Lectures 10 and 11: Constrained optimization problems and their optimality conditions

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

Problems and solutions

$$\text{minimize } f(x) \quad \text{subject to } x \in \Omega \subseteq \mathbb{R}^n.$$

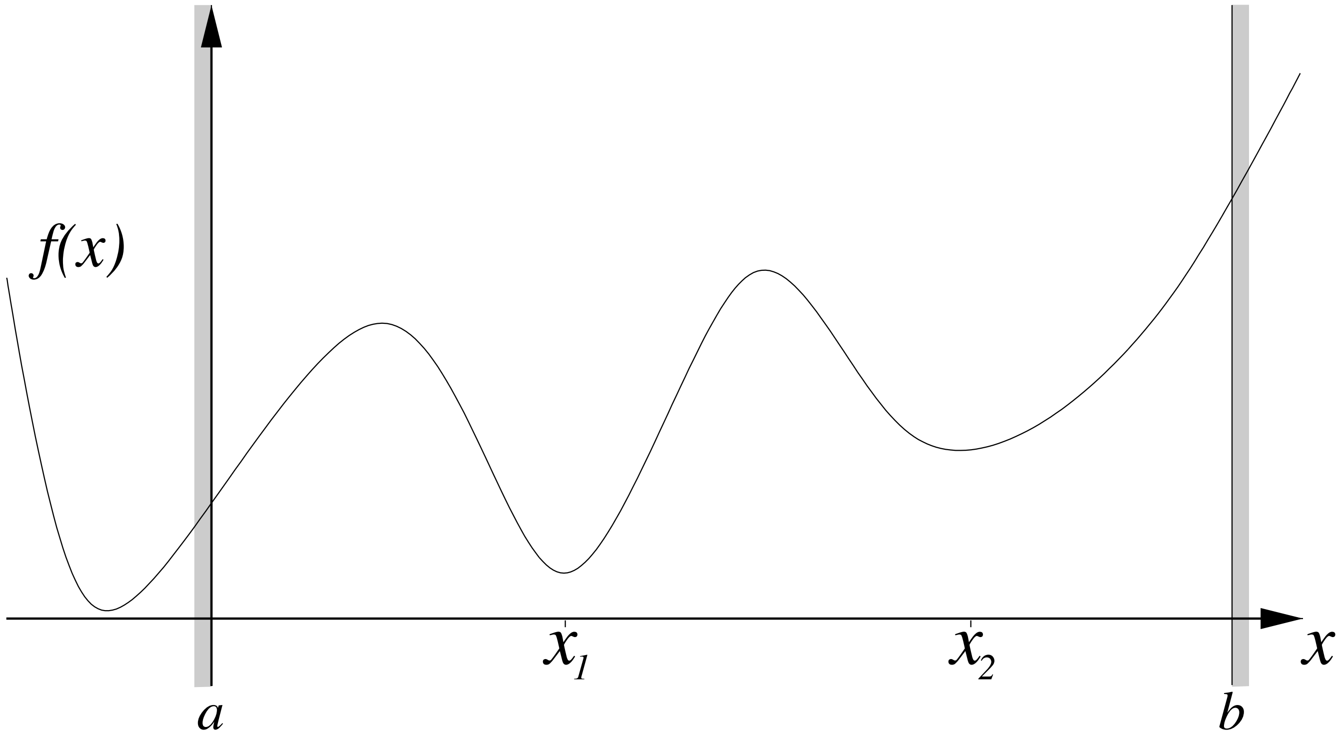
- $f : \Omega \to \mathbb{R}$ is (sufficiently) smooth.
- f objective; x variables.
- Ω feasible set determined by finitely many (equality and/or inequality) constraints.

x^* global minimizer of f over $\Omega \implies f(x) \geq f(x^*), \forall x \in \Omega$.

x^* local minimizer of f over $\Omega \implies$
$\exists N(x^*, \delta)$ such that $f(x) \geq f(x^*)$, for all $x \in \Omega \cap N(x^*, \delta)$.
• $N(x^*, \delta) := \{x \in \mathbb{R}^n : \|x - x^*\| \leq \delta\}$.

Example problem in one dimension

$$\textbf{Example :} \quad \min f(x) \quad \text{subject to} \quad a \le x \le b.$$

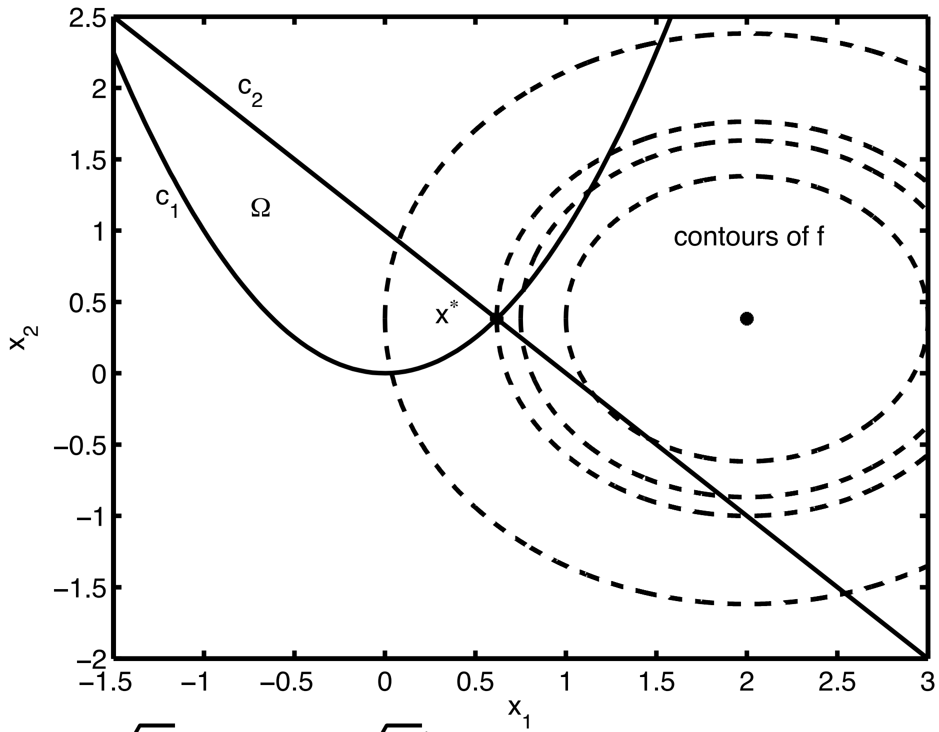


- The feasible region Ω is the interval $[a, b]$.
- The point x_1 is the global minimizer; x_2 is a local (non-global) minimizer; $x = a$ is a constrained local minimizer.

An example of a nonlinear constrained problem

$$\min_{x\in\mathbb{R}^2} (x_1 - 2)^2 + (x_2 - 0.5(3-\sqrt{5}))^2 \quad \text{subject to}$$

$$-x_1 - x_2 + 1 \geq 0, \;\; x_2 - x_1^2 \geq 0.$$



$x^* = 0.5(-1+\sqrt{5}, 3-\sqrt{5})$; Ω feasible set.

Optimality conditions for constrained problems

== algebraic characterizations of solutions $\longrightarrow$ suitable for computations.

- provide a way to guarantee that a candidate point is optimal (sufficient conditions)
- indicate when a point is not optimal (necessary conditions)

$$\text{minimize}_{x\in\mathbb{R}^n} \quad f(x) \quad \text{subject to} \quad c_E(x) = 0, \quad c_I(x) \geq 0. \tag{CP}$$

- $f : \mathbb{R}^n \to \mathbb{R}$, $c_E : \mathbb{R}^n \to \mathbb{R}^m$ and $c_I : \mathbb{R}^n \to \mathbb{R}^p$ (suff.) smooth;

• $c_I(x) \geq 0 \Leftrightarrow c_i(x) \geq 0$, $i \in I$.

• $\Omega := \{x : \ c_E(x) = 0, \ c_I(x) \geq 0\}$ feasible set of the problem.

Optimality conditions for constrained problems

unconstrained problem $\longrightarrow \hat{x}$ stationary point ($\nabla f(\hat{x}) = 0$).

constrained problem $\longrightarrow$ $\hat{x}$ Karush-Kuhn-Tucker (KKT) point.

Definition: $\hat{x}$ KKT point of (CP) if there exist $\hat{y} \in \mathbb{R}^m$ and $\hat{\lambda} \in \mathbb{R}^p$ such that $(\hat{x}, \hat{y}, \hat{\lambda})$ satisfies

$$\nabla f(\hat{x}) = \sum_{j \in E} \hat{y}_j \nabla c_j(\hat{x}) + \sum_{i \in I} \hat{\lambda}_i \nabla c_i(\hat{x}),$$

$$c_E(\hat{x}) = 0, \quad c_I(\hat{x}) \geq 0,$$

$$\hat{\lambda}_i \geq 0, \quad \hat{\lambda}_i c_i(\hat{x}) = 0, \quad \text{for all } i \in I.$$

Optimality conditions for constrained problems

unconstrained problem $\longrightarrow \hat{x}$ stationary point ($\nabla f(\hat{x}) = 0$).

constrained problem $\longrightarrow$ $\hat{x}$ Karush-Kuhn-Tucker (KKT) point.

Definition: $\hat{x}$ KKT point of (CP) if there exist $\hat{y} \in \mathbb{R}^m$ and $\hat{\lambda} \in \mathbb{R}^p$ such that $(\hat{x}, \hat{y}, \hat{\lambda})$ satisfies

$$\nabla f(\hat{x}) = \sum_{j \in E} \hat{y}_j \nabla c_j(\hat{x}) + \sum_{i \in I} \hat{\lambda}_i \nabla c_i(\hat{x}),$$

$$c_E(\hat{x}) = 0, \quad c_I(\hat{x}) \geq 0,$$

$$\hat{\lambda}_i \geq 0, \quad \hat{\lambda}_i c_i(\hat{x}) = 0, \quad \text{for all } i \in I.$$

• Let $\mathcal{A} := E \cup \{i \in I : c_i(\hat{x}) = 0\}$ index set of active constraints at $\hat{x}$; $c_j(\hat{x}) > 0$ inactive constraint at $\hat{x} \Rightarrow \hat{\lambda}_j = 0$. Then $\sum_{i \in I} \hat{\lambda}_i \nabla c_i(\hat{x}) = \sum_{i \in I \cap \mathcal{A}} \hat{\lambda}_i \nabla c_i(\hat{x})$.

Optimality conditions for constrained problems

unconstrained problem $\longrightarrow \hat{x}$ stationary point ($\nabla f(\hat{x}) = 0$).

constrained problem $\longrightarrow$ $\hat{x}$ Karush-Kuhn-Tucker (KKT) point.

Definition: $\hat{x}$ KKT point of (CP) if there exist $\hat{y} \in \mathbb{R}^m$ and $\hat{\lambda} \in \mathbb{R}^p$ such that $(\hat{x}, \hat{y}, \hat{\lambda})$ satisfies

$$\nabla f(\hat{x}) = \sum_{j \in E} \hat{y}_j \nabla c_j(\hat{x}) + \sum_{i \in I} \hat{\lambda}_i \nabla c_i(\hat{x}),$$

$$c_E(\hat{x}) = 0, \quad c_I(\hat{x}) \geq 0,$$

$$\hat{\lambda}_i \geq 0, \quad \hat{\lambda}_i c_i(\hat{x}) = 0, \quad \text{for all } i \in I.$$

• Let $\mathcal{A} := E \cup \{i \in I : c_i(\hat{x}) = 0\}$ index set of active constraints at $\hat{x}$; $c_j(\hat{x}) > 0$ inactive constraint at $\hat{x} \Rightarrow \hat{\lambda}_j = 0$. Then $\sum_{i \in I} \hat{\lambda}_i \nabla c_i(\hat{x}) = \sum_{i \in I \cap \mathcal{A}} \hat{\lambda}_i \nabla c_i(\hat{x})$.

• $J(x) = \left(\nabla c_i(x)^T\right)_i$ Jacobian matrix of constraints c. Thus $\sum_{j \in E} \hat{y}_j \nabla c_j(\hat{x}) = J_E(x)^T \hat{y}$ and $\sum_{i \in I} \hat{\lambda}_i \nabla c_i(\hat{x}) = J_I(x)^T \hat{\lambda}$.

Optimality conditions for constrained problems ...

$\hat{x}$ KKT point $\longrightarrow$ $\hat{y}$ and $\hat{\lambda}$ Lagrange multipliers of the equality and inequality constraints, respectively.

$\hat{y}$ and $\hat{\lambda}$ $\longrightarrow$ sensitivity analysis.

$\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ Lagrangian function of (CP),

$$\mathcal{L}(x, y, \lambda) := f(x) - y^\top c_E(x) - \lambda^\top c_I(x), \quad x \in \mathbb{R}^n.$$

Thus $\nabla_x \mathcal{L}(x, y, \lambda) = \nabla f(x) - J_E(x)^\top y - J_I(x)^\top \lambda$,

and $\hat{x}$ KKT point of (CP) $\Longrightarrow$ $\nabla_x \mathcal{L}(\hat{x}, \hat{y}, \hat{\lambda}) = 0$
(i. e., $\hat{x}$ is a stationary point of $\mathcal{L}(\cdot, \hat{y}, \hat{\lambda})$).

- duality theory...

An illustration of the KKT conditions

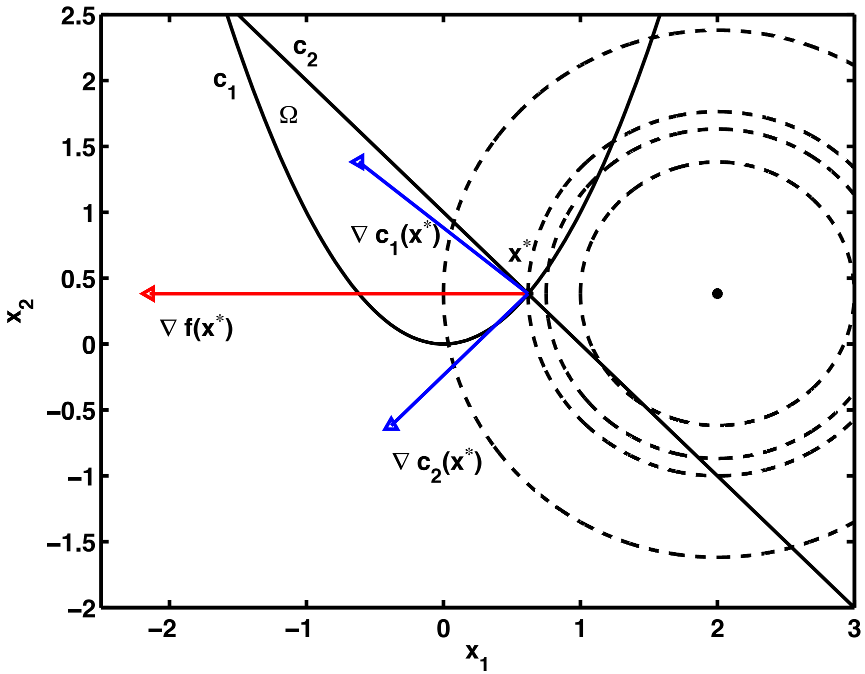
$$\min_{x\in\mathbb{R}^2} (x_1 - 2)^2 + (x_2 - 0.5(3 - \sqrt{5}))^2 \quad \text{subject to}$$

$$-x_1 - x_2 + 1 \geq 0, \; x_2 - x_1^2 \geq 0. \qquad (*)$$

$x^* = \frac{1}{2}(-1 + \sqrt{5}, 3 - \sqrt{5})^\top$:

- global solution of $(*)$,
- KKT point of $(*)$.

$\nabla f(x^*) = (-5 + \sqrt{5}, 0)^\top$,
$\nabla c_1(x^*) = (1 - \sqrt{5}, 1)^\top$,
$\nabla c_2(x^*) = (-1, -1)^\top$.



$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*) + \lambda_2^* \nabla c_2(x^*)$, with $\lambda_1^* = \lambda_2^* = \sqrt{5} - 1 > 0$.
$c_1(x^*) = c_2(x^*) = 0$: constraints are active at x^*.

An illustration of the KKT conditions ...

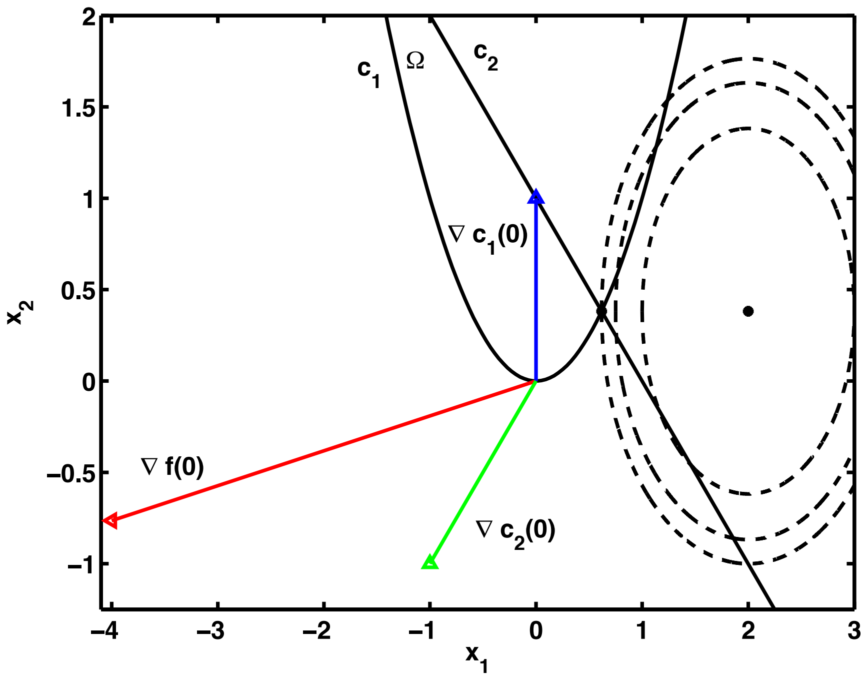
$$\min_{x\in\mathbb{R}^2} (x_1-2)^2 + (x_2 - 0.5(3-\sqrt{5}))^2 \quad \text{subject to}$$

$$-x_1 - x_2 + 1 \geq 0,\ x_2 - x_1^2 \geq 0. \qquad (*)$$

$x := (0,0)^\top$
is NOT a KKT point of $(*)$!

$c_1(x) = 0$: active at x.
$c_2(x) = 1$: inactive at x.
$\Longrightarrow \lambda_2 = 0$ and
$\nabla f(x) = \lambda_1 \nabla c_1(x)$,
with $\lambda_1 \geq 0$.

$\Downarrow$



Contradiction with $\nabla f(x) = (-4, \sqrt{5}-3)^\top$ and
$\nabla c_1(x) = (0,1)^\top$.

Optimality conditions for constrained problems ...

In general, need constraints/feasible set of (CP) to satisfy regularity assumption called constraint qualification in order to derive optimality conditions.

Theorem 16 (First order necessary conditions) Under suitable constraint qualifications,

x^* local minimizer of (CP) $\Longrightarrow$ x^* KKT point of (CP).

Optimality conditions for constrained problems ...

In general, need constraints/feasible set of (CP) to satisfy regularity assumption called constraint qualification in order to derive optimality conditions.

Theorem 16 (First order necessary conditions) Under suitable constraint qualifications,
x^* local minimizer of (CP) $\Longrightarrow$ x^* KKT point of (CP).

Proof of Theorem 16 (for equality constraints only): Let $I = \emptyset$. Then the KKT conditions become: $c_E(x^*) = 0$ (which is trivial as x^* feasible) and $\nabla f(x^*) = J_E(x^*)^T y^*$ for some $y^* \in \mathbb{R}^m$, where J_E is the Jacobian matrix of the constraints c_E.

Optimality conditions for constrained problems ...

In general, need constraints/feasible set of (CP) to satisfy regularity assumption called constraint qualification in order to derive optimality conditions.

Theorem 16 (First order necessary conditions) Under suitable constraint qualifications,
x^* local minimizer of (CP) $\Longrightarrow$ x^* KKT point of (CP).

Proof of Theorem 16 (for equality constraints only): Let $I = \emptyset$. Then the KKT conditions become: $c_E(x^*) = 0$ (which is trivial as x^* feasible) and $\nabla f(x^*) = J_E(x^*)^T y^*$ for some $y^* \in \mathbb{R}^m$, where J_E is the Jacobian matrix of the constraints c_E.

Consider feasible perturbations/paths $x(\alpha)$ around x^*, where α (sufficiently small) scalar, $x(\alpha) \in \mathcal{C}^1(\mathbb{R}^n)$ and

$$x(0) = x^*,\ x(\alpha) = x^* + \alpha s + \mathcal{O}(\alpha^2),\ s \neq 0 \text{ and } c(x(\alpha)) = 0^{(\dagger)}.$$

(†) requires constraint qualifications, namely, assuming the existence of $s \neq 0$ with above properties.

Optimality conditions for constrained problems ...

Proof of Theorem 16 (for equality constraints only): (continued)
For any $i \in E$, by Taylor's theorem for $c_i(x(\alpha))$ around x^*,

$$\begin{aligned} 0 &= c_i(x(\alpha)) = c_i(x^* + \alpha s + \mathcal{O}(\alpha^2)) \\ &= c_i(x^*) + \nabla c_i(x^*)^T(x^* + \alpha s - x^*) + \mathcal{O}(\alpha^2) \\ &= \alpha \nabla c_i(x^*)^T s + \mathcal{O}(\alpha^2), \end{aligned}$$

where we used $c_i(x^*) = 0$.

Optimality conditions for constrained problems ...

Proof of Theorem 16 (for equality constraints only): (continued)
For any $i \in E$, by Taylor's theorem for $c_i(x(\alpha))$ around x^*,

$$\begin{aligned} 0 &= c_i(x(\alpha)) = c_i(x^* + \alpha s + \mathcal{O}(\alpha^2)) \\ &= c_i(x^*) + \nabla c_i(x^*)^T(x^* + \alpha s - x^*) + \mathcal{O}(\alpha^2) \\ &= \alpha \nabla c_i(x^*)^T s + \mathcal{O}(\alpha^2), \end{aligned}$$

where we used $c_i(x^*) = 0$. Dividing both sides by α, we deduce

$$0 = \nabla c_i(x^*)^T s + \mathcal{O}(\alpha),$$

for all α sufficiently small. Letting $\alpha \to 0$, we obtain

$$\nabla c_i(x^*)^T s = 0 \text{ for all } i \in E,$$

and so $J_E(x^*)s = 0$. [In other words, any feasible direction s (which is assumed to exist) satisfies $J_E(x^*)s = 0$.]

Optimality conditions for constrained problems ...

Proof of Theorem 16 (for equality constraints only): (continued)

Now expanding f, we deduce

$$\begin{aligned} f(x(\alpha)) &= f(x^*) + \nabla f(x^*)^T (x^* + \alpha s - s^*) + \mathcal{O}(\alpha^2) \\ &= f(x^*) + \alpha \nabla f(x^*)^T s + \mathcal{O}(\alpha^2). \end{aligned}$$

Optimality conditions for constrained problems ...

Proof of Theorem 16 (for equality constraints only): (continued)
Now expanding f, we deduce

$$f(x(\alpha)) = f(x^*) + \nabla f(x^*)^T (x^* + \alpha s - s^*) + \mathcal{O}(\alpha^2)$$
$$= f(x^*) + \alpha \nabla f(x^*)^T s + \mathcal{O}(\alpha^2).$$

Since x^* is a local minimizer of f, we have $f(x(\alpha)) \geq f(x^*)$ for all α sufficiently small. Thus $\alpha \nabla f(x^*)^T s + \mathcal{O}(\alpha^2) \geq 0$ for all α sufficiently small.

Optimality conditions for constrained problems ...

Proof of Theorem 16 (for equality constraints only): (continued)
Now expanding f, we deduce

$$\begin{aligned} f(x(\alpha)) &= f(x^*) + \nabla f(x^*)^T(x^* + \alpha s - s^*) + \mathcal{O}(\alpha^2) \\ &= f(x^*) + \alpha \nabla f(x^*)^T s + \mathcal{O}(\alpha^2). \end{aligned}$$

Since x^* is a local minimizer of f, we have $f(x(\alpha)) \geq f(x^*)$ for all α sufficiently small. Thus $\alpha \nabla f(x^*)^T s + \mathcal{O}(\alpha^2) \geq 0$ for all α sufficiently small. Considering $\alpha > 0$, we divide by α to obtain $\nabla f(x^*)^T s + \mathcal{O}(\alpha) \geq 0$; now letting $\alpha \to 0$, we deduce $\nabla f(x^*)^T s \geq 0$.

Optimality conditions for constrained problems ...

Proof of Theorem 16 (for equality constraints only): (continued)
Now expanding f, we deduce

$$\begin{aligned} f(x(\alpha)) &= f(x^*) + \nabla f(x^*)^T (x^* + \alpha s - s^*) + \mathcal{O}(\alpha^2) \\ &= f(x^*) + \alpha \nabla f(x^*)^T s + \mathcal{O}(\alpha^2). \end{aligned}$$

Since x^* is a local minimizer of f, we have $f(x(\alpha)) \geq f(x^*)$ for all α sufficiently small. Thus $\alpha \nabla f(x^*)^T s + \mathcal{O}(\alpha^2) \geq 0$ for all α sufficiently small. Considering $\alpha > 0$, we divide by α to obtain $\nabla f(x^*)^T s + \mathcal{O}(\alpha) \geq 0$; now letting $\alpha \to 0$, we deduce $\nabla f(x^*)^T s \geq 0$. Similarly, considering $\alpha < 0$, we obtain $\nabla f(x^*)^T s \leq 0$. Thus

$$\nabla f(x^*)^T s = 0 \text{ for all } s \text{ such that } J_E(x^*) s = 0. \qquad (1)$$

Optimality conditions for constrained problems ...

Proof of Theorem 16 (for equality constraints only): (continued)
Now expanding f, we deduce

$$\begin{aligned} f(x(\alpha)) &= f(x^*) + \nabla f(x^*)^T (x^* + \alpha s - s^*) + \mathcal{O}(\alpha^2) \\ &= f(x^*) + \alpha \nabla f(x^*)^T s + \mathcal{O}(\alpha^2). \end{aligned}$$

Since x^* is a local minimizer of f, we have $f(x(\alpha)) \geq f(x^*)$ for all α sufficiently small. Thus $\alpha \nabla f(x^*)^T s + \mathcal{O}(\alpha^2) \geq 0$ for all α sufficiently small. Considering $\alpha > 0$, we divide by α to obtain $\nabla f(x^*)^T s + \mathcal{O}(\alpha) \geq 0$; now letting $\alpha \to 0$, we deduce $\nabla f(x^*)^T s \geq 0$. Similarly, considering $\alpha < 0$, we obtain $\nabla f(x^*)^T s \leq 0$. Thus

$$\nabla f(x^*)^T s = 0 \text{ for all } s \text{ such that } J_E(x^*) s = 0. \qquad (1)$$

By rank-nullity theorem, (1) implies that $\nabla f(x^*)$ must belong to the range space of $J_E(x^*)^T$ (ie, span of columns of $J_E(x^*)^T$), and so $\nabla f(x^*) = J_E(x^*)^T y^*$ for some y^*. The next slide details this argument.

Optimality conditions for constrained problems ...

Proof of Theorem 16 (for equality constraints only): (continued)
By rank-nullity theorem, there exists $y^* \in \mathbb{R}^m$ and $s^* \in \mathbb{R}^n$ such that

$$\nabla f(x^*) = J_E(x^*)^T y^* + s^*, \qquad (2)$$

where s^* belongs to the null space of $J_E(x^*)$ (so $J_E(x^*)s^* = 0$).

Optimality conditions for constrained problems ...

Proof of Theorem 16 (for equality constraints only): (continued)

By rank-nullity theorem, there exists $y^* \in \mathbb{R}^m$ and $s^* \in \mathbb{R}^n$ such that

$$\nabla f(x^*) = J_E(x^*)^T y^* + s^*, \qquad (2)$$

where s^* belongs to the null space of $J_E(x^*)$ (so $J_E(x^*)s^* = 0$).

Taking the inner product of (2) with s^*, we deduce

$(s^*)^T \nabla f(x^*) = (s^*)^T J_E(x^*)^T y^* + (s^*)^T s^*$, or equivalently,

$$(s^*)^T \nabla f(x^*) = (y^*)^T J_E(x^*) s^* + \|s^*\|^2.$$

Optimality conditions for constrained problems ...

Proof of Theorem 16 (for equality constraints only): (continued)
By rank-nullity theorem, there exists $y^* \in \mathbb{R}^m$ and $s^* \in \mathbb{R}^n$ such that

$$\nabla f(x^*) = J_E(x^*)^T y^* + s^*, \qquad (2)$$

where s^* belongs to the null space of $J_E(x^*)$ (so $J_E(x^*)s^* = 0$).
Taking the inner product of (2) with s^*, we deduce
$(s^*)^T \nabla f(x^*) = (s^*)^T J_E(x^*)^T y^* + (s^*)^T s^*$, or equivalently,

$$(s^*)^T \nabla f(x^*) = (y^*)^T J_E(x^*) s^* + \|s^*\|^2.$$

From (1) and $J_E(x^*)s^* = 0$, we deduce $(s^*)^T \nabla f(x^*) = 0$. Thus $\|s^*\|^2 = 0$ and so $s^* = 0$. Again from (2): $\nabla f(x^*) = J_E(x^*)^T y^*$. □

Optimality conditions for constrained problems ...

Proof of Theorem 16 (for equality constraints only): (continued)
By rank-nullity theorem, there exists $y^* \in \mathbb{R}^m$ and $s^* \in \mathbb{R}^n$ such that

$$\nabla f(x^*) = J_E(x^*)^T y^* + s^*, \qquad (2)$$

where s^* belongs to the null space of $J_E(x^*)$ (so $J_E(x^*)s^* = 0$).
Taking the inner product of (2) with s^*, we deduce
$(s^*)^T \nabla f(x^*) = (s^*)^T J_E(x^*)^T y^* + (s^*)^T s^*$, or equivalently,

$$(s^*)^T \nabla f(x^*) = (y^*)^T J_E(x^*) s^* + \|s^*\|^2.$$

From (1) and $J_E(x^*)s^* = 0$, we deduce $(s^*)^T \nabla f(x^*) = 0$. Thus $\|s^*\|^2 = 0$ and so $s^* = 0$. Again from (2): $\nabla f(x^*) = J_E(x^*)^T y^*$. □

- Let (CP) with equalities only ($I = \emptyset$). Then feasible descent direction s at $x \in \Omega$ if $\nabla f(x)^T s < 0$ and $J_E(x)s = 0$.
- Let (CP). Then feasible descent direction s at $x \in \Omega$ if $\nabla f(x)^T s < 0$, $J_E(x)s = 0$ and $\nabla c_i(x)^T s \geq 0$ for all $i \in I \cap \mathcal{A}(x)$.

Constraint qualifications

- Proof of Th 16: used (first-order) Taylor to linearize f and c_i along feasible paths/perturbations $x(\alpha)$ etc. Only correct if linearized approximation covers the essential geometry of the feasible set. CQs ensure this is the case.

Constraint qualifications

- Proof of Th 16: used (first-order) Taylor to linearize f and c_i along feasible paths/perturbations $x(\alpha)$ etc. Only correct if linearized approximation covers the essential geometry of the feasible set. CQs ensure this is the case.

  Examples:

- (CP) satisfies the Slater Constraint Qualification (SCQ) $\Longleftrightarrow$ if $\exists\, x$ s.t. $c_E(x) = Ax - b = 0$ and $c_I(x) > 0$ (i.e., $c_i(x) > 0$, $i \in I$).
- (CP) satisfies the Linear Independence Constraint Qualification (LICQ) $\Longleftrightarrow$ $\nabla c_i(x)$, $i \in \mathcal{A}(x)$, are linearly independent (at relevant x).

Constraint qualifications

- Proof of Th 16: used (first-order) Taylor to linearize f and c_i along feasible paths/perturbations $x(\alpha)$ etc. Only correct if linearized approximation covers the essential geometry of the feasible set. CQs ensure this is the case.

  Examples:

- (CP) satisfies the Slater Constraint Qualification (SCQ) $\Longleftrightarrow$ if $\exists\, x$ s.t. $c_E(x) = Ax - b = 0$ and $c_I(x) > 0$ (i.e., $c_i(x) > 0$, $i \in I$).
- (CP) satisfies the Linear Independence Constraint Qualification (LICQ) $\Longleftrightarrow$ $\nabla c_i(x)$, $i \in \mathcal{A}(x)$, are linearly independent (at relevant x).

Both SCQ and LICQ fail for

$$\Omega = \{(x_1, x_2) : c_1(x) = 1 - x_1^2 - (x_2 - 1)^2 \geq 0;\ c_2(x) = -x_2 \geq 0\}.$$

$T_\Omega(x) = \{(0,0)\}$ and $\mathcal{F}(x) = \{(s_1, 0) : s_1 \in \mathbb{R}\}$. Thus $T_\Omega(x) \neq \mathcal{F}(x)$.

Constraint qualifications...

Tangent cone to Ω at x: [See Chapter 12, Nocedal & Wright]

$T_\Omega(x) = \{s : \text{limiting direction of feasible sequence}\}$ ['geometry' of Ω]

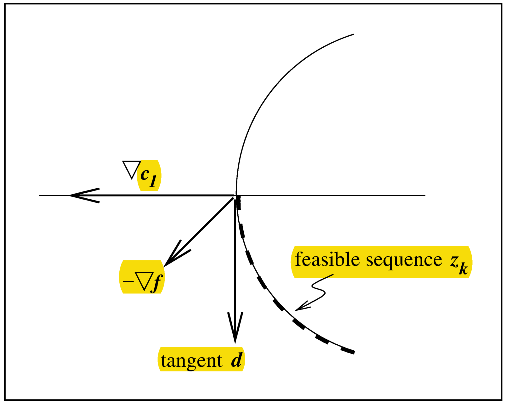
$s = \lim_{k\to\infty} \frac{z^k - x}{t^k}$ where $z^k \in \Omega$, $t^k > 0$, $t^k \to 0$ and $z^k \to x$ as $k \to \infty$.

Set of linearized feasible directions: ['algebra' of Ω]

$\mathcal{F}(x) = \{s : s^T \nabla c_i(x) = 0, i \in E;\ s^T \nabla c_i(x) \geq 0, i \in I \cap \mathcal{A}(x)\}$

Want $T_\Omega(x) = \mathcal{F}(x) \longleftarrow$[ensured if a CQ holds]

$\min_{(x_1,x_2)} x_1 + x_2$
s.t. $x_1^2 + x_2^2 - 2 = 0.$



Optimality conditions for constrained problems ...

If the constraints of (CP) are linear in the variables, no constraint qualification is required.

Theorem 17 (First order necessary conditions for linearly constrained problems) Let $(c_E, c_I)(x) := Ax - b$ in (CP). Then x^* local minimizer of (CP) $\Longrightarrow$ x^* KKT point of (CP).

Optimality conditions for constrained problems ...

If the constraints of (CP) are linear in the variables, no constraint qualification is required.

Theorem 17 (First order necessary conditions for linearly constrained problems) Let $(c_E, c_I)(x) := Ax - b$ in (CP). Then x^* local minimizer of (CP) $\Longrightarrow$ x^* KKT point of (CP).

Let $A = (A_E, A_I)$ and $b = (b_E, b_I)$ corresponding to equality and inequality constraints.
KKT conditions for linearly-constrained (CP): x^* KKT point $\Leftrightarrow$ there exists (y^*, λ^*) such that

$$\nabla f(x^*) = A_E^T y^* + A_I^T \lambda^*,$$
$$A_E x^* - b_E = 0, \quad A_I x^* - b_I \geq 0,$$
$$\lambda^* \geq 0, \quad (\lambda^*)^T (A_I x^* - b_I) = 0.$$

Optimality conditions for convex problems

(CP) is a convex programming problem if and only if $f(x)$ is a convex function, $c_i(x)$ is a concave function for all $i \in I$ and $c_E(x) = Ax - b$.

- c_i is a concave function $\Leftrightarrow$ $(-c_i)$ is a convex function.
- (CP) convex problem $\Rightarrow$ Ω is a convex set.
- (CP) convex problem $\Rightarrow$ any local minimizer of (CP) is global.

Convex constrained problems

(CP) $\min_x f(x)$ s.t. $c_E(x) = 0$ and $c_I(x) \geq 0$. ← general

(CP) convex programming problem ⟺ f convex function (over Ω) and Ω convex set.

If f convex function (over Ω) and

- $c_E(x) = Ax - b$ ← linear
- $c_i(x)$ concave function for all $i \in I$,

THEN (CP) convex programming problem.

Proof: Need to show Ω is a convex set, namely, let $x, y \in \Omega$, and show $z = (1-\alpha)x + \alpha y \in \Omega$, $\forall \alpha \in [0,1]$.

- $c_E(z) = A((1-\alpha)x + \alpha y) - b = (1-\alpha)\underbrace{[Ax - b]}_{=0} + \alpha \underbrace{[Ay - b]}_{=0} = 0$.
- $i \in I$, $c_i(z) = c_i((1-\alpha)x + \alpha y) \geq \underbrace{(1-\alpha)}_{\geq 0} \underbrace{c_i(x)}_{\geq 0} + \underbrace{\alpha}_{\geq 0} \underbrace{c_i(y)}_{\geq 0} \geq 0$.

  ($\geq$ ↓ c_i concave)

Thus $z \in \Omega$. □

Optimality conditions for convex problems

(CP) is a convex programming problem if and only if $f(x)$ is a convex function, $c_i(x)$ is a concave function for all $i \in I$ and $c_E(x) = Ax - b$.

- c_i is a concave function $\Leftrightarrow$ $(-c_i)$ is a convex function.
- (CP) convex problem $\Rightarrow$ Ω is a convex set.
- (CP) convex problem $\Rightarrow$ any local minimizer of (CP) is global.

First order necessary conditions are also sufficient for optimality when (CP) is convex.

Theorem 18. (Sufficient optimality conditions for convex problems: Let (CP) be a convex programming problem.
$\hat{x}$ KKT point of (CP) $\Longrightarrow$ $\hat{x}$ is a (global) minimizer of (CP). □

Optimality conditions for convex problems

Proof of Theorem 18.

f convex $\Longrightarrow f(x) \geq f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x})$, for all $x \in \mathbb{R}^n$. (3)

Optimality conditions for convex problems

Proof of Theorem 18.

f convex $\Longrightarrow f(x) \geq f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x})$, for all $x \in \mathbb{R}^n$. (3)

(3)+[$\nabla f(\hat{x}) = A^\top \hat{y} + \sum_{i \in I} \hat{\lambda}_i \nabla c_i(\hat{x})$] $\Longrightarrow$

Optimality conditions for convex problems

Proof of Theorem 18.

f convex $\Longrightarrow f(x) \geq f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x})$, for all $x \in \mathbb{R}^n$. (3)

(3)+$[\nabla f(\hat{x}) = A^\top \hat{y} + \sum_{i \in I} \hat{\lambda}_i \nabla c_i(\hat{x})] \Longrightarrow$

$f(x) \geq f(\hat{x}) + (A^\top \hat{y})^\top (x - \hat{x}) + \sum_{i \in I} \hat{\lambda}_i (\nabla c_i(\hat{x})^\top (x - \hat{x})),$

Optimality conditions for convex problems

Proof of Theorem 18.

f convex $\Longrightarrow f(x) \geq f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x})$, for all $x \in \mathbb{R}^n$. (3)

(3)+$[\nabla f(\hat{x}) = A^\top \hat{y} + \sum_{i\in I} \hat{\lambda}_i \nabla c_i(\hat{x})] \Longrightarrow$

$f(x) \geq f(\hat{x}) + (A^\top \hat{y})^\top (x - \hat{x}) + \sum_{i\in I} \hat{\lambda}_i (\nabla c_i(\hat{x})^\top (x - \hat{x}))$,

$f(x) \geq f(\hat{x}) + \hat{y}^\top A(x - \hat{x}) + \sum_{i\in I} \hat{\lambda}_i (\nabla c_i(\hat{x})^\top (x - \hat{x}))$ (4).

Optimality conditions for convex problems

Proof of Theorem 18.

f convex $\Longrightarrow f(x) \geq f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x})$, for all $x \in \mathbb{R}^n$. (3)

(3)+$[\nabla f(\hat{x}) = A^\top \hat{y} + \sum_{i \in I} \hat{\lambda}_i \nabla c_i(\hat{x})] \Longrightarrow$

$f(x) \geq f(\hat{x}) + (A^\top \hat{y})^\top (x - \hat{x}) + \sum_{i \in I} \hat{\lambda}_i (\nabla c_i(\hat{x})^\top (x - \hat{x}))$,

$f(x) \geq f(\hat{x}) + \hat{y}^\top A(x - \hat{x}) + \sum_{i \in I} \hat{\lambda}_i (\nabla c_i(\hat{x})^\top (x - \hat{x}))$ (4).

Let $x \in \Omega$ arbitrary $\Longrightarrow Ax = b$ and $c(x) \geq 0$.

$Ax = b$ and $A\hat{x} = b \Longrightarrow A(x - \hat{x}) = 0$. (5)

Optimality conditions for convex problems

Proof of Theorem 18.

f convex $\Longrightarrow f(x) \geq f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x})$, for all $x \in \mathbb{R}^n$. (3)

(3)+[$\nabla f(\hat{x}) = A^\top \hat{y} + \sum_{i \in I} \hat{\lambda}_i \nabla c_i(\hat{x})$] $\Longrightarrow$

$f(x) \geq f(\hat{x}) + (A^\top \hat{y})^\top (x - \hat{x}) + \sum_{i \in I} \hat{\lambda}_i (\nabla c_i(\hat{x})^\top (x - \hat{x}))$,

$f(x) \geq f(\hat{x}) + \hat{y}^\top A(x - \hat{x}) + \sum_{i \in I} \hat{\lambda}_i (\nabla c_i(\hat{x})^\top (x - \hat{x}))$ (4).

Let $x \in \Omega$ arbitrary $\Longrightarrow Ax = b$ and $c(x) \geq 0$.
$Ax = b$ and $A\hat{x} = b \Longrightarrow A(x - \hat{x}) = 0$. (5)

c_i concave $\Longrightarrow c_i(x) \leq c_i(\hat{x}) + \nabla c_i(\hat{x})^\top (x - \hat{x})$.

$\Longrightarrow \nabla c_i(\hat{x})^\top (x - \hat{x}) \geq c_i(x) - c_i(\hat{x})$.

$\Longrightarrow \hat{\lambda}_i (\nabla c_i(\hat{x})^\top (x - \hat{x})) \geq \hat{\lambda}_i (c_i(x) - c_i(\hat{x})) = \hat{\lambda}_i c_i(x) \geq 0$,

since $\hat{\lambda} \geq 0$, $\hat{\lambda}_i c_i(x) = 0$ and $c(x) \geq 0$.

Optimality conditions for convex problems

Proof of Theorem 18.

f convex $\Longrightarrow f(x) \geq f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x})$, for all $x \in \mathbb{R}^n$. (3)

(3)+[$\nabla f(\hat{x}) = A^\top \hat{y} + \sum_{i \in I} \hat{\lambda}_i \nabla c_i(\hat{x})$] $\Longrightarrow$

$f(x) \geq f(\hat{x}) + (A^\top \hat{y})^\top (x - \hat{x}) + \sum_{i \in I} \hat{\lambda}_i (\nabla c_i(\hat{x})^\top (x - \hat{x}))$,

$f(x) \geq f(\hat{x}) + \hat{y}^\top A(x - \hat{x}) + \sum_{i \in I} \hat{\lambda}_i (\nabla c_i(\hat{x})^\top (x - \hat{x}))$ (4).

Let $x \in \Omega$ arbitrary $\Longrightarrow Ax = b$ and $c(x) \geq 0$.
$Ax = b$ and $A\hat{x} = b \Longrightarrow A(x - \hat{x}) = 0$. (5)

c_i concave $\Longrightarrow c_i(x) \leq c_i(\hat{x}) + \nabla c_i(\hat{x})^\top (x - \hat{x})$.
$\Longrightarrow \nabla c_i(\hat{x})^\top (x - \hat{x}) \geq c_i(x) - c_i(\hat{x})$.
$\Longrightarrow \hat{\lambda}_i (\nabla c_i(\hat{x})^\top (x - \hat{x})) \geq \hat{\lambda}_i (c_i(x) - c_i(\hat{x})) = \hat{\lambda}_i c_i(x) \geq 0$,
since $\hat{\lambda} \geq 0$, $\hat{\lambda}_i c_i(x) = 0$ and $c(x) \geq 0$.
Thus, from (4) and (5), $f(x) \geq f(\hat{x})$. $\square$

Example: Optimality conditions for QP problems

A Quadratic Programming (QP) problem has the form

$$\text{minimize}_{x\in\mathbb{R}^n}\ c^\top x + \frac{1}{2}x^\top H x \ \text{ s. t. }\ Ax = b,\ \tilde{A}x \geq \tilde{b}. \quad \text{(QP)}$$

H symm. pos. semidefinite $\Longrightarrow$ (QP) convex problem.

The KKT conditions for (QP):

$\hat{x}$ KKT point of (QP) $\iff \exists\,(\hat{y},\hat{\lambda}) \in \mathbb{R}^m \times \mathbb{R}^p$ such that

$$\begin{aligned} &H\hat{x} + c = A^\top \hat{y} + \tilde{A}^\top \hat{\lambda}, \\ &A\hat{x} = b,\ \tilde{A}\hat{x} \geq \tilde{b}, \\ &\hat{\lambda} \geq 0,\ \hat{\lambda}^\top(\tilde{A}\hat{x} - \tilde{b}) = 0. \end{aligned}$$

- "An example of a nonlinear constrained problem" is convex; removing the constraint $x_2 - x_1^2 \geq 0$ makes it a convex (QP).

Example: Duality theory for QP problems

For simplicity, let $A := 0$ and $H \succ 0$ in (QP): primal problem:

$$\text{minimize}_{x \in \mathbb{R}^n}\ c^\top x + \frac{1}{2} x^\top H x \ \text{ s. t. } \tilde{A}x \geq \tilde{b}. \quad \text{(QP)}$$

The KKT conditions for (QP):

$$H\hat{x} + c = \tilde{A}^\top \hat{\lambda},$$
$$\tilde{A}\hat{x} \geq \tilde{b},$$
$$\hat{\lambda} \geq 0,\ \hat{\lambda}^\top (\tilde{A}\hat{x} - \tilde{b}) = 0.$$

Dual problem:

$$\text{maximize}_{(x,\lambda)} - \frac{1}{2} x^T H x + \tilde{b}^T \lambda \ \text{s.t.} \ -Hx + \tilde{A}^\top \lambda = c \ \text{and} \ \lambda \geq 0.$$

Optimal value of primal pb=optimal value of dual pb (provided they exist).

Optimality conditions for nonconvex problems

- When (CP) is not convex, the KKT conditions are not in general sufficient for optimality
$\longrightarrow$ need positive definite Hessian of the Lagrangian function along "feasible" directions.

- More on second-order optimality conditions later on.

Second-order optimality conditions

- When (CP) is not convex, the KKT conditions are not in general sufficient for optimality.

- Assume some CQ holds. Then at a given point x^*: the set of feasible directions for (CP) at x^*:

$$\mathcal{F}(x^*) = \left\{s : J_E(x^*)s = 0,\ s^T\nabla c_i(x^*) \geq 0, i \in \mathcal{A}(x^*) \cap I\right\}.$$

- If x^* is a KKT point, then for any $s \in \mathcal{F}(x^*)$,

$$\begin{aligned} s^T\nabla f(x^*) &= s^T J_E(x^*)^T y^* + \textstyle\sum_{i\in\mathcal{A}(x^*)\cap I} \lambda_i s^T \nabla c_i(x^*) \\ &= (J_E(x^*)s)^T y^* + \textstyle\sum_{i\in\mathcal{A}(x^*)\cap I} \lambda_i s^T \nabla c_i(x^*) \\ &= \textstyle\sum_{i\in\mathcal{A}(x^*)\cap I} \lambda_i s^T \nabla c_i(x^*) \geq 0. \quad (6) \end{aligned}$$

Second-order optimality conditions...

- If x^* is a KKT point, then for any $s \in \mathcal{F}(x^*)$, either

$$s^T \nabla f(x^*) > 0$$

$\longrightarrow$ so f can only increase and stay feasible along s

$$\text{or} \quad s^T \nabla f(x^*) = 0$$

$\longrightarrow$ cannot decide from 1st order info if f increases or not along such s.

From (6), we see that the directions of interest are:
$J_E(x^*)s = 0$ and $s^T \nabla c_i(x^*) = 0$, $\forall i \in \mathcal{A}(x^*) \cap I$ with $\lambda_i > 0$.

$$F(\lambda^*) = \{s \in \mathcal{F}(x^*) : s^T \nabla c_i(x^*) = 0, \ \forall i \in \mathcal{A}(x^*) \cap I \text{ with } \lambda_i^* > 0\},$$

where λ^* is a Lagrange multiplier of the inequality constraints.

Then note that $s^T \nabla f(x^*) = 0$ for all $s \in F(\lambda^*)$.

Second-order optimality conditions ...

Theorem 19 (Second-order necessary conditions)
Let some CQ hold for (CP). Let x^* be a local minimizer of (CP), and (y^*, λ^*) Lagrange multipliers of the KKT conditions at x^*. Then

$$s^T \nabla^2_{xx} \mathcal{L}(x^*, y^*, \lambda^*) s \geq 0 \text{ for all } s \in F(\lambda^*),$$

where $\mathcal{L}(x, y, \lambda) = f(x) - y^T c_E(x) - \lambda^T c_I(x)$ is the Lagrangian function and so
$\nabla^2_{xx} \mathcal{L}(x, y, \lambda) = \nabla^2 f(x) - \sum_{j=1}^m y_j \nabla^2 c_j(x) - \sum_{i=1}^p \lambda_i c_i(x)]$.

Second-order optimality conditions ...

Theorem 19 (Second-order necessary conditions)
Let some CQ hold for (CP). Let x^* be a local minimizer of (CP), and (y^*, λ^*) Lagrange multipliers of the KKT conditions at x^*. Then

$$s^T \nabla^2_{xx} \mathcal{L}(x^*, y^*, \lambda^*) s \geq 0 \text{ for all } s \in F(\lambda^*),$$

where $\mathcal{L}(x, y, \lambda) = f(x) - y^T c_E(x) - \lambda^T c_I(x)$ is the Lagrangian function and so
$\nabla^2_{xx} \mathcal{L}(x, y, \lambda) = \nabla^2 f(x) - \sum_{j=1}^m y_j \nabla^2 c_j(x) - \sum_{i=1}^p \lambda_i c_i(x)]$.

Theorem 20 (Second-order sufficient conditions)
Assume that x^* is a feasible point of (CP) and (y^*, λ^*) are such that the KKT conditions are satisfied by (x^*, y^*, λ^*). If

$$s^T \nabla^2_{xx} \mathcal{L}(x^*, y^*, \lambda^*) s > 0 \text{ for all } s \in F(\lambda^*),\ s \neq 0,$$

then x^* is a local minimizer of (CP). [See proofs in Nocedal & Wright]

Some simple approaches for solving (CP)

Equality-constrained problems: direct elimination (a simple approach that may help/work sometimes; cannot be automated in general)

Method of Lagrange multipliers: using the KKT and second order conditions to find minimizers (again, cannot be automated in general)

[see Pb Sheet 4]