

Part 1: Optimality conditions and why they are important

Nick Gould (RAL)

$$c(x) \geq 0, \quad g(x) + A^T(x)y = 0, \quad y \geq 0$$

Misc course on nonlinear optimization

OPTIMIZATION PROBLEMS

Unconstrained minimization:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

where the **objective function** $f : \mathbb{R}^n \rightarrow \mathbb{R}$

Equality constrained minimization:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \text{ subject to } c(x) = 0$$

where the **constraints** $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ($m \leq n$)

Inequality constrained minimization:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \text{ subject to } c(x) \geq 0$$

where $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$ (m may be larger than n)

NOTATION

Use the following throughout the course:

$g(x) \stackrel{\text{def}}{=} H(x)$	$\nabla_x f(x)$	gradient of f
$a_i(x) \stackrel{\text{def}}{=} H_i(x)$	$\nabla_{x_i} f(x)$	Hessian matrix of f
	$\nabla_{x_i} c_i(x)$	gradient of ith constraint
	$\nabla_{x_i} c_i(x)$	Hessian of ith constraint
$A(x) \stackrel{\text{def}}{=} \nabla_x c(x) \equiv \begin{pmatrix} a_1^T(x) \\ \dots \\ a_m^T(x) \end{pmatrix}$		Jacobian matrix of c
$\ell(x, y) \stackrel{\text{def}}{=} f(x) - y^T c(x)$		Lagrangian function, where y are Lagrange multipliers
$H(x, y) \stackrel{\text{def}}{=} \nabla_{xx} \ell(x, y) \equiv H(x) - \sum_{i=1}^m y_i H_i(x)$		Hessian of the Lagrangian

LIPSCHITZ CONTINUITY

- \mathcal{X} and \mathcal{Y} open sets
- $F : \mathcal{X} \rightarrow \mathcal{Y}$
- $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Y}}$ are norms

Then

- F is **Lipschitz continuous at $x \in \mathcal{X}$** if $\exists \gamma(x)$ such that

$$\|F(z) - F(x)\|_{\mathcal{Y}} \leq \gamma(x) \|z - x\|_{\mathcal{X}}$$

for all $z \in \mathcal{X}$.

- F is **Lipschitz continuous throughout/in \mathcal{X}** if $\exists \gamma$ such that

$$\|F(z) - F(x)\|_{\mathcal{Y}} \leq \gamma \|z - x\|_{\mathcal{X}}$$

for all x and $z \in \mathcal{X}$.

USEFUL TAYLOR APPROXIMATIONS

Theorem 1.1. Let \mathcal{S} be an open subset of \mathbb{R}^n , and suppose $f : \mathcal{S} \rightarrow \mathbb{R}$ is continuously differentiable throughout \mathcal{S} . Suppose further that $g(x)$ is Lipschitz continuous at x , with Lipschitz constant $\gamma^L(x)$ in some appropriate vector norm. Then, if the segment $x + \theta s \in \mathcal{S}$ for all $\theta \in [0, 1]$,

$$|f(x + s) - m^L(x + s)| \leq \frac{1}{2} \gamma^L(x) \|s\|^2, \quad \text{where}$$

$$m^L(x + s) = f(x) + g(x)^T s.$$

If f is twice continuously differentiable throughout \mathcal{S} and $H(x)$ is Lipschitz continuous at x , with Lipschitz constant $\gamma^Q(x)$,

$$|f(x + s) - m^Q(x + s)| \leq \frac{1}{6} \gamma^Q(x) \|s\|^3, \quad \text{where}$$
$$m^Q(x + s) = f(x) + g(x)^T s + \frac{1}{2} s^T H(x) s.$$

ANOTHER USEFUL TAYLOR APPROXIMATION

Theorem 1.3. Let \mathcal{S} be an open subset of \mathbb{R}^n , and suppose $F : \mathcal{S} \rightarrow \mathbb{R}^m$ is continuously differentiable throughout \mathcal{S} . Suppose further that $\nabla_x F(x)$ is Lipschitz continuous at x , with Lipschitz constant $\gamma^L(x)$ in some appropriate vector norm and its induced matrix norm. Then, if the segment $x + \theta s \in \mathcal{S}$ for all $\theta \in [0, 1]$,

$$\|F(x + s) - M^L(x + s)\| \leq \frac{1}{2} \gamma^L(x) \|s\|^2,$$

where

$$M^L(x + s) = F(x) + \nabla_x F(x) s.$$

MEAN VALUE THEOREM

Theorem 1.2. Let \mathcal{S} be an open subset of \mathbb{R}^n , and suppose $f : \mathcal{S} \rightarrow \mathbb{R}$ is twice continuously differentiable throughout \mathcal{S} . Suppose further that $s \neq 0$, and that the interval $[x, x + s] \in \mathcal{S}$. Then

$$f(x + s) = f(x) + g(x)^T s + \frac{1}{2} s^T H(z) s$$

for some $z \in (x, x + s)$.

OPTIMALITY CONDITIONS

Optimality conditions are useful because:

- they provide a means of guaranteeing that a candidate solution is indeed optimal (**sufficient conditions**), and
- they indicate when a point is not optimal (**necessary conditions**)

Furthermore they

- guide in the design of algorithms, since
- lack of optimality \iff indication of improvement

UNCONSTRAINED MINIMIZATION

First-order necessary optimality:

Theorem 1.4. Suppose that $f \in C^1$, and that x_* is a local minimizer of $f(x)$. Then

$$g(x_*) = 0.$$

Second-order necessary optimality:

Theorem 1.5. Suppose that $f \in C^2$, and that x_* is a local minimizer of $f(x)$. Then $g(x_*) = 0$ and $H(x_*)$ is positive semi-definite, that is

$$s^T H(x_*) s \geq 0 \text{ for all } s \in \mathbb{R}^n.$$

PROOF OF THEOREM 1.4

Suppose otherwise, that $g(x_*) \neq 0$.

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

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

This contradicts hypothesis that x_* is a local minimizer.

UNCONSTRAINED MINIMIZATION (cont.)

Second-order sufficient optimality:

Theorem 1.6. Suppose that $f \in C^2$, that x_* satisfies the condition $g(x_*) = 0$, and that additionally $H(x_*)$ is positive definite, that is

$$s^T H(x_*) s > 0 \text{ for all } s \neq 0 \in \mathbb{R}^n.$$

Then x_* is an isolated local minimizer of f .

PROOF OF THEOREM 1.5

Suppose otherwise that $s^T H(x_*) s < 0$.

Taylor expansion in the direction s gives

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

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

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

This contradicts hypothesis that x_* is a local minimizer.

PROOF OF THEOREM 1.6

Continuity $\implies H(x)$ positive definite $\forall x$ in open ball \mathcal{N} around x^* .

$x_* + s \in \mathcal{N}$ + generalized mean value theorem $\implies \exists z$ between x_* and $x_* + s$ for which

$$\begin{aligned} f(x_* + s) &= f(x_*) + g(x_*)^T s + \frac{1}{2} s^T H(z) s \\ &= f(x_*) + \frac{1}{2} s^T H(z) s \\ &> f(x_*) \end{aligned}$$

$\forall s \neq 0 \implies x_*$ is an isolated local minimizer.

EQUALITY CONSTRAINED MINIMIZATION

First-order necessary optimality:

Theorem 1.7. Suppose that f , $c \in C^1$, and that x_* is a local minimizer of $f(x)$ subject to $c(x) = 0$. Then, so long as a first-order constraint qualification holds, there exist a vector of Lagrange multipliers y_* such that

$$\begin{aligned} c(x_*) &= 0 \quad (\text{primal feasibility}) \text{ and} \\ g(x_*) - A^T(x_*)y_* &= 0 \quad (\text{dual feasibility}). \end{aligned}$$

PROOF OF THEOREM 1.7

Constraint qualification $\implies \exists$ vector valued C^2 (C^3 for Theorem 1.8) function $x(\alpha)$ of the scalar α for which

$$x(0) = x_* \text{ and } c(x(\alpha)) = 0$$

and

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

+ Taylor's theorem \implies

$$\begin{aligned} 0 &= c_1(x(\alpha)) = c(x_* + \alpha s + \frac{1}{2} \alpha^2 p + O(\alpha^3)) \\ &= c_1(x_*) + a_1^T(x_*) (\alpha s + \frac{1}{2} \alpha^2 p) + \frac{1}{2} \alpha^2 s^T H_1(x_*) s + O(\alpha^3) \\ &= \alpha a_1^T(x_*) s + \frac{1}{2} \alpha^2 (a_1^T(x_*) p + s^T H_1(x_*) s) + O(\alpha^3) \end{aligned}$$

Matching similar asymptotic terms \implies

$$A(x_*) s = 0 \tag{1}$$

and

$$a_i^T(x_*) p + s^T H_i(x_*) s = 0 \quad \forall i = 1, \dots, m \tag{2}$$

Now consider objective function

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

$f(x)$ unconstrained along $x(\alpha) \implies$

$$g(x_*)^T s = 0 \text{ for all } s \text{ such that } A(x_*) s = 0. \tag{4}$$

Let S be a basis for null space of $A(x_*) \implies$

$$g(x_*) = A^T(x_*) y_* + S z_* \tag{5}$$

for some y_* and z_* . (4) $\implies g^T(x_*) S = 0 + A(x_*) S = 0 \implies$

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

$\implies S^T S z_* = 0 + S$ full rank $\implies z_* = 0 + (5) \implies$

$$g(x_*) - A^T(x_*) y_* = 0.$$

EQUALITY CONSTRAINED MINIMIZATION (cont.)

Second-order necessary optimality:

Theorem 1.8. Suppose that f , $c \in C^2$, and that x_* is a local minimizer of $f(x)$ subject to $c(x) = 0$. Then, provided that first- and second-order constraint qualifications hold, there exist a vector of Lagrange multipliers y_* such that

$$s^T H(x_*, y_*) s \geq 0 \text{ for all } s \in \mathcal{N}$$

where

$$\mathcal{N} = \{s \in \mathbb{R}^n \mid A(x_*)s = 0\}.$$

INEQUALITY CONSTRAINED MINIMIZATION

First-order necessary optimality:

Theorem 1.9. Suppose that f , $c \in C^1$, and that x_* is a local minimizer of $f(x)$ subject to $c(x) \geq 0$. Then, provided that a first-order constraint qualification holds, there exist a vector of Lagrange multipliers y_* such that

$$\begin{aligned} c(x_*) &\geq 0 \quad (\text{primal feasibility}), \\ g(x_*) - A^T(x_*)y_* &= 0 \quad (\text{dual feasibility}) \text{ and} \\ \text{and } y_* &\geq 0 \\ c_i(x_*)[y_*]_i &= 0 \quad (\text{complementary slackness}). \end{aligned}$$

Often known as the **Karush-Kuhn-Tucker (KKT)** conditions

PROOF OF THEOREM 1.8

$$g(x_*) - A^T(x_*)y_* = 0. \quad (6)$$

while (3) \implies

$$f(x(\alpha)) = f(x_*) + \frac{1}{2}\alpha^2 (p^T g(x_*) + s^T H(x_*)s) + O(\alpha^3) \quad (7)$$

for all s and p satisfying $A(x_*)s = 0$ and

$$a_i^T(x_*)p + s^T H_i(x_*)s = 0 \quad \forall i = 1, \dots, m. \quad (8)$$

Hence, necessarily,

$$p^T g(x_*) + s^T H(x_*)s \geq 0 \quad (9)$$

But (6) + (8) \implies

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

\implies (9) 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 $A(x_*)s = 0$.

PROOF OF THEOREM 1.9

Consider feasible perturbations about x_* . $c_i(x_*) > 0 \implies c_i(x) > 0$ for small perturbations \implies need only consider perturbations that are constrained by $c_i(x) \geq 0$ for $i \in \mathcal{A} \stackrel{\text{def}}{=} \{i : c_i(x_*) = 0\}$.

Consider $x(\alpha)$: $x(0) = x_*$, $c_i(x(\alpha)) \geq 0$ for $i \in \mathcal{A}$ and

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

\implies

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

$\forall i \in \mathcal{A} \implies$

$$s^T a_i(x_*) \geq 0 \quad \forall i \in \mathcal{A} \quad (10)$$

and

$$p^T a_i(x_*) + s^T H_i(x_*)s \geq 0 \text{ when } s^T a_i(x_*) = 0 \quad \forall i \in \mathcal{A} \quad (11)$$

Expansion (3) of $f(x(\alpha))$

$$f(x(\alpha)) = f(x_*) + \alpha g(x_*)^T s + \frac{1}{2}\alpha^2 (g(x_*)^T p + s^T H(x_*) s) + O(\alpha^3)$$

$\implies x_*$ can only be a local minimizer if

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

Result then follows directly from Farkas' lemma:

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

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

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

Expansion

$$f(x(\alpha)) = f(x_*) + \alpha g(x_*)^T s + \frac{1}{2}\alpha^2 (g(x_*)^T p + s^T H(x_*) s) + O(\alpha^3)$$

for change in objective function dominated by $\alpha s^T g(x_*)$ for feasible perturbations unless $s^T g(x_*) = 0$, in which case the expansion

$$f(x(\alpha)) = f(x_*) + \frac{1}{2}\alpha^2 (p^T g(x_*) + s^T H(x_*) s) + O(\alpha^3)$$

is relevant \implies

$$p^T g(x_*) + s^T H(x_*) s \geq 0 \quad (12)$$

holds for all feasible s for which $s^T g(x_*) = 0 \implies$

$$0 = s^T g(x_*) = \sum_{i \in \mathcal{A}} (y_*)_i s^T a_i(x_*) \implies \text{either } (y_*)_i = 0 \text{ or } a_i(x_*)^T s = 0.$$

\implies second-order feasible perturbations characterised by $s \in \mathcal{N}_+$.

INEQUALITY CONSTRAINED MINIMIZATION (cont.)

Second-order necessary optimality:

Theorem 1.10. Suppose that $f, c \in C^2$, and that x_* is a local minimizer of $f(x)$ subject to $c(x) \geq 0$. Then, provided that first- and second-order constraint qualifications hold, there exist a vector of Lagrange multipliers y_* for which primal/dual feasibility and complementary slackness requirements hold as well as

$$s^T H(x_*, y_*) s \geq 0 \text{ for all } s \in \mathcal{N}_+$$

where

$$\mathcal{N}_+ = \left\{ s \in \mathbb{R}^n \mid \begin{array}{l} s^T a_i(x_*) = 0 \text{ if } c_i(x_*) = 0 \ \& \ [y_*]_i > 0 \ \& \\ s^T a_i(x_*) \geq 0 \text{ if } c_i(x_*) = 0 \ \& \ [y_*]_i = 0 \end{array} \right\}.$$

Focus on *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 $i \in \mathcal{A} \implies s \in \mathcal{N}_+$.

When $c_i(x(\alpha)) = 0 \implies$

$$a_i^T(x_*) p + s^T H_i(x_*) s = 0$$

\implies

$$\begin{aligned} p^T g(x_*) &= \sum_{i \in \mathcal{A}} (y_*)_i p^T a_i(x_*) = \sum_{i \in \mathcal{A}} (y_*)_i p^T a_i(x_*) \\ &= - \sum_{\substack{i \in \mathcal{A} \\ (y_*)_i > 0}} (y_*)_i s^T H_i(x_*) s = - \sum_{i \in \mathcal{A}} (y_*)_i s^T H_i(x_*) s \end{aligned}$$

$$\begin{aligned} + (12) \implies s^T H(x_*, y_*) s &\equiv s^T \left(H(x_*) - \sum_{i=1}^m (y_*)_i H_i(x_*) \right) s \\ &= p^T g(x_*) + s^T H(x_*) s \geq 0. \end{aligned}$$

for all $s \in \mathcal{N}_+$

INEQUALITY CONSTRAINED MINIMIZATION (cont.)

Second-order sufficient optimality:

Theorem 1.11. Suppose that $f, c \in C^2$, that x_* and a vector of Lagrange multipliers y_* satisfy

$$c(x_*) \geq 0, g(x_*) - A^T(x_*)y_* = 0, y_* \geq 0, \text{ and } c_i(x_*)[y_*]_i = 0$$

and that

$$s^T H(x_*, y_*) s > 0$$

for all s in the set

$$\mathcal{N}_+ = \left\{ s \in \mathbb{R}^n \mid \begin{array}{l} s^T a_i(x_*) = 0 \text{ if } c_i(x_*) = 0 \ \& \ [y_*]_i > 0 \ \& \\ s^T a_i(x_*) \geq 0 \text{ if } c_i(x_*) = 0 \ \& \ [y_*]_i = 0. \end{array} \right\}.$$

Then x_* is an isolated local minimizer of $f(x)$ subject to $c(x) \geq 0$.

PROOF OF THEOREM 1.11

Consider any feasible arc $x(\alpha)$. Already shown

$$s^T a_i(x_*) \geq 0 \quad \forall i \in \mathcal{A} \tag{13}$$

and

$$p^T a_i(x_*) + s^T H_i(x_*) s \geq 0 \quad \text{when } s^T a_i(x_*) = 0 \quad \forall i \in \mathcal{A} \tag{14}$$

and that second-order feasible perturbations are characterized by \mathcal{N}_+ .

$$\begin{aligned} (14) \implies p^T g(x_*) &= \sum_{i \in \mathcal{A}} (y_*)_i p^T a_i(x_*) = \sum_{i \in \mathcal{A}} (y_*)_i p^T a_i(x_*) \\ &\geq - \sum_{\substack{i \in \mathcal{A} \\ s^T a_i(x_*)=0}} (y_*)_i s^T H_i(x_*) s = - \sum_{i \in \mathcal{A}} (y_*)_i s^T H_i(x_*) s, \end{aligned}$$

and hence by assumption that

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

$\forall s \in \mathcal{N}_+ + (3) + (13) \implies f(x(\alpha)) > f(x_*) \ \forall$ sufficiently small α .