \mathcal{N}_+$. In this case

$$p^T g(x_*) = \sum_{i \in \mathcal{A}} y_i p^T a_i(x_*) = \sum_{\substack{i \in \mathcal{A} \\ y_i > 0}} y_i p^T a_i(x_*) = - \sum_{\substack{i \in \mathcal{A} \\ y_i > 0}} y_i s^T H_i(x_*) s = - \sum_{i \in \mathcal{A}} y_i s^T H_i(x_*) s$$

This combines with (1.10) to give that

$$s^T H(x_*, y_*) s \equiv s^T \left(H(x_*) - \sum_{i=1}^{m} (y_*)_i H_i(x_*) \right) s = p^T g(x_*) + s^T H(x_*) s \geq 0.$$

for all $s \in \mathcal{N}_+$, which is the required result.

Proof of Theorem 1.11

Consider any feasible arc $x(\alpha)$. We have seen that (1.11) and (1.12) hold, and that first-order feasible perturbations are characterized by $\mathcal{N}_+$. It then follows from (1.12) that

$$p^T g(x_*) = \sum_{i \in \mathcal{A}} y_i p^T a_i(x_*) = \sum_{\substack{i \in \mathcal{A} \\ s^T a_i(x_*) = 0}} y_i p^T a_i(x_*) \geq - \sum_{\substack{i \in \mathcal{A} \\ s^T a_i(x_*) = 0}} y_i s^T H_i(x_*) s = - \sum_{i \in \mathcal{A}} y_i s^T H_i(x_*) s,$$

and hence by assumption that

$$p^T g(x_*) + s^T H(x_*) s \geq s^T \left(H(x_*) - \sum_{i=1}^{m} (y_*)_i H_i(x_*) \right) s \equiv s^T H(x_*, y_*) s > 0$$

for all $s \in \mathcal{N}_+$. But this then combines with (1.5) and (1.11) to show that $f(x(\alpha)) > f(x_*)$ for all sufficiently small α.