Optimality Conditions for Nonlinear Optimisation

Lecture 7, Numerical Linear Algebra and Optimisation
Oxford University Computing Laboratory, MT 2007
Dr Raphael Hauser (hauser@comlab.ox.ac.uk)

What is a Continuous Optimisation Problem?

Unconstrained minimization:

$$\min_{x \in \mathbb{R}^n} f(x)$$

where the *objective function* $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is sufficiently smooth (often C^2 or C^2 with Lipschitz continuous second derivatives).

Equality constrained minimization:

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} f(x) \\ &\text{s.t. } c(x) = 0, \end{aligned}$$

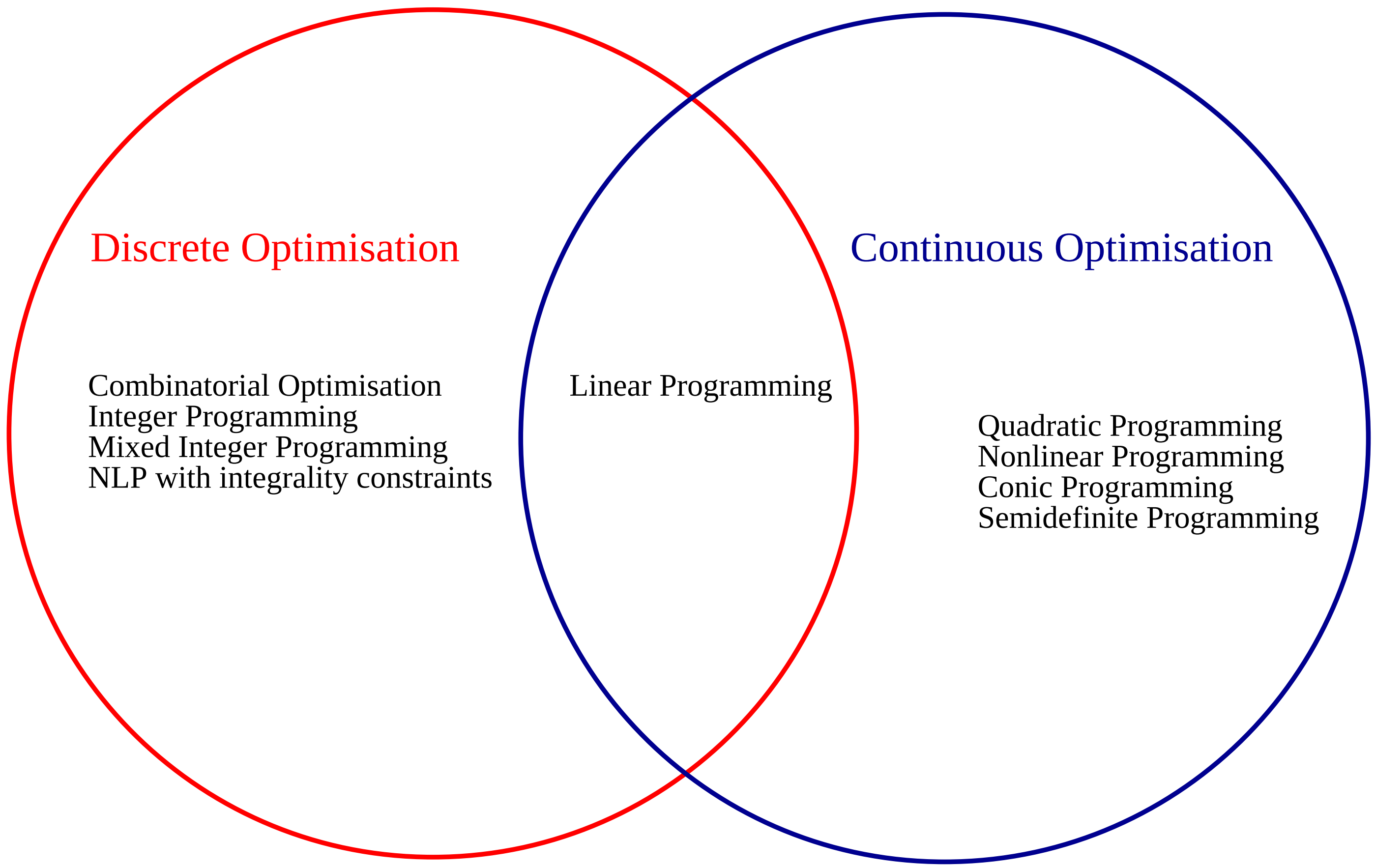
where the *equality constraints* $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$, are defined by sufficiently smooth functions, and $m \leq n$.

Inequality constrained minimization:

$$\begin{aligned}
\min_{x \in \mathbb{R}^n} \ & f(x) \\
\text{s.t.} \ & c_I(x) \geq 0, \\
& c_E(x) = 0,
\end{aligned}$$

where $c_I : \mathbb{R}^n \to \mathbb{R}^{m_I}$ and $c_E : \mathbb{R}^n \to \mathbb{R}^{m_E}$ are sufficiently smooth functions, $m_E \leq n$, and m_I may be larger than n.

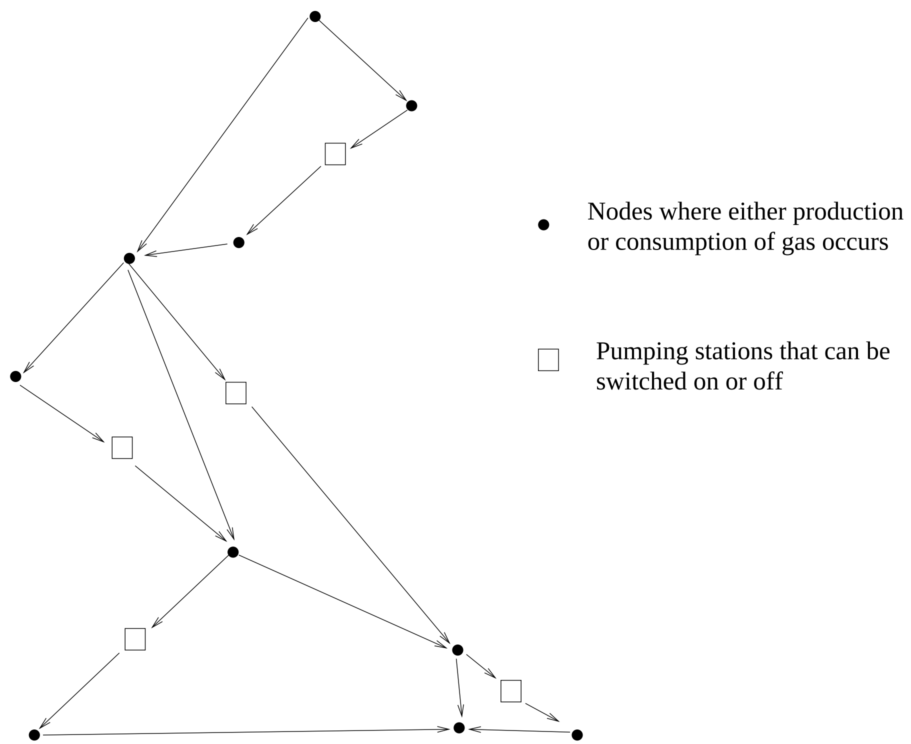
We also write $c = \begin{bmatrix} c_I \\ c_E \end{bmatrix}$.


Discrete Optimisation
Continuous Optimisation
Combinatorial Optimisation
Integer Programming
Mixed Integer Programming
NLP with integrality constraints
Linear Programming
Quadratic Programming
Nonlinear Programming
Conic Programming
Semidefinite Programming

Application Areas:

- minimum energy problems
- structural design problems
- traffic equilibrium models
- production scheduling problems
- portfolio selection
- parameter determination in financial markets
- hydro-electric power scheduling
- gas production models
- computer tomography (image reconstruction)
- efficient models of alternative energy sources
- etc. etc.

Example 1 (Optimising a Gas Pipeline Network)**.**



- Given is the production rate d_i of gas at each node i ($d_i < 0$ corresponds to consumption).

- Decision variables
  - p_i the pressure at node i,
  - q_{ij} the flow rate between nodes i and j,
  - $z_j = 1$ if pumping station j is switched on, $z_j = 0$ otherwise.

- Constraints
  - $\sum_{k \neq i} q_{ki} + d_i = \sum_{j \neq i} q_{ij}$ (conservation of gas),
  - $p_k^2 = p_i^2 + \kappa_{ki} q_{ki}^{2.8359}$ for all k, i connected via a pipe (pipe equations),

    $$A_1^\top p^2 + K q^{2.8359} = 0$$

    nonlinear, sparse structured system of equations,
  - $q_{ij} - q_{jk} + z_j \cdot c_j(p_i, q_{ij}, p_k, q_{jk}) \geq 0$ for all nodes i, k connected via compressors j (compressor constraints),

    $$A_2^\top q + z \cdot c(p, q) \geq 0,$$

    nonlinear, sparse structured system of inequalities with binary variables,

- $p_{\text{min}} \le p \le p_{\text{max}}$, $q_{\text{min}} \le q \le q_{\text{max}}$ (bound constriaints).

- Objectives
  - minimise sum of pressures,
  - or minimise compressor fuel costs,

- Statistics of the British Gas National Transmission System
  - 199 nodes
  - 196 pipes
  - 21 machines
  - for steady state problem, ≈ 400 variables,
  - for 24-hour variable demand problem with 10 minute discretization, $\approx 58,000$ variables.

  Problem to be solved in real time.

This problem is typical of real-world, large-scale applications

- simple bounds
- linear constraints
- nonlinear constraints
- structure
- global solution "required"
- integer variables
- discretisation

Notation and Basic Tools

$$\begin{aligned}
g(x) &= \nabla f(x) = [\mathsf{D}_x\, f(x)]^{\mathsf{T}}, \quad \text{the gradient of } f,\\
H(x) &= \mathsf{D}_{xx}\, f(x), \quad \text{the Hessian of } f,\\
a_i(x) &= \nabla c_i(x), \quad \text{the gradient of the } i\text{-th constraint function},\\
H_i(x) &= \mathsf{D}_{xx}\, c_i(x), \quad \text{the Hessian of the } i\text{-th constraint function},\\
A(x) &= \mathsf{D}_x\, c(x) = \begin{bmatrix} a_1(x) & \dots & a_m(x) \end{bmatrix}^{\mathsf{T}}, \quad \text{the Jacobian of } c,\\
\ell(x,y) &= f(x) - y^{\mathsf{T}} c(x), \quad \text{the Lagrangian function},\\
H(x,y) &= \mathsf{D}_{xx}\, \ell(x,y) = H(x) - \sum_{i=1}^{m} y_i H_i(x), \quad \text{the } x\text{-Hessian of } \ell.
\end{aligned}$$

The variables y that appear in ℓ and H are called *Lagrange multipliers*.

Definition 2 (Lipschitz Continuity)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be open sets in two normed spaces $(N_{\mathcal{X}}, \|\cdot\|_{\mathcal{X}})$ and $(N_{\mathcal{Y}}, \|\cdot\|_{\mathcal{Y}})$. A function $F : \mathcal{X} \to \mathcal{Y}$ is called

i) *Lipschitz continuous* at $x \in \mathcal{X}$ if there exists a $\gamma(x) > 0$ such that

$$\|F(z) - F(x)\|_{\mathcal{Y}} \leq \gamma(x)\|z - x\|_{\mathcal{X}}, \quad (z \in \mathcal{X}), \tag{1}$$

ii) *uniformly Lipschitz continuous* in $\mathcal{X}$ if there exists a $\gamma > 0$ such that (1) holds true with $\gamma(x) = \gamma$ for all $x \in \mathcal{X}$.

Theorem 3. *[A Useful Taylor Approximation] Let $\mathcal{S}$ be an open subset of $\mathbb{R}^n$, $x, s \in \mathbb{R}^n$ such that $x + \theta s \in \mathcal{S}$ for all $\theta \in [0, 1]$.*

i) If $f \in C^1(\mathcal{S}, \mathbb{R})$, and its gradient $g(x)$ is Lipschitz continuous at x with Lipschitz constant $\gamma^{\mathsf{L}}(x)$, then

$$\left|f(x+s) - \mathsf{m}^{\mathsf{L}}(x+s)\right| \leq \frac{1}{2}\gamma^{\mathsf{L}}(x)\|s\|^2,$$

where $\mathsf{m}^{\mathsf{L}}(x+s) = f(x) + g(x)^{\mathsf{T}} s$ is the first-order Taylor approximation of f at x (a linear model).

ii) (Vectorisation of i)) If $F \in C^1(\mathcal{S}, \mathbb{R}^m)$, and its Jacobian $\mathsf{D}_x F(x)$ is Lipschitz continuous at x with Lipschitz constant $\gamma^{\mathsf{L}}(x)$ (using the matrix operator norm induced by the norms on $\mathbb{R}^n$ and $\mathbb{R}^m$), then

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

where $\mathsf{M}^{\mathsf{L}}(x+s) = F(x) + \mathsf{D}_x F(x) s$ is the first-order Taylor approximation of f at x.

iii) If $f \in C^2(\mathcal{S}, \mathbb{R})$, and its Hessian $H(x)$ is Lipschitz continuous at x with Lipschitz constant $\gamma^{\mathsf{Q}}(x)$, then

$$\left| f(x+s) - \mathsf{m}^{\mathsf{Q}}(x+s) \right| \leq \frac{1}{6}\gamma^{\mathsf{Q}}(x)\|s\|^3,$$

where $\mathsf{m}^{\mathsf{Q}}(x+s) = f(x) + s^\top g(x) + \frac{1}{2}s^\top H(x)s$ is the second-order Taylor approximation of f at x (a quadratic model).

Optimality Conditions for Continuous Optimisation

Optimality conditions are useful for the following reasons:

- They provide a means of guaranteeing that a candidate solution is indeed optimal ($\rightarrow$ sufficient conditions).

- They indicate when a point is not optimal ($\rightarrow$ necessary conditions).

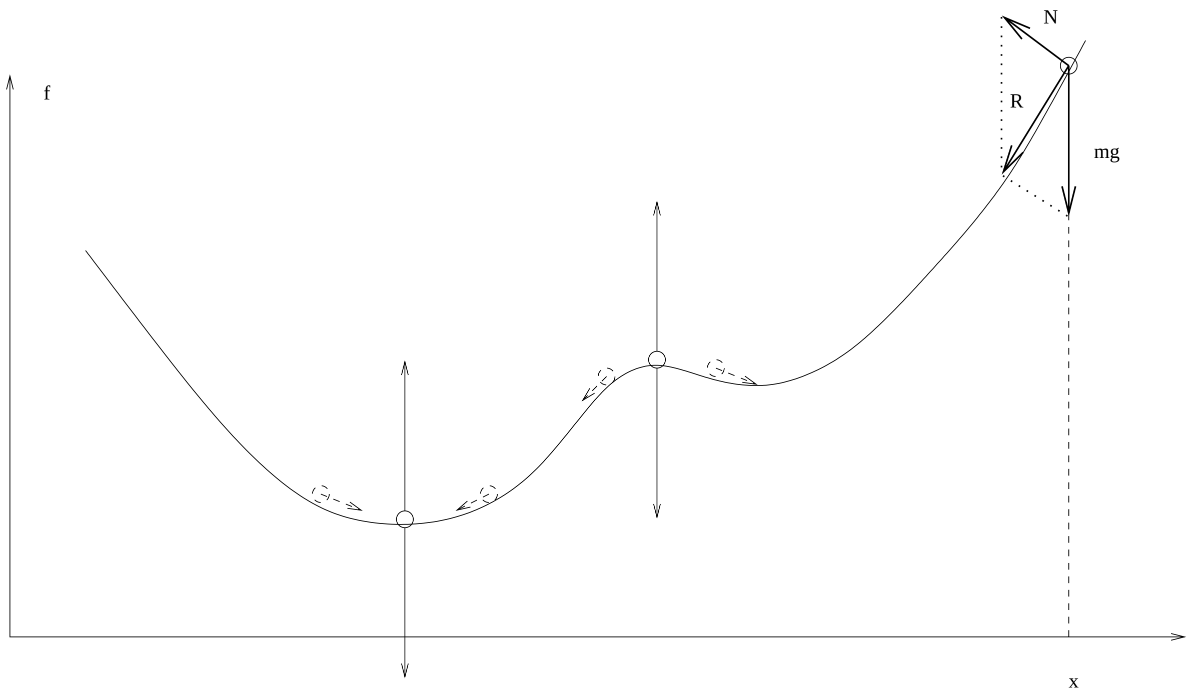
- They form a guide in the design of algorithms, since lack of optimality is an indication of potential for improvement.

We first consider the unconstrained minimisation problem

$$\text{(UCM)} \quad \min_{x \in \mathbb{R}^n} f(x),$$

where $f \in C^1(\mathbb{R}^n, \mathbb{R})$.

Definition 4. A *local minimiser* for problem (UCM) is a point $x^* \in \mathbb{R}^n$ for which there exists $\varrho > 0$ such that $f(x) \geq f(x^*)$ for all $x \in \mathrm{B}_\varrho(x^*)$.



Theorem 5 (Necessary Optimality Conditions for (UCM)). *Let x^* be a local minimiser for problem (UCM).*

i) Then the following first order necessary condition *must hold,*

$$g(x^*) = 0.$$

ii) If furthermore $f \in C^2$, then the following second order necessary condition *must also hold,*

$$s^{\mathsf{T}} H(x^*) s \geq 0, \quad (s \in \mathbb{R}^n),$$

that is, $H(x^*)$ is positive semidefinite.

.

Theorem 6 (Sufficient Optimality Conditions for (UCM)). *Let* $f \in C^2$*, and let* $x^* \in \mathbb{R}^n$ *be a point where the following* sufficient optimality conditions *are satisfied,*

$$g(x^*) = 0,$$
$$s^\top H(x^*) s > 0 \quad (s \in \mathbb{R}^n \setminus \{0\}),$$

that is, $H(x^*)$ *is positive definite. Then* x^* *is an isolated local minimiser of* f*.*

The situation is more complex in the case of the equality constrained minimisation problem

$$\text{(ECM)} \quad \min_{x \in \mathbb{R}^n} f(x) \\ \text{s.t. } c(x) = 0,$$

where $f \in C^1(\mathbb{R}^n, \mathbb{R})$ and $c \in C^1(\mathbb{R}^n, \mathbb{R}^m)$.

Definition 7. A *local minimiser* for problem (ECM) is a point $x^* \in \mathbb{R}^n$ for which there exists $\varrho > 0$ such that $f(x) \geq f(x^*)$ for all $x \in \mathrm{B}_\varrho(x^*) \cap \{x : c(x) = 0\}$.

Definition 8 (LICQ)**.** The *linear independence constraint qualification* is satisfied at x^* if the set of gradient vectors $\{a_i(x^*) : i = 1, \ldots, m\}$ defined by the constraint functions is linearly independent.

Theorem 9 (Necessary Optimality Conditions for (ECM)). *Let x^* be a local minimiser for problem (ECM) where the LICQ holds.*

i) Then the following first order necessary conditions *must hold: There exists a vector $y^* \in \mathbb{R}^m$ of* Lagrange multipliers *such that*

$$c(x^*) = 0, \quad \textit{(primal feasibility)}$$
$$\nabla_x \ell(x^*, y^*) = g(x^*) - A^\top(x^*) y^* = 0 \quad \textit{(dual feasibility)}.$$

(Recall that $g(x^*) - A^\top(x^*)y^* = \nabla f(x^*) - \sum_{i=1}^m y_i^* \nabla c_i(x^*)$, so that dual feasibility says that $\nabla f(x^*)$ is "counterbalanced" by a linear combination of the $\nabla c_i(x^*)$).

ii) Furthermore, if $f, c \in C^2$, then for all $s \in \mathbb{R}^n$ such that

$$a_i(x^*)^\top s = 0, \quad (i = 1, \ldots, m), \tag{2}$$

the following second order necessary condition *must also hold,*

$$s^\top H(x^*, y^*) s \geq 0, \tag{3}$$

that is, $H(x^*)$ is positive semidefinite in the directions that lie in the tangent space of the feasible manifold.

Note that Theorem 5 is merely the special case $m = 0$ of Theorem 9. Let us now further generalise the result so that it applies to minimisation problems with inequality constraints,

$$\text{(ICM)} \qquad \begin{aligned} &\min_{x \in \mathbb{R}^n} f(x) \\ &\text{s.t. } c_I(x) \geq 0, \\ &\qquad c_E(x) = 0, \end{aligned}$$

where $f \in C^1(\mathbb{R}^n, \mathbb{R})$, $c_I \in C^1(\mathbb{R}^n, \mathbb{R}^{m_I})$ and $c_E \in C^1(\mathbb{R}^n, \mathbb{R}^{m_E})$.

Definition 10. A *local minimiser* for problem (ICM) is a point $x^* \in \mathbb{R}^n$ for which there exists $\varrho > 0$ such that $f(x) \geq f(x^*)$ for all $x \in \mathrm{B}_\varrho(x^*) \cap \{x : c_E(x) = 0,\ c_I(x) \geq 0\}$.

Definition 11. Let x^* be feasible for (ICM). The *active set* of constraints at x^* is the set of indices

$$\mathscr{A}(x^*) = E \cup \{i \in I : c_i(x^*) = 0\}.$$

The *linear independence constraint qualification* is satisfied at x^* if the set of vectors $\{\nabla c_i(x^*) : i \in \mathscr{A}(x^*)\}$ is linearly independent.

Theorem 12 (Necessary Optimality Conditions for (ICM)). *Let x^* be a local minimiser for problem (ICM) where the LICQ holds.*

i) *Then the following* first order necessary conditions *must hold: There exists a vector $y^* \in \mathbb{R}^m$ of* Lagrange multipliers *such that*

$$
\begin{aligned}
c_E(x^*) &= 0, \quad \textit{(primal feasibility 1)},\\
c_I(x^*) &\geq 0, \quad \textit{(primal feasibility 2)},\\
\nabla_x \ell(x^*, y^*) &= g(x^*) - A^\mathsf{T}(x^*) y^* = 0 \quad \textit{(dual feasibility 1)},\\
y_i^* &\geq 0, \quad (i \in I) \quad \textit{(dual feasibility 2)},\\
c_i(x^*) y_i^* &= 0, \quad (i \in E \cup I) \quad \textit{(complementarity)}.
\end{aligned}
$$

(These conditions are also called Karush-Kuhn-Tucker *(KKT) conditions. Complementarity guarantees that $y_i^* = 0$ for all $i \notin \mathscr{A}(x^*)$.)*

ii) *Furthermore, if $f, c \in C^2$, then for all $s \in \mathbb{R}^n$ such that*

$$
\begin{aligned}
a_i(x^*)^\mathsf{T} s &= 0, \quad (i \in E \cup \{i \in I : i \in \mathscr{A}(x^*), y_i^* > 0\}),\\
a_i(x^*)^\mathsf{T} s &\geq 0, \quad (i \in \{i \in I : i \in \mathscr{A}(x^*), y_i^* = 0\}),
\end{aligned}
$$

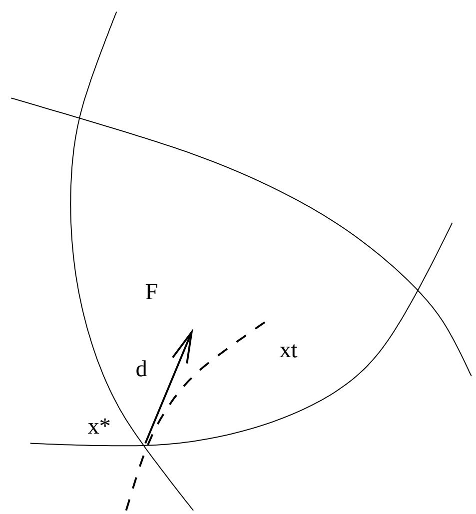
the following second order necessary condition *must also hold,*

$$s^\mathsf{T} H(x^*, y^*) s \geq 0.$$

Remark 13. The second order optimality analysis is based on the following observation:

If x^* is a local minimiser of (ICM) and $x(t)$ is a feasible exit path from x^* with tangent s at x^*, then x^* must also be a local minimiser for the univariate constrained optimisation problem

$$\begin{aligned} &\min f(x(t)) \\ &\text{s.t. } t \geq 0 \end{aligned}$$



Theorem 14 (Sufficient Optimality Conditions for (ICM)). *Let x^* be a feasible point for (ICM) at which the LICQ holds, where it is assumed that $f, c \in C^2$. Let $y^* \in \mathbb{R}^m$ be a vector of Lagrange multipliers such that (x^*, y^*) satisfy the KKT conditions (see Theorem 12). If it is furthermore the case that*

$$s^{\mathsf{T}} H(x^*, y^*) s > 0$$

for all $s \in \mathbb{R}^n$ that satisfy

$$\begin{aligned} a_i(x^*)^{\mathsf{T}} s = 0, &\quad (i \in E \cup \{i \in I : i \in \mathscr{A}(x^*), y_i^* > 0\}), \\ a_i(x^*)^{\mathsf{T}} s \geq 0, &\quad (i \in \{i \in I : i \in \mathscr{A}(x^*), y_i^* = 0\}), \end{aligned}$$

then x^* is a local minimiser for (ICM).