Math 6880/7875: Advanced Optimization
Background and Review: Optimization

Akil Narayan[1]

[1]Department of Mathematics, and Scientific Computing and Imaging (SCI) Institute
University of Utah

January 11, 2022

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

"Prerequisites"

Several topics are background for this course:

- (Numerical) linear algebra
- Probability/statistics
- "Basic" optimization knowledge

We'll spend some time briefly reviewing portions of these.

Optimization

General optimization problem:

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \end{aligned}$$

$$S := \left\{ x \mid g_i(x) \leqslant 0, \; i \in [m] \right\}$$

- $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ is the *optimization* or *design* variable
- $f : \mathbb{R}^n \to \mathbb{R}$ is the *objective* function
- $g_i : \mathbb{R}^n \to \mathbb{R}$, $i \in [m]$, are the *constraints*
- S is the *feasible* set
- We will always consider $n < \infty$, but we will occasionally allow $m \uparrow \infty$
- If $m > 0$, the problem is **constrained**; otherwise it is **unconstrained**

Optimization

General optimization problem:

$$\text{minimize } f(x)$$
$$\text{subject to } x \in S$$
$$S := \{x \mid g_i(x) \leqslant 0, \; i \in [m]\}$$

{1, 2, .. m}

- $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ is the *optimization* or *design* variable
- $f : \mathbb{R}^n \to \mathbb{R}$ is the *objective* function
- $g_i : \mathbb{R}^n \to \mathbb{R}$, $i \in [m]$, are the *constraints*
- S is the *feasible* set
- We will always consider $n < \infty$, but we will occasionally allow $m \uparrow \infty$
- If $m > 0$, the problem is **constrained**; otherwise it is **unconstrained**

(implicitly $S = \mathbb{R}^n$)

Optimization

General optimization problem:

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \end{aligned}$$

$$S := \left\{ x \mid g_i(x) \leqslant 0, \ i \in [m] \right\}$$

Solutions to optimization problems have their own taxonomy and properties:

- A point $x \in \mathbb{R}^n$ is **feasible** if $x \in S$.
- A point $x^* \in \mathbb{R}^n$ is a ("global") **solution**, **optimum**, or **optimal point** if $f(x^*) \leqslant f(x)$ for every $x \in S$.
- A point $x^* \in \mathbb{R}^n$ is a "local" solution, optimum, or optimal point if $\exists \epsilon > 0$ such that $f(x^*) \leqslant f(x)$ for all $x \in B_\epsilon(x^*) \cap S$, where

$$B_\epsilon(x^*) \cap S := \left\{ x \in S \mid \|x - x^*\| < \epsilon \right\}.$$

Optimization

General optimization problem:

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } x \in S \\ &S := \left\{ x \mid g_i(x) \leqslant 0,\ i \in [m] \right\} \end{aligned}$$

Solutions to optimization problems have their own taxonomy and properties:

- A point $x \in \mathbb{R}^n$ is **feasible** if $x \in S$.
- A point $x^* \in \mathbb{R}^n$ is a ("global") **solution**, **optimum**, or **optimal point** if $f(x^*) \leqslant f(x)$ for every $x \in S$.
- A point $x^* \in \mathbb{R}^n$ is a "local" solution, optimum, or optimal point if $\exists \epsilon > 0$ such that $f(x^*) \leqslant f(x)$ for all $x \in B_\epsilon(x^*) \cap S$, where

$$B_\epsilon(x^*) \cap S := \left\{ x \in S \mid \|x - x^*\| < \epsilon \right\}.$$

Optimization

General optimization problem:

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \end{aligned}$$

$$S := \{x \mid g_i(x) \leqslant 0, \; i \in [m]\}$$

Solutions to optimization problems have their own taxonomy and properties:

- "Optimum"/"extremum" $\longleftrightarrow$ "maximum"/"minimum", as appropriate
- Maximization of f is minimization of $-f$
- Optimization problems can have zero, one, or many solutions. Which of these is true is rarely obvious.

Generally our goal is to find/compute an optimal solution. A local one could suffice.

- equality constraints are doable: $h(x) = 0$

$$g_1(x) = h(x)$$
$$g_2(x) = -h(x)$$

Optimization

General optimization problem:

$$\text{minimize } f(x)$$
$$\text{subject to } x \in S$$
$$S := \{x \mid g_i(x) \leqslant 0,\ i \in [m]\}$$

Solutions to optimization problems have their own taxonomy and properties:

- "Optimum"/"extremum" $\longleftrightarrow$ "maximum"/"minimum", as appropriate
- Maximization of f is minimization of $-f$
- Optimization problems can have zero, one, or many solutions. Which of these is true is rarely obvious.

Generally our goal is to find/compute an optimal solution. A local one could suffice.

Optimization

General optimization problem:

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \end{aligned}$$

$$S := \left\{ x \mid g_i(x) \leqslant 0, \; i \in [m] \right\}$$

Solutions to optimization problems have their own taxonomy and properties:

- "Optimum"/"extremum" $\longleftrightarrow$ "maximum"/"minimum", as appropriate
- Maximization of f is minimization of $-f$
- Optimization problems can have zero, one, or many solutions. Which of these is true is rarely obvious.

Generally our goal is to find/compute an optimal solution. A local one could suffice.

Simple examples

Example

One, unique solution

$$\min_{x \in \mathbb{R}} |x|$$
$$\text{subject to } \; x \geqslant -1$$

Example

No solutions – infeasible

$$\min_{x \in \mathbb{R}} x^2$$
$$\text{subject to } \; |x| \leqslant -1$$

Simple examples

Example

One, unique solution

$$\min_{x \in \mathbb{R}} |x|$$

$$\text{subject to } \; x \geqslant -1$$

Example

No solutions – infeasible

$$\min_{x \in \mathbb{R}} x^2$$

$$\text{subject to } \; |x| \leqslant -1$$

Simple examples

Example

No solutions – unbounded

$$\max_{x \in \mathbb{R}} x^2$$

Example

Many solutions

$$\min_{x \in \mathbb{R}} \sin x$$

$$\text{subject to } \; |x| \geqslant \pi$$

Simple examples

Example

No solutions – unbounded

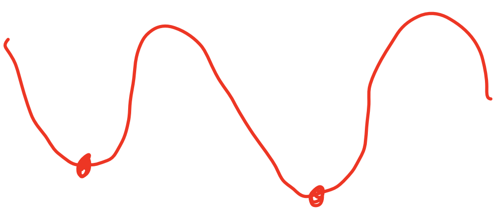
$$\max_{x \in \mathbb{R}} x^2$$

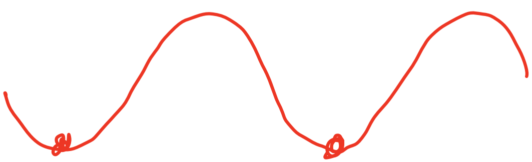
Example

Many solutions

$$\min_{x \in \mathbb{R}} \; \sin x$$

$$\text{subject to} \;\; |x| \geqslant \pi$$





Ascertaining optimality

In some special cases, with some effort, one can conclude global optimality.

- Direct methods – Analytically prove global optimality. (E.g., $f(x) = (x^2 - 1)^{10}$)
- Quadratic functions – f is quadratic with a positive-definite Hessian.
- Coercive functions – Global optimality in some ball B, and show that that f outside B dominates f inside B.
- Globally convex functions – Ensures that local minima are global minima.

Caveats:

- All the above are "easier" for unconstrained optimization, and become much more technical and difficult for constrained optimization.
- *Global* optimality requires some *global* knowledge of the objective and constraints.
- In high dimensions (n large), globally certifying any property of generic functions is hard.

The depressing fact of life: without relatively strong assumptions, local optimality is the best we know how to establish.

Ascertaining optimality

In some special cases, with some effort, one can conclude global optimality.

- Direct methods – Analytically prove global optimality. (E.g., $f(x) = (x^2 - 1)^{10}$)
- Quadratic functions – f is quadratic with a positive-definite Hessian.
- Coercive functions – Global optimality in some ball B, and show that that f outside B dominates f inside B.
- Globally convex functions – Ensures that local minima are global minima.

Caveats:

- All the above are "easier" for unconstrained optimization, and become much more technical and difficult for constrained optimization.
- *Global* optimality requires some *global* knowledge of the objective and constraints.
- In high dimensions (n large), globally certifying any property of generic functions is hard.

The depressing fact of life: without relatively strong assumptions, local optimality is the best we know how to establish.

Ascertaining optimality

In some special cases, with some effort, one can conclude global optimality.

- Direct methods – Analytically prove global optimality. (E.g., $f(x) = (x^2 - 1)^{10}$)
- Quadratic functions – f is quadratic with a positive-definite Hessian.
- Coercive functions – Global optimality in some ball B, and show that that f outside B dominates f inside B.
- Globally convex functions – Ensures that local minima are global minima.

Caveats:

- All the above are "easier" for unconstrained optimization, and become much more technical and difficult for constrained optimization.
- *Global* optimality requires some *global* knowledge of the objective and constraints.
- In high dimensions (n large), globally certifying any property of generic functions is hard.

The depressing fact of life: without relatively strong assumptions, local optimality is the best we know how to establish.

Local optimality

There are a handful of **optimality conditions** that can be sufficient and/or necessary to determine local optimality.

- First-order optimality conditions – conditions involving the gradients of f and/or g_i.
- Second-order optimality conditions – conditions involving Hessians. Less computationally useful due to complexity/storage requirements.

It's much easier to discuss these conditions for unconstrained optimization first.

Local optimality

There are a handful of **optimality conditions** that can be sufficient and/or necessary to determine local optimality.

- First-order optimality conditions – conditions involving the gradients of f and/or g_i.
- Second-order optimality conditions – conditions involving Hessians. Less computationally useful due to complexity/storage requirements.

It's much easier to discuss these conditions for unconstrained optimization first.

First-order local optimality: unconstrained optimization

Unconstrained optimization:

$$\text{minimize } f(x),$$

which implicitly allows $x \in \mathbb{R}^n$. (I.e., the feasible set is $S = \mathbb{R}^n$.)

The simplest first-order local optimality condition is a necessary one.

Theorem

If $f \in C^1(\mathbb{R}^n)$, then x^* is a local minimum only if $\nabla f(x^*) = 0$.

Proof.

Fix $i \in [n]$. Let x_i be free, but fix $x_{\setminus i} = x^*_{\setminus i}$.
The resulting one-dimensional function f_i must have a local minimum at $x_i = x_i^*$, where its univariate derivative vanishes.
Repeat for every $i \implies \nabla f(x^*) = 0$. □

Notes:

- $\nabla f(x^*) = 0$ is not sufficient to conclude anything.
- $\nabla f(x^*) = 0$ is also a necessary condition for local minimization over $S \subset \mathbb{R}^n$ so long as $x^* \in \operatorname{int}(S)$.

First-order local optimality: unconstrained optimization

Unconstrained optimization:

$$\text{minimize } f(x),$$

which implicitly allows $x \in \mathbb{R}^n$. (I.e., the feasible set is $S = \mathbb{R}^n$.)

The simplest first-order local optimality condition is a necessary one.

Theorem

If $f \in C^1(\mathbb{R}^n)$, then x^* is a local minimum only if $\nabla f(x^*) = 0$.

Proof.

Fix $i \in [n]$. Let x_i be free, but fix $x_{\backslash i} = x^*_{\backslash i}$.
The resulting one-dimensional function f_i must have a local minimum at $x_i = x_i^*$, where its univariate derivative vanishes.
Repeat for every $i \Longrightarrow \nabla f(x^*) = 0$. □

Notes:

- $\nabla f(x^*) = 0$ is not sufficient to conclude anything.
- $\nabla f(x^*) = 0$ is also a necessary condition for local minimization over $S \subset \mathbb{R}^n$ so long as $x^* \in \text{int}(S)$.

First-order local optimality: unconstrained optimization

Unconstrained optimization:

$$\text{minimize } f(x),$$

which implicitly allows $x \in \mathbb{R}^n$. (I.e., the feasible set is $S = \mathbb{R}^n$.)

The simplest first-order local optimality condition is a necessary one.

Theorem

If $f \in C^1(\mathbb{R}^n)$, then x^* is a local minimum only if $\nabla f(x^*) = 0$.

Proof.

Fix $i \in [n]$. Let x_i be free, but fix $x_{\setminus i} = x^*_{\setminus i}$.
The resulting one-dimensional function f_i must have a local minimum at $x_i = x^*_i$, where its univariate derivative vanishes.
Repeat for every $i \Longrightarrow \nabla f(x^*) = 0$. □

Notes:

- $\nabla f(x^*) = 0$ is not sufficient to conclude anything.
- $\nabla f(x^*) = 0$ is also a necessary condition for local minimization over $S \subset \mathbb{R}^n$ so long as $x^* \in \text{int}(S)$.

Stationary points and definite matrices

Given $f : \mathbb{R}^n \to \mathbb{R}$ that is differentiable, a point x satisfying $\nabla f(x) = 0$ is a **stationary point**.

- Stationary points can be local/global minima.
- Stationary points can be local/global maxima.
- Stationary points can be saddle points (neither a maximum nor a minimum).

Many computational methods attempt to compute stationary points, even if we can't classify the result.

Stationary points are not necessarily easy to compute....

Stationary points and definite matrices

Given $f : \mathbb{R}^n \to \mathbb{R}$ that is differentiable, a point x satisfying $\nabla f(x) = 0$ is a **stationary point**.

- Stationary points can be local/global minima.
- Stationary points can be local/global maxima.
- Stationary points can be saddle points (neither a maximum nor a minimum).

Many computational methods attempt to compute stationary points, even if we can't classify the result.

Stationary points are not necessarily easy to compute....

Stationary points and definite matrices

Given $f : \mathbb{R}^n \to \mathbb{R}$ that is differentiable, a point x satisfying $\nabla f(x) = 0$ is a **stationary point**.

- Stationary points can be local/global minima.
- Stationary points can be local/global maxima.
- Stationary points can be saddle points (neither a maximum nor a minimum).

Many computational methods attempt to compute stationary points, even if we can't classify the result.

Stationary points are not necessarily easy to compute....

Definite matrices

Quadratic classification of matrices are needed for second-order conditions:

- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive definite** if $x^T A x > 0$ for every $x \in \mathbb{R}^n \backslash \{0\}$.
  We write $A \succ 0$.
  (Equivalently, the inequality holds for all x with unit norm.)
- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive semi-definite** if $x^T A x \geqslant 0$ for every $x \in \mathbb{R}^n \backslash \{0\}$.
  We write $A \succeq 0$.
- Similar definitions for negative definite, and negative semi-definite. ($A \prec 0$, $A \preceq 0$, respectively)
- Matrices that are not positive/negative definite are **indefinite**.

We will, in particular, utilize these characterizations for Hessian matrices, $\nabla^2 f$.

Definite matrices

Quadratic classification of matrices are needed for second-order conditions:

- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive definite** if $x^T A x > 0$ for every $x \in \mathbb{R}^n \backslash \{0\}$.
  We write $A \succ 0$.
  (Equivalently, the inequality holds for all x with unit norm.)
- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive semi-definite** if $x^T A x \geqslant 0$ for every $x \in \mathbb{R}^n \backslash \{0\}$.
  We write $A \succeq 0$.
- Similar definitions for negative definite, and negative semi-definite. ($A \prec 0$, $A \preceq 0$, respectively)
- Matrices that are not positive/negative definite are **indefinite**.

We will, in particular, utilize these characterizations for Hessian matrices, $\nabla^2 f$.

Definite matrices

Quadratic classification of matrices are needed for second-order conditions:

- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive definite** if $x^T A x > 0$ for every $x \in \mathbb{R}^n \backslash \{0\}$.
  We write $A \succ 0$.
  (Equivalently, the inequality holds for all x with unit norm.)
- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive semi-definite** if $x^T A x \geqslant 0$ for every $x \in \mathbb{R}^n \backslash \{0\}$.
  We write $A \succeq 0$.
- Similar definitions for negative definite, and negative semi-definite. ($A \prec 0$, $A \preceq 0$, respectively)
- Matrices that are not positive/negative definite are **indefinite**.

We will, in particular, utilize these characterizations for Hessian matrices, $\nabla^2 f$.

Definite matrices

Quadratic classification of matrices are needed for second-order conditions:

- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive definite** if $x^T A x > 0$ for every $x \in \mathbb{R}^n \backslash \{0\}$.
  We write $A \succ 0$.
  (Equivalently, the inequality holds for all x with unit norm.)
- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive semi-definite** if $x^T A x \geqslant 0$ for every $x \in \mathbb{R}^n \backslash \{0\}$.
  We write $A \succeq 0$.
- Similar definitions for negative definite, and negative semi-definite. ($A \prec 0$, $A \preceq 0$, respectively)
- Matrices that are not positive/negative definite are **indefinite**.

We will, in particular, utilize these characterizations for Hessian matrices, $\nabla^2 f$.

Second-order local optimality: unconstrained optimization

Unconstrained optimization:

$$\text{minimize } f(x),$$

An initial, necessary second-order condition:

Theorem

Assume $f \in C^2(\mathbb{R}^n)$. If x^* is a local minimum, then $\nabla f^2(x^*) \geq 0$.

$\nabla^2 f(x^*)$

Proof sketch.

Take second-order Taylor expansion of f around x^*,

$$f(x) \approx f(x^*) + \nabla f(x^*)^T (x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^*)(x - x^*).$$

x^* must be a stationary point for f, and the above holds for all x sufficiently close to x^*. □

As before, this necessary condition holds if x^* is in the interior of a feasible set for a constrained optimization problem.

$f(x) = x^3, \quad x^* = 0, \quad f''(x^*) = 0 \not\Rightarrow x^*$ is a local min

Second-order local optimality: unconstrained optimization

Unconstrained optimization:

$$\text{minimize } f(x),$$

An initial, necessary second-order condition:

Theorem

Assume $f \in C^2(\mathbb{R}^n)$. If x^* is a local minimum, then $\nabla f^2(x^*) \succeq 0$.

Proof sketch.

Take second-order Taylor expansion of f around x^*,

$$f(x) \approx f(x^*) + \nabla f(x^*)^T(x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^*)(x - x^*).$$

x^* must be a stationary point for f, and the above holds for all x sufficiently close to x^*. □

As before, this necessary condition holds if x^* is in the interior of a feasible set for a constrained optimization problem.

Sufficient second-order optimality

Unconstrained optimization:

$$\text{minimize } f(x),$$

Theorem

Assume $f \in C^2(\mathbb{R}^n)$. *If* x^* *is a stationary point for* f *and* $\nabla f^2(x^*) \succ 0$, *then* x^* *is a local minimum.*

Proof sketch.

Another second-order Taylor expansion of f for x close to x^*:

$$f(x) \approx f(x^*) + \nabla f(x^*)^T (x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^*)(x - x^*).$$

x^* is a stationary point for f, and $(x - x^*)^T \nabla^2 f(x^*)(x - x^*) > 0$. □

Optimality for constrained optimization

Constrained optimization:

$$\text{minimize } f(x)$$
$$\text{subject to } x \in S$$
$$S := \left\{x \mid g_i(x) \leqslant 0,\ i \in [m]\right\}$$

One major complication with constrained vs. unconstrained optimization: local optima on the boundary of the feasible set must be handled with care.

Given a local optimum x^*, we divide $[m]$ into *active* and *inactive* constraint sets:

- $\mathcal{A}(x^*) = \left\{i \in [m] \mid g_i(x) = 0\right\}$
- $\mathcal{I}(x^*) = \left\{i \in [m] \mid g_i(x) < 0\right\}$

No feasible descent: at a local minimum x^*, we cannot find a direction for travel that simultaneously decreases f and all element of $g_{\mathcal{A}(x^*)}$.

Optimality for constrained optimization

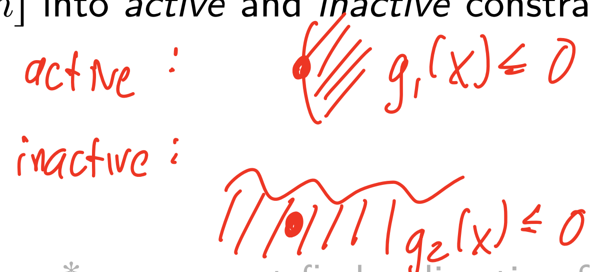
Constrained optimization:

$$\text{minimize } f(x)$$
$$\text{subject to } x \in S$$
$$S := \{x \mid g_i(x) \leqslant 0,\ i \in [m]\}$$

One major complication with constrained vs. unconstrained optimization: local optima on the boundary of the feasible set must be handled with care.

Given a local optimum x^*, we divide $[m]$ into *active* and *inactive* constraint sets:

- $\mathcal{A}(x^*) = \{i \in [m] \mid g_i(x) = 0\}$
- $\mathcal{I}(x^*) = \{i \in [m] \mid g_i(x) < 0\}$



No feasible descent: at a local minimum x^*, we cannot find a direction for travel that simultaneously decreases f and all element of $g_{\mathcal{A}(x^*)}$.

Optimality for constrained optimization

Constrained optimization:

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \end{aligned}$$

$$S := \{x \mid g_i(x) \leqslant 0, \; i \in [m]\}$$

One major complication with constrained vs. unconstrained optimization: local optima on the boundary of the feasible set must be handled with care.

Given a local optimum x^*, we divide $[m]$ into *active* and *inactive* constraint sets:

- $\mathcal{A}(x^*) = \{i \in [m] \mid g_i(x) = 0\}$
- $\mathcal{I}(x^*) = \{i \in [m] \mid g_i(x) < 0\}$

No feasible descent: at a local minimum x^*, we cannot find a direction for travel that simultaneously decreases f and all element of $g_{\mathcal{A}(x^*)}$.

Constraint qualification

$$\text{minimize } f(x)$$
$$\text{subject to } x \in S$$
$$S := \{x \mid g_i(x) \leqslant 0, \ i \in [m]\}$$

To state (useful versions of) first-order optimality, we require an additional concept.

A local minimum x^* satisfies the **linear independence constraint qualification** (LICQ) condition if

$$\{\nabla g_i(x^*)\}_{i \in \mathcal{A}(x^*)},$$

is a collection of linearly independent vectors.

The LICQ condition is used to strengthen necessary local optimality conditions.

Constraint qualification

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \end{aligned}$$
$$S := \left\{ x \;\middle|\; g_i(x) \leqslant 0,\ i \in [m] \right\}$$

To state (useful versions of) first-order optimality, we require an additional concept.

A local minimum x^* satisfies the **linear independence constraint qualification** (LICQ) condition if

$$\left\{ \nabla g_i(x^*) \right\}_{i \in \mathcal{A}(x^*)},$$

is a collection of linearly independent vectors.

The LICQ condition is used to strengthen necessary local optimality conditions.

Constraint qualification

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \end{aligned}$$

$$S := \left\{ x \,\middle|\, g_i(x) \leqslant 0,\ i \in [m] \right\}$$

To state (useful versions of) first-order optimality, we require an additional concept.

A local minimum x^* satisfies the **linear independence constraint qualification** (LICQ) condition if

$$\left\{ \nabla g_i(x^*) \right\}_{i \in \mathcal{A}(x^*)},$$

is a collection of linearly independent vectors.

The LICQ condition is used to strengthen necessary local optimality conditions.

Constrained optimization: first-order optimality

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \end{aligned}$$

$$S := \{x \mid g_i(x) \leqslant 0,\ i \in [m]\}$$

Theorem (Karush-Kuhn-Tucker)

Assume both f and g_i are $C^1(S)$ for every $i \in [m]$. Assume x^* is a local minimum of the above optimization problem that satisfies the LICQ condition. Then there exists a $\lambda \in \mathbb{R}^m$ such that (x^*, λ) satisfies,

$$\nabla f(x^*) + \sum_{i \in [m]} \lambda_i \nabla g_i(x^*) = 0 \qquad \textit{(Stationarity)}$$

$$\lambda_i g_i(x^*) = 0,\ i \in [m] \qquad \textit{(Complementary Slackness)}$$

$$g_i(x^*) \leqslant 0,\ i \in [m] \qquad \textit{(Primal feasibility)}$$

$$\lambda_i \geqslant 0,\ i \in [m] \qquad \textit{(Dual feasibility)}$$

The above are called the **KKT conditions**, and any point (x, λ) satisfying these conditions (even if x is not a local minimum) is a **KKT point**.

Equality constraint (1):

$$h(x)=0 \longrightarrow \left.\begin{array}{l} g_1(x)=h \\ g_2(x)=-h \end{array}\right\} \Rightarrow \begin{array}{l} g_1(x)\le 0 \\ g_2(x)\le 0 \end{array} \Rightarrow h(x)=0$$

$$\Downarrow$$

$$g_1(x)=0$$

$$g_2(x)=0$$

KKT conditions proof idea

One proof of the KKT conditions is a combination of three ideas/techniques:

- *No feasible descent*: We cannot find any direction $\boldsymbol{d} \in \mathbb{R}^n$ such that all the following hold:

$$\begin{aligned} \nabla f(x^*)^T \boldsymbol{d} &< 0 \\ \nabla g_i(x^*)^T \boldsymbol{d} &< 0, & i \in \mathcal{A}(x^*) \end{aligned}$$

- *Theorems of the alternative*: If there does not exist a $\boldsymbol{d}$ satisfying the above, then there must exist a $\lambda \in \mathbb{R}^{m+1}$ with positive components satisfying

$$\lambda_0 \nabla f(x^*) + \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla g_i(x^*) = 0, \qquad \lambda_i = 0, \;\; i \in \mathcal{I}(x^*)$$

  In particular, the above exercises *Gordan's Theorem of the alternative*.

- *Constraint qualification*: Under LICQ, we can set $\lambda_0 = 1$ without loss.

KKT conditions proof idea

One proof of the KKT conditions is a combination of three ideas/techniques:

- *No feasible descent*: We cannot find any direction $\boldsymbol{d} \in \mathbb{R}^n$ such that all the following hold:

$$
\begin{aligned}
\nabla f(x^*)^T \boldsymbol{d} &< 0 \\
\nabla g_i(x^*)^T \boldsymbol{d} &< 0, \qquad\qquad i \in \mathcal{A}(x^*)
\end{aligned}
$$

- *Theorems of the alternative*: If there does not exist a $\boldsymbol{d}$ satisfying the above, then there must exist a $\lambda \in \mathbb{R}^{m+1}$ with positive components satisfying

$$
\lambda_0 \nabla f(x^*) + \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla g_i(x^*) = 0, \qquad \lambda_i = 0, \;\; i \in \mathcal{I}(x^*)
$$

  In particular, the above exercises *Gordan's Theorem of the alternative*.

- *Constraint qualification*: Under LICQ, we can set $\lambda_0 = 1$ without loss.

KKT conditions proof idea

One proof of the KKT conditions is a combination of three ideas/techniques:

- *No feasible descent*: We cannot find any direction $\boldsymbol{d} \in \mathbb{R}^n$ such that all the following hold:

$$\begin{aligned} \nabla f(x^*)^T \boldsymbol{d} &< 0 \\ \nabla g_i(x^*)^T \boldsymbol{d} &< 0, & i \in \mathcal{A}(x^*) \end{aligned}$$

- *Theorems of the alternative*: If there does not exist a $\boldsymbol{d}$ satisfying the above, then there must exist a $\lambda \in \mathbb{R}^{m+1}$ with positive components satisfying

$$\lambda_0 \nabla f(x^*) + \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla g_i(x^*) = 0, \qquad \lambda_i = 0, \;\; i \in \mathcal{I}(x^*)$$

  In particular, the above exercises *Gordan's Theorem of the alternative*.

- *Constraint qualification*: Under LICQ, we can set $\lambda_0 = 1$ without loss.

KKT conditions

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \end{aligned}$$

$$S := \{x \mid g_i(x) \leqslant 0, \; i \in [m]\}$$

The KKT conditions

- are necessary first-order optimality conditions
- are a lot more complicated than unconstrained optimization conditions
- extend to equality constraints (associated dual inequality constraints are always active)
- are also necessary with other (typically weaker) types of constraint qualification
- technically don't require constraint qualification (Fritz-John conditions), but this makes them less useful
- are really only explicitly used to analytically solve problems, but serve as the basis for some algorithms

KKT conditions

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in S \end{aligned}$$

$$S := \left\{ x \mid g_i(x) \leqslant 0, \ i \in [m] \right\}$$

The KKT conditions

- are necessary first-order optimality conditions
- are a lot more complicated than unconstrained optimization conditions
- extend to equality constraints (associated dual inequality constraints are always active)
- are also necessary with other (typically weaker) types of constraint qualification
- technically don't require constraint qualification (Fritz-John conditions), but this makes them less useful
- are really only explicitly used to analytically solve problems, but serve as the basis for some algorithms

KKT conditions

$$\text{minimize } f(x)$$
$$\text{subject to } x \in S$$
$$S := \{x \mid g_i(x) \leqslant 0,\ i \in [m]\}$$

The KKT conditions

- are necessary first-order optimality conditions
- are a lot more complicated than unconstrained optimization conditions
- extend to equality constraints (associated dual inequality constraints are always active)
- are also necessary with other (typically weaker) types of constraint qualification
- technically don't require constraint qualification (Fritz-John conditions), but this makes them less useful
- are really only explicitly used to analytically solve problems, but serve as the basis for some algorithms

KKT conditions

$$\text{minimize } f(x)$$
$$\text{subject to } x \in S$$
$$S := \{x \mid g_i(x) \leqslant 0,\ i \in [m]\}$$

The KKT conditions

- are necessary first-order optimality conditions
- are a lot more complicated than unconstrained optimization conditions
- extend to equality constraints (associated dual inequality constraints are always active)
- are also necessary with other (typically weaker) types of constraint qualification
- technically don't require constraint qualification (Fritz-John conditions), but this makes them less useful
- are really only explicitly used to analytically solve problems, but serve as the basis for some algorithms

References I

- Amir Beck, *Introduction to Nonlinear Optimization*, MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics, October 2014.
- ———, *First-Order Methods in Optimization*, MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics, October 2017.
- Stephen Boyd and Lieven Vandenberghe, *Convex Optimization, With Corrections 2008*, 1 edition ed., Cambridge University Press, Cambridge, UK ; New York, March 2004 (English).