POLITECNICO
DI MILANO

Math Tools:

Basics on constrained and convex optimization – Part 1

Maria Prandini

Outline

- Constrained and convex optimization
- Optimality conditions
- Descent iterative methods: gradient algorithms
- Convergence results

Main references:

D. Bertsekas. Nonlinear programming. Athena scientific, 1999

D. Bertsekas. Convex Optimization Theory. Athena Scientific, 2009

Remark: pictures are taken from the reference books

Constrained optimization

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \to R$ is a continuously differentiable function over X

X is a non-empty closed convex set of R^n

Constrained optimization

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \to R$ is a continuously differentiable function over X

X is a non-empty closed convex set of R^n

Convex sets

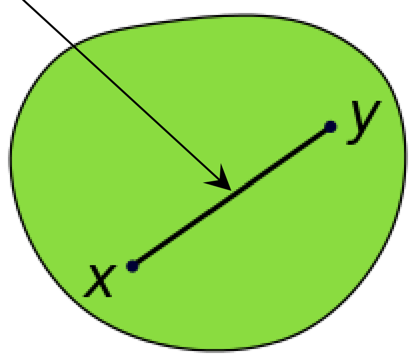
A set $X \subseteq R^n$ is convex if

$$\alpha x + (1-\alpha) y \in X, \forall x, y \in X, \forall \alpha \in [0,1]$$

Convex sets

A set $X \subseteq R^n$ is convex if

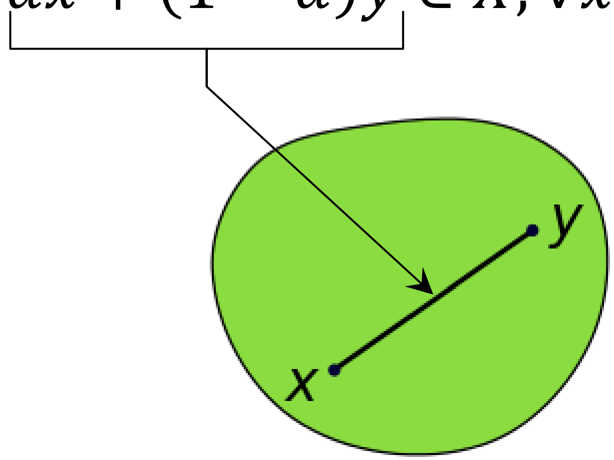
$$\alpha x + (1-\alpha)y \in X, \forall x, y \in X, \forall \alpha \in [0,1]$$



Convex sets

A set $X \subseteq R^n$ is convex if

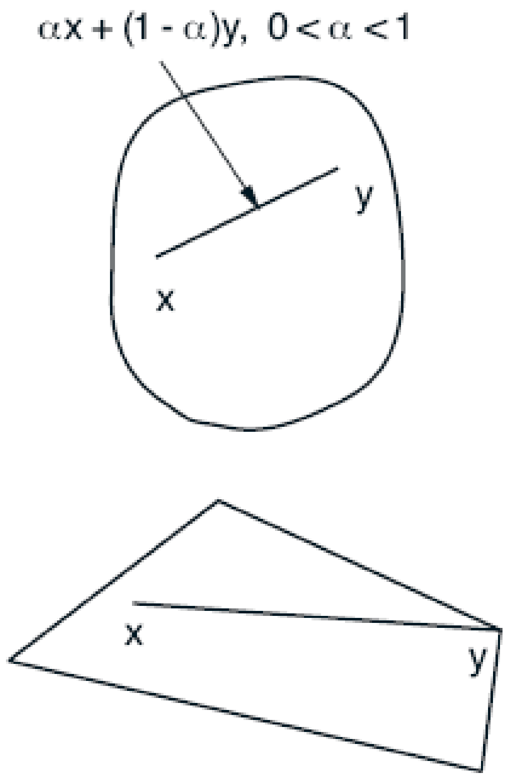
$$\alpha x + (1-\alpha)y \in X, \forall x, y \in X, \forall \alpha \in [0,1]$$



The empty set and R^n are convex

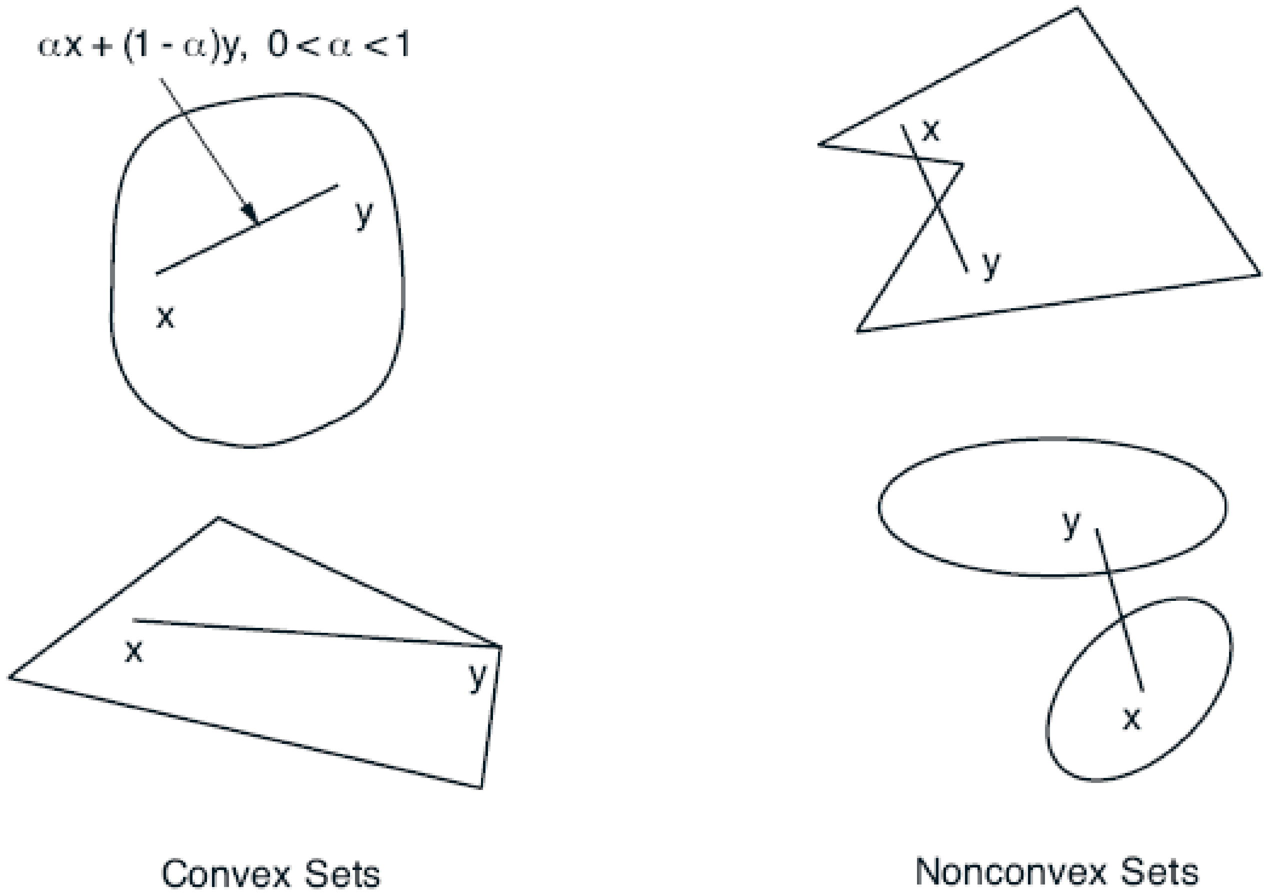
The intersection of any collection of convex sets is convex

Convex sets



Convex Sets

Convex sets



Convex sets

A convex set has nice “shape”:

- connected

Convex sets

A convex set has nice “shape”:

- connected
- at any point $x \in X$, there is a feasible direction

Convex sets

A convex set has nice "shape":

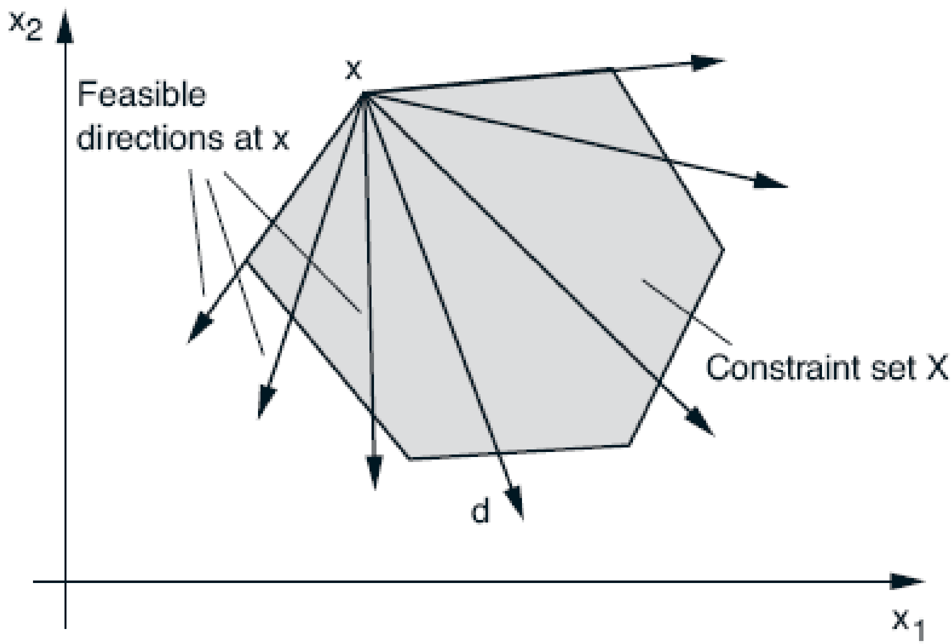
- connected
- at any point $x \in X$, there is a feasible direction

Def. $d \in R^n$ is a feasible direction at $x \in X$ if $x + \alpha d \in X$ for all $\alpha > 0$ that are sufficiently small.

Convex sets

A convex set has nice "shape":

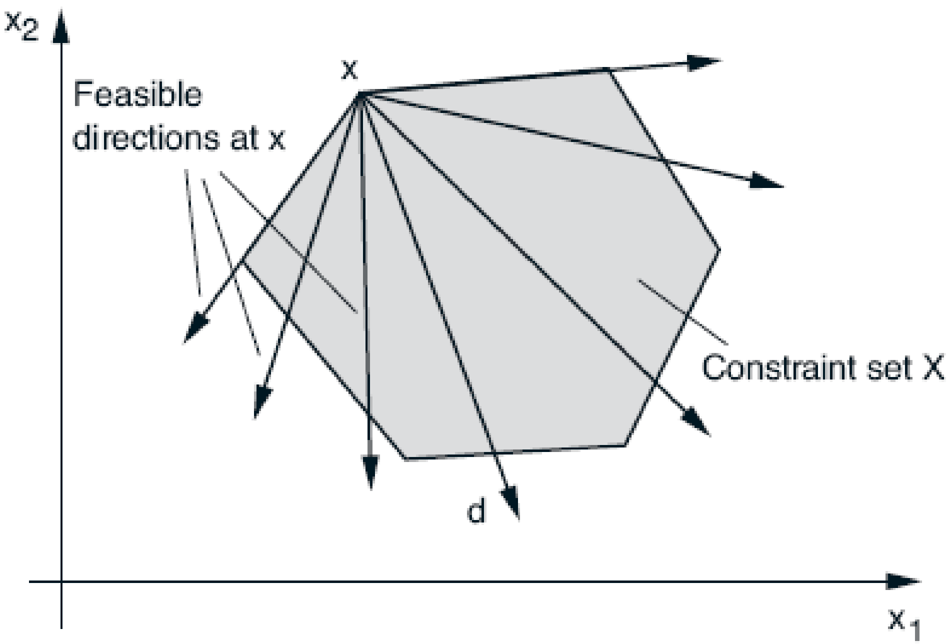
- connected
- at any point $x \in X$, there is a feasible direction



Convex sets

A convex set has nice “shape”:

- connected
- at any point $x \in X$, there is a feasible direction



If X convex, feasible directions are given by $d = y - x$ with $y \in X$

Constrained optimization

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \rightarrow R$ is a continuously differentiable function over X

X is a non-empty closed convex subset of R^n

Constrained optimization

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \rightarrow R$ is a continuously differentiable function over X

X is a non-empty closed convex subset of R^n

$x \in X$ is a *feasible solution* for the optimization problem

if $X = R^n$, then the optimization problem is *unconstrained*

Local and global minima

A feasible $x^* \in X$ is

- a ***local minimum of f over the set X*** if

  $\exists \varepsilon > 0 \; such \; that \; f(x^*) \leq f(x), \forall x \in X \; with \; \|x - x^*\| \leq \varepsilon$

Local and global minima

A feasible $x^* \in X$ is

- a ***local minimum of f over the set X*** if
  $\exists \varepsilon > 0 \; such \; that \; f(x^*) \leq f(x), \forall x \in X \; with \; \|x - x^*\| \leq \varepsilon$
- a global ***minimum of f over the set X*** if
  $f(x^*) \leq f(x), \forall x \in X$

Local and global minima

A feasible $x^* \in X$ is

- a ***local minimum of f over the set X*** if

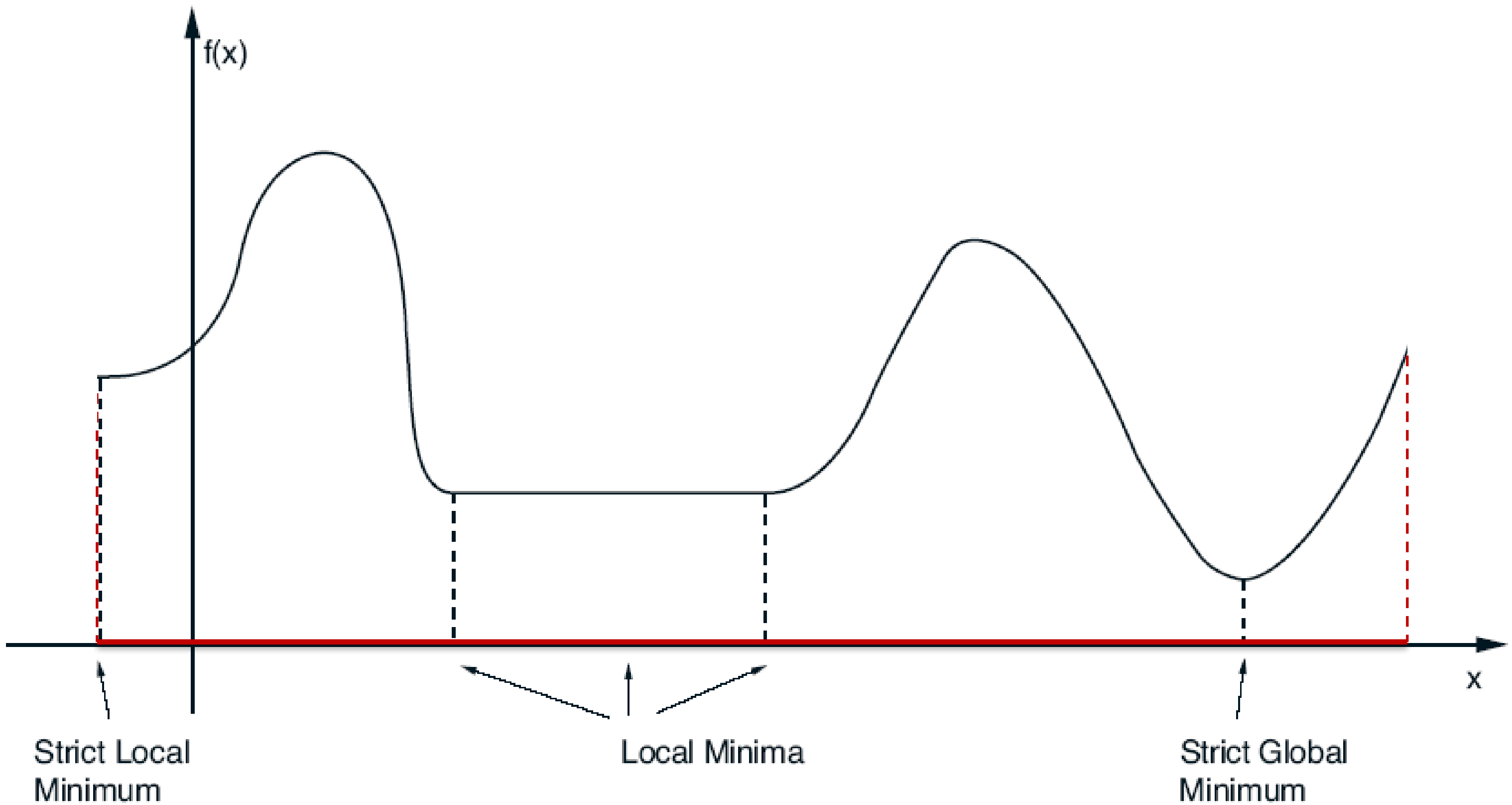
  $\exists \varepsilon > 0$ $such\ that$ $f(x^*) \leq f(x), \forall x \in X$ $with$ $\|x - x^*\| \leq \varepsilon$

- a global ***minimum of f over the set X*** if

  $f(x^*) \leq f(x), \forall x \in X$

A local/global minimum is **strict** if $f(x^*) < f(x)$ for $x \neq x^*$

Local and global minima



X

Existence of a minimum

$f(x) = x$ and $f(x) = e^x$ have no minima over $X = R$

Existence of a minimum

$f(x) = x$ and $f(x) = e^x$ have no minima over $X = R$

How shall we know that at least a (global) minimum of a function f over X does exist?

Existence of a minimum

$f(x) = x$ and $f(x) = e^x$ have no minima over $X = R$

How shall we know that at least a (global) minimum of a function f over X does exist?

Sufficient conditions for the existence of a minimum:

i) f continuous and X compact (closed and bounded)

ii) f continuous, X closed, and f coercive ($\lim_{\|x\| \to \infty} f(x) = +\infty$)

Existence of a minimum

$f(x) = x$ and $f(x) = e^x$ have no minima over $X = R$

How shall we know that at least a (global) minimum of a function f over X does exist?

Sufficient conditions for the existence of a minimum:

i) f continuous and X compact (closed and bounded)

ii) f continuous, X closed, and f coercive ($\lim_{\|x\| \to \infty} f(x) = +\infty$)

What about local versus global minima?

Convex functions and minima

Let $X \subseteq R^n$ be a convex set and $f: X \to R$ a convex function.

Then, a local minimum x^* of f over X is also a global minimum.

Convex functions and minima

Let $X \subseteq R^n$ be a convex set and $f: X \to R$ a convex function.

Then, a local minimum x^* of f over X is also a global minimum.

Convex functions

$f: X \to R$ is a ***convex function*** if X convex and

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \forall x, y \in X, \forall \alpha \in [0,1]$$

Convex functions

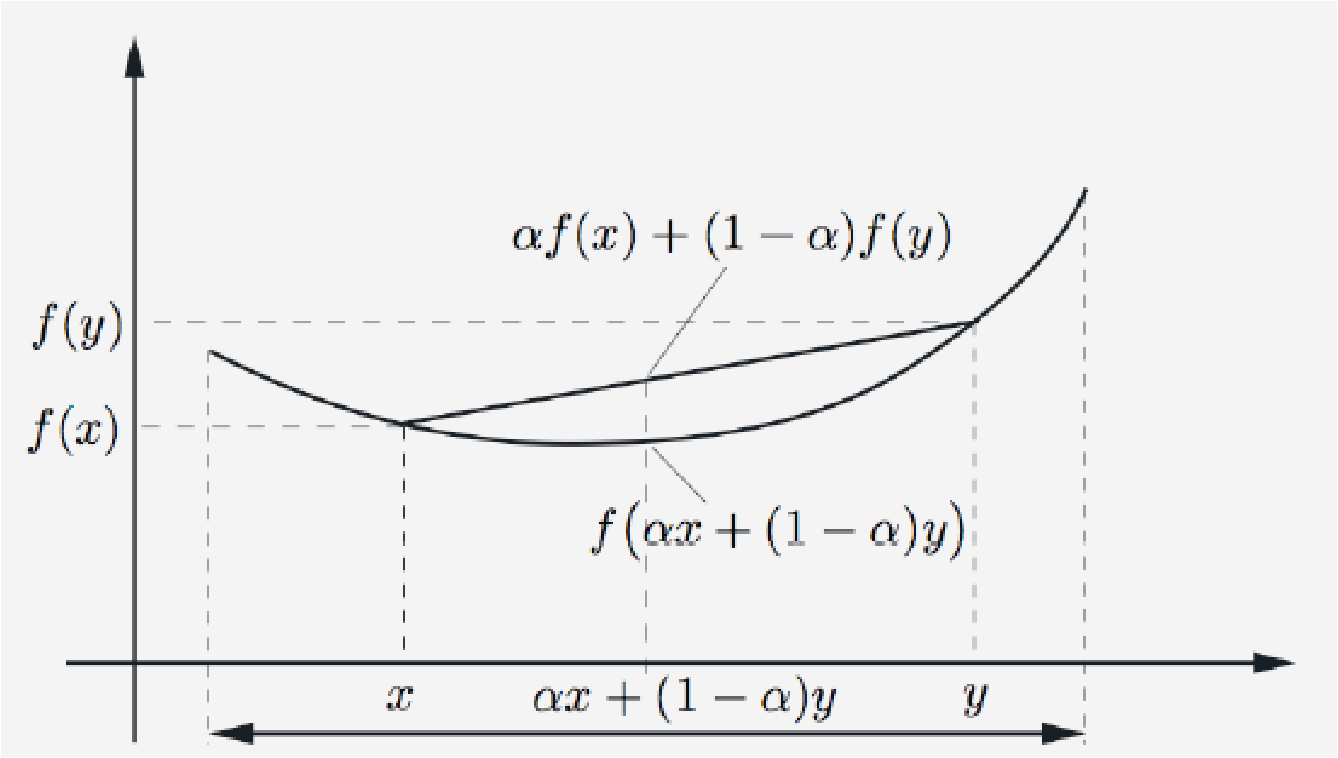
$f: X \to R$ is a ***convex function*** if X convex and

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y), \forall x, y \in X, \forall \alpha \in [0,1]$$

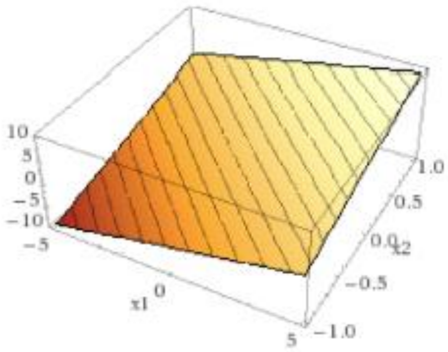
Convex functions

$f: X \to R$ is a ***convex function*** if X convex and

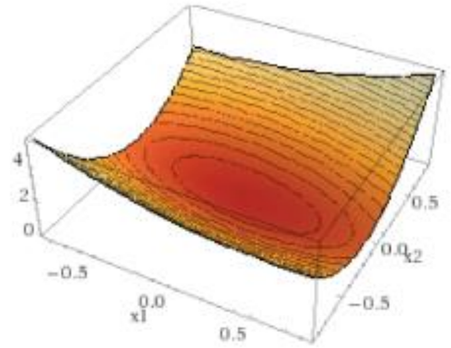
$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y), \forall x, y \in X, \forall \alpha \in [0,1]$$



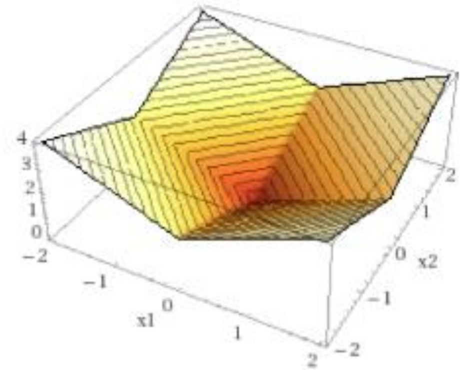
Convex functions



(a) An affine function

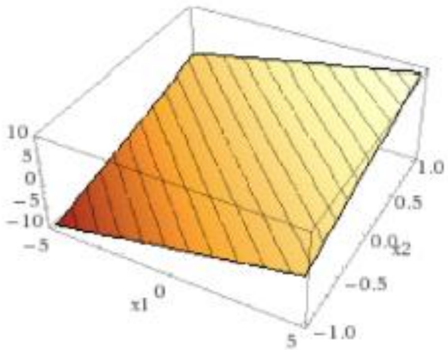


(b) A quadratic function

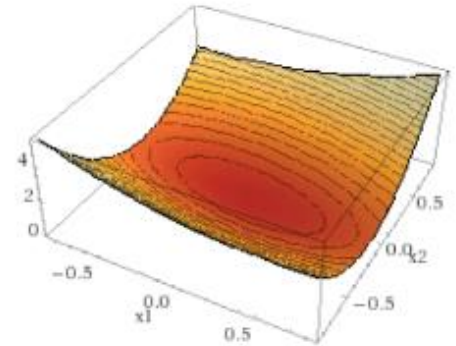


(c) The 1-norm

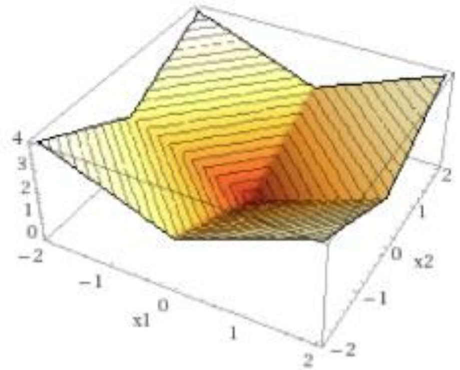
Convex functions



(a) An affine function



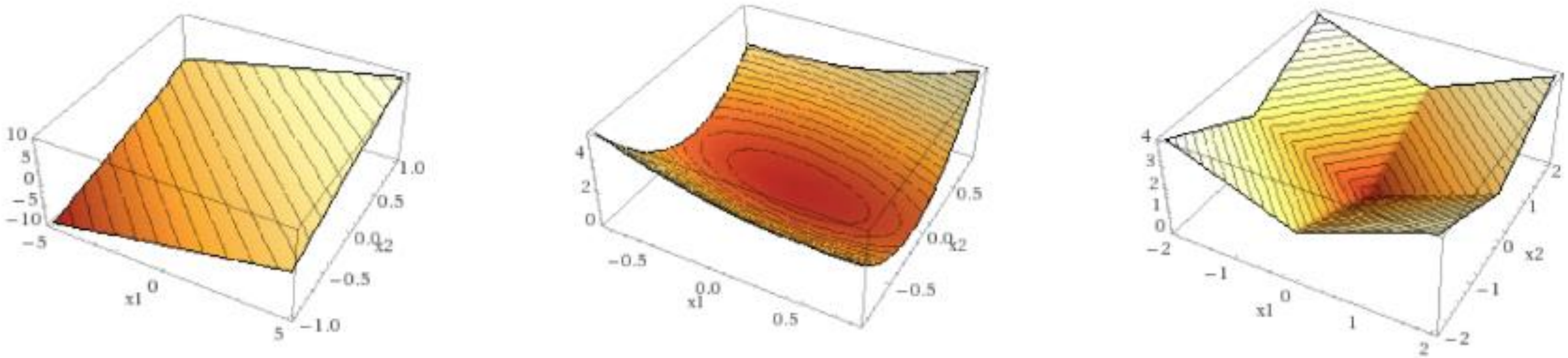
(b) A quadratic function



(c) The 1-norm

All norms are convex functions

Convex functions



(a) An affine function (b) A quadratic function (c) The 1-norm

All norms are convex functions

$$
\begin{aligned}
f(\alpha x + (1-\alpha)y) &\le f(\alpha x) + f\big((1-\alpha)y\big) \; [\textit{triangle inequality}] \\
&= \alpha f(x) + (1-\alpha) f(y) \; [\textit{homogeneity}]
\end{aligned}
$$

Convex functions and minima

Let $X \subseteq R^n$ be a convex set and $f: X \to R$ a convex function.

Then, a local minimum x^* of f over X is also a global minimum.

Convex functions and minima

Let $X \subseteq R^n$ be a convex set and $f: X \rightarrow R$ a convex function.
Then, a local minimum x^* of f over X is also a global minimum.

Proof [by contradiction]:

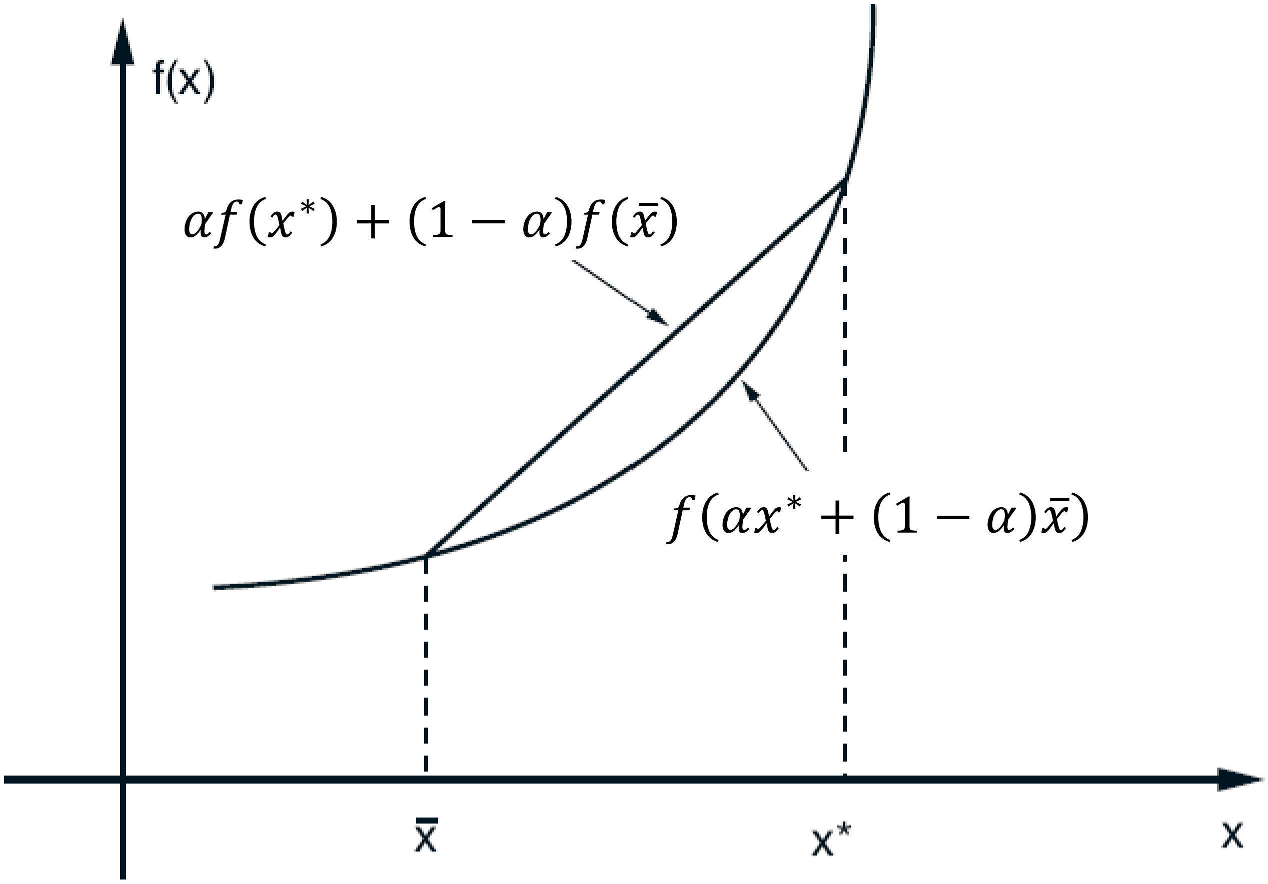
Suppose that there exists $\bar{x} \in X$ such that $f(\bar{x}) < f(x^*)$.

Then,

$$f(\alpha x^* + (1-\alpha)\bar{x}) \leq \alpha f(x^*) + (1-\alpha) f(\bar{x}) < f(x^*), \forall \alpha$$

which contradicts that fact that x^* is a local minimum.

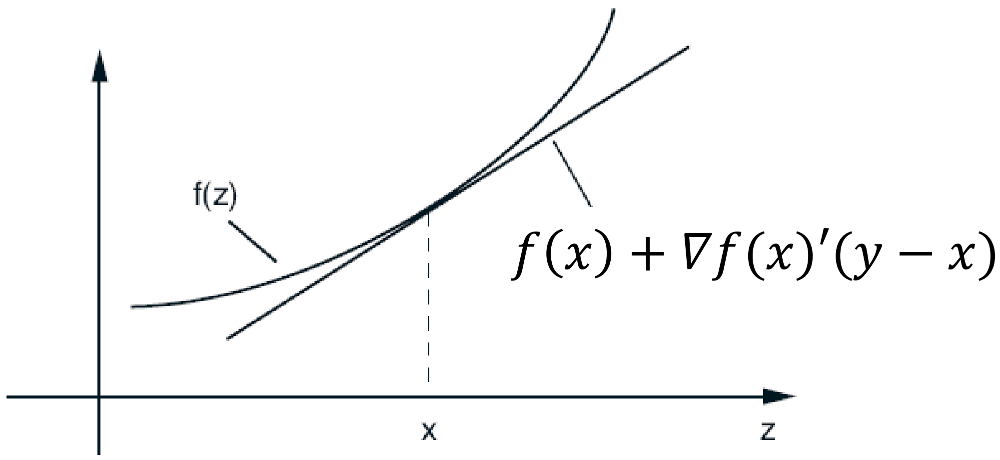
Convex functions and minima



Convex functions: first order characterization

$f: X \to R$ differentiable is a convex function if and only if

$$f(y) \geq f(x) + \nabla f(x)'(y - x), \forall y \in X, \forall x \in X$$



the first order Taylor expansion at any point is a global under-estimator of the function

Convex functions: first order characterization

Proof [only if]

$f(\alpha y + (1-\alpha)x) \geq \alpha f(y) + (1-\alpha) f(x), \forall \alpha \in [0,1], \forall x, y \in X$

By rewriting, we get

$$f(x + \alpha(y-x)) \geq f(x) + \alpha(f(y) - f(x))$$

from which it follows

$$f(y) - f(x) \geq \frac{f(x + \alpha(y-x)) - f(x)}{\alpha(y-x)}(y-x)$$

as $\alpha \to 0^+$, $f(y) - f(x) \geq \nabla f(x)'(y-x)$

Convex functions: first order characterization

Proof [if]

$$z = \alpha y + (1-\alpha)x$$

From

$$f(y) \geq f(z) + \nabla f(z)'(y - z)$$
$$f(x) \geq f(z) + \nabla f(z)'(x - z)$$

we obtain

$$\alpha f(y) + (1-\alpha)f(x) \geq f(z) + \nabla f(z)'(\alpha y + (1-\alpha)x - z) = f(z)$$

which rewrites as

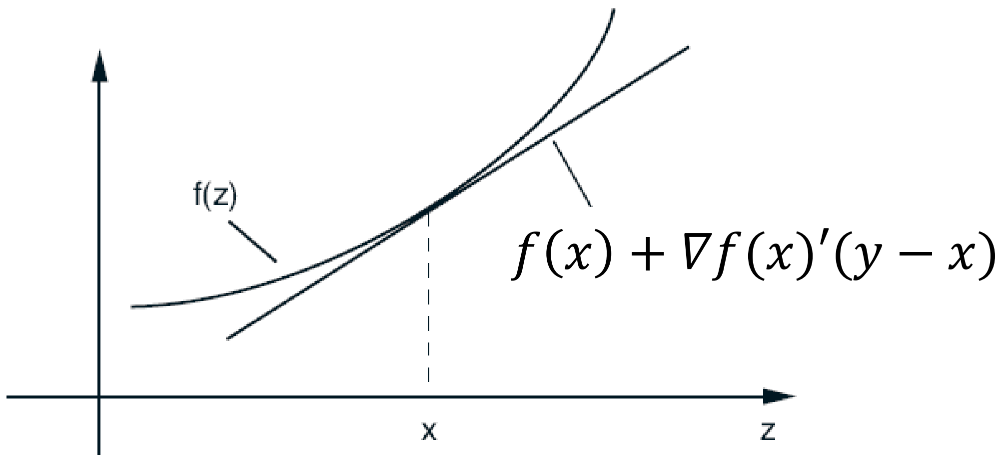
$$\alpha f(y) + (1-\alpha)f(x) \geq f(\alpha y + (1-\alpha)x)$$

i.e., f is convex

Convex functions: first order characterization

$f: X \rightarrow R$ differentiable is a convex function if and only if

$$f(y) \geq f(x) + \nabla f(x)'(y - x), \forall y \in X, \forall x \in X$$



the first order Taylor expansion at any point is a global under estimator of the function

growth of a convex function is at least linear

Convex functions

$f: X \to R$ is a ***convex function*** if

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \forall x, y \in X, \forall \alpha \in [0,1]$$

Convex functions

$f: X \rightarrow R$ is a ***convex function*** if

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y), \forall x, y \in X, \forall \alpha \in [0,1]$$

$f: X \rightarrow R$ is a ***strictly convex function*** if

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y),$$
$$\forall x, y \in X, x \neq y, \forall \alpha \in (0,1)$$

Convex functions

$f: X \to R$ is a ***convex function*** if

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \forall x, y \in X, \forall \alpha \in [0,1]$$

$f: X \to R$ is a ***strictly convex function*** if

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y),$$
$$\forall x, y \in X, x \neq y, \forall \alpha \in (0,1)$$

$$f(y) > f(x) + \nabla f(x)'(y - x), \forall y \in X, \forall x \in X, x \neq y$$

growth of a strictly convex function is more than linear

Convex functions and minima

Let $X \subseteq R^n$ be a convex set and $f: X \to R$ a convex function.

Then, a local minimum x^* of f over X is also a global minimum.

If f is also **strictly convex**,

then there exists **at most a global minimum** of f over X.

Convex functions and minima

Let $X \subseteq R^n$ be a convex set and $f: X \rightarrow R$ a convex function.

Then, a local minimum x^* of f over X is also a global minimum.

If f is also **strictly convex**,

then there exists **at most a global minimum** of f over X.

Proof [by contradiction]:

Suppose that x^* and $y^*$are both global minima.

Then, by strict convexity:

$$f(0.5x^* + 0.5y^*) < 0.5f(x^*) + 0.5f(y^*) = f(x^*)$$

which contradicts the fact that x^* is a global minimum

Convex functions

$f: X \to R$ is a ***convex function*** if

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \forall x, y \in X, \forall \alpha \in [0,1]$$

$f: X \to R$ is a ***strictly convex function*** if

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y),$$
$$\forall x, y \in X, x \neq y, \forall \alpha \in (0,1)$$

$f: X \to R$ is a ***strongly convex function*** if there exists $\mu > 0$ such that

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \frac{\alpha(1-\alpha)\mu}{2}\|x-y\|^2, \ \alpha \in [0,1]$$

Convex functions

$f: X \to R$ is a ***convex function*** if

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \forall x, y \in X, \forall \alpha \in [0,1]$$

$f: X \to R$ is a ***strictly convex function*** if

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y),$$
$$\forall x, y \in X, x \neq y, \forall \alpha \in (0,1)$$

$f: X \to R$ is a ***strongly convex function*** if there exists $\mu > 0$ such that

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \frac{\alpha(1-\alpha)\mu}{2}\|x-y\|^2, \ \alpha \in [0,1]$$

strongly convex $\Rightarrow$ strictly convex $\Rightarrow$ convex

Convex functions

$f: X \to R$ is a ***strongly convex function*** if there exists $\mu > 0$ such that

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) - \frac{\alpha(1-\alpha)\mu}{2}\|x-y\|^2,\ \ \alpha \in [0,1]$$

Equivalently,

$$g(x) = f(x) - \frac{\mu}{2}\|x\|^2 \text{ is convex}$$

Convex functions

$f: X \rightarrow R$ is a ***strongly convex function*** if there exists $\mu > 0$ such that

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \frac{\alpha(1-\alpha)\mu}{2}\|x-y\|^2, \ \alpha \in [0,1]$$

Equivalently,

$$g(x) = f(x) - \frac{\mu}{2}\|x\|^2 \text{ is convex}$$

Proof [exercise]:

follows from the definition of convex function for $g(x)$

Convex functions

$f: X \to R$ is a ***strongly convex function*** if there exists $\mu > 0$ such that

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) - \frac{\alpha(1-\alpha)\mu}{2}\|x-y\|^2, \ \alpha \in [0,1]$$

Equivalently,

$$g(x) = f(x) - \frac{\mu}{2}\|x\|^2 \text{ is convex}$$

and if f is differentiable

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2}\|y-x\|^2, \ \forall x, y \in X$$

Convex functions

$f: X \to R$ is a ***strongly convex function*** if there exists $\mu > 0$ such that

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \frac{\alpha(1-\alpha)\mu}{2}\|x-y\|^2, \ \alpha \in [0,1]$$

Equivalently,

$$g(x) = f(x) - \frac{\mu}{2}\|x\|^2 \text{ is convex}$$

and if f is differentiable

$$f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{\mu}{2}\|y-x\|^2, \ \forall x, y \in X$$

Proof [exercise]:

follows from the first-order condition for convexity of $g(x)$, $i.e.$

$$g(y) \geq g(x) + \nabla g(x)^T(y-x), \ \forall x, y$$

Strong convexity

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2} \|y - x\|^2, \ \forall x, y \in X$$

In practice, strong convexity means that there exists a quadratic lower bound on the growth of the function.

Strong convexity

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2} \|y - x\|^2, \ \forall x, y \in X$$

In practice, strong convexity means that there exists a quadratic lower bound on the growth of the function

for a convex function the growth is at least linear

for a strictly convex function the growth is more than linear

for a strongly convex function the growth is at least quadratic

Convex functions

If $f: X \rightarrow R$ with $X \subseteq R$ is twice continuously differentiable, then we can characterize convexity, strict convexity and strong convexity as follows:

i) f convex if and only if $\frac{d^2 f}{dx^2}(x) \geq 0, \forall x \epsilon X$

ii) f strictly convex if $\frac{d^2 f}{dx^2}(x) > 0, \forall x \epsilon X$

iii) f is μ-strongly convex if and only if $\frac{d^2 f}{dx^2}(x) \geq \mu, \forall x \epsilon X$

Convex functions

If $f: X \rightarrow R$ with $X \subseteq R$ is twice continuously differentiable, then we can characterize convexity, strict convexity and strong convexity as follows:

i) f convex if and only if $\frac{d^2 f}{dx^2}(x) \geq 0, \forall x \epsilon X$

ii) f strictly convex if $\frac{d^2 f}{dx^2}(x) > 0, \forall x \epsilon X$

iii) f is μ-strongly convex if and only if $\frac{d^2 f}{dx^2}(x) \geq \mu, \forall x \epsilon X$

Remark:

ii) is a sufficient but not necessary condition

Example: $f(x) = x^4$ is strictly convex but $\frac{d^2 f}{dx^2}(x) = 12x^2$

Outline

- Constrained and convex optimization ☺
- Optimality conditions
- Descent iterative methods: gradient algorithms
- Convergence results
- Non differentiable setting

Optimality conditions

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \rightarrow R$ is a continuously differentiable function over X

X is a non-empty closed convex subset of R^n

Characterization of local minima through necessary and/or sufficient conditions in:

- the unconstrained case ($X = R^n$)
- the constrained case ($X \subset R^n$)

Optimality conditions – unconstrained opt

Necessary conditions for x^* to be a local minimum of f over R^n

- ***First order condition:*** zero slope at x^*

$$\nabla f(x^*) = 0$$

  where ∇f is the gradient of f (i.e., $\nabla f_i = \frac{\partial f}{\partial x_i}$)

Optimality conditions – unconstrained opt

Necessary conditions for x^* to be a local minimum of f over R^n

- ***First order condition:*** zero slope at x^*

$$\nabla f(x^*) = 0$$

  where ∇f is the gradient of f (i.e., $\nabla f_i = \frac{\partial f}{\partial x_i}$)

Def. x* is a *stationary point* if it satisfies the first order condition.

Optimality conditions – unconstrained opt

Necessary conditions for x^* to be a local minimum of f over R^n

- ***First order condition:*** zero slope at x^*

$$\nabla f(x^*) = 0$$

  where ∇f is the gradient of f (i.e., $\nabla f_i = \frac{\partial f}{\partial x_i}$)

Def. x* is a *stationary point* if it satisfies the first order condition.

- ***Second order condition:*** nonnegative curvature at x^*

$$\nabla^2 f(x^*) \text{ positive semidefinite}$$

  where $\nabla^2 f$ is the Hessian matrix of f (i.e., $\nabla^2 f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$)

Optimality conditions – unconstrained opt

First order cost variation

$$f(x^* + \Delta x) \cong f(x^*) + \nabla f(x^*)' \Delta x$$

Second order cost variation

$$f(x^* + \Delta x) \cong f(x^*) + \nabla f(x^*)' \Delta x + \frac{1}{2} \Delta x' \nabla^2 f(x^*) \Delta x$$

Optimality conditions – unconstrained opt

First order cost variation non-negative

$$\nabla f(x^*)'\Delta x = \sum_{i=1}^{n} \frac{\partial f(x^*)}{\partial x_i} \Delta x_i \geq 0, \forall \Delta x$$

→ first order condition follows

Optimality conditions – unconstrained opt

First order cost variation non-negative

$$\nabla f(x^*)'\Delta x = \sum_{i=1}^{n} \frac{\partial f(x^*)}{\partial x_i} \Delta x_i \geq 0, \forall \Delta x$$

→ first order condition follows

Second order cost variation non-negative

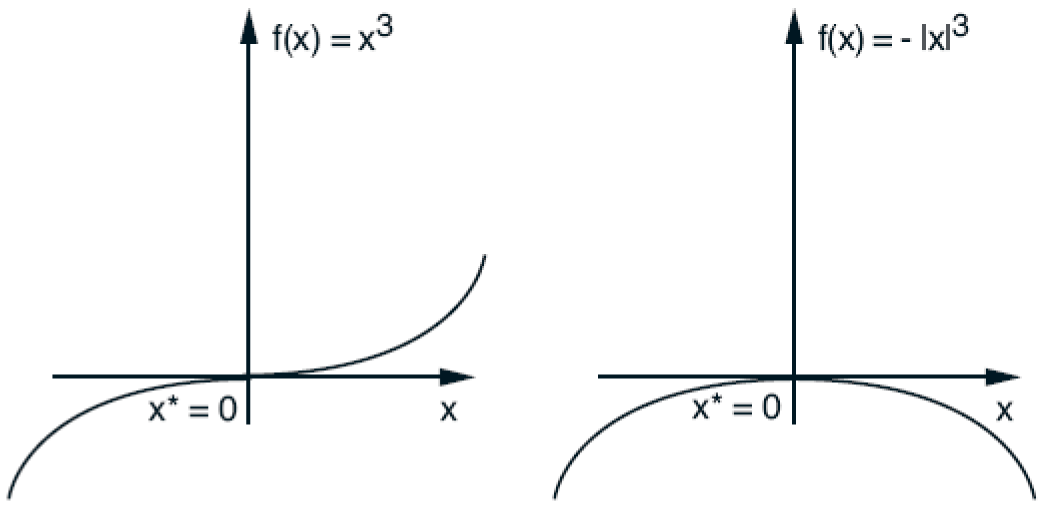
$$\nabla f(x^*)'\Delta x + \frac{1}{2} \Delta x' \nabla^2 f(x^*) \Delta x \geq 0, \forall \Delta x$$

→ second order condition follows

Optimality conditions – unconstrained opt

These optimality conditions are necessary but not sufficient

→ there may exists points that satisfy both conditions but are not local minima



Optimality conditions – unconstrained opt

Sufficient conditions for x^* to be a local minimum of f over R^n

- ***First order condition:*** zero slope at x^*

$$\nabla f(x^*) = 0$$

where ∇f is the gradient of f (i.e., $\nabla f_i = \frac{\partial f}{\partial x_i}$)

- ***Second order condition:*** positive curvature at x^*

$$\nabla^2 f(x^*) \text{ positive definite}$$

where $\nabla^2 f$ is the Hessian matrix of f (i.e., $\nabla^2 f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$)

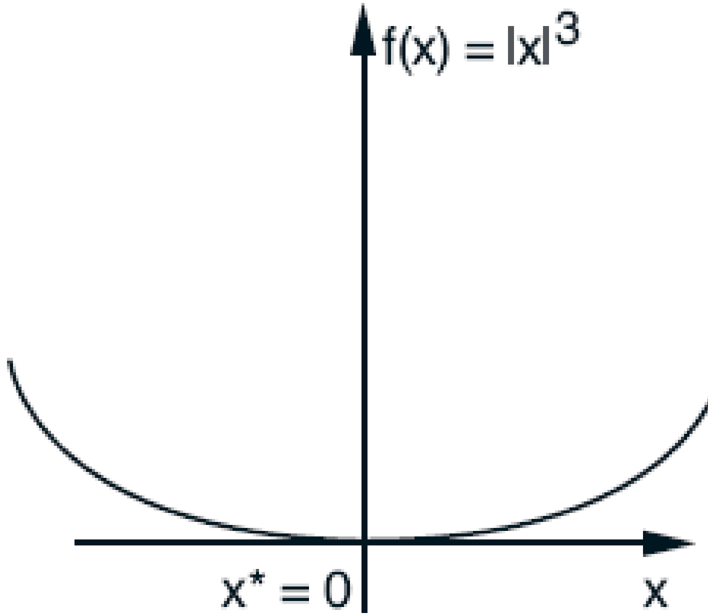
Optimality conditions – unconstrained opt

Second order cost variation positive

$$\nabla f(x^*)'\Delta x + \frac{1}{2}\Delta x'\nabla^2 f(x^*)\Delta x > 0, \forall \Delta x \neq 0$$

Optimality conditions – unconstrained opt

Sufficient but not necessary…



Optimality conditions – unconstrained convex

Let $X \subseteq R^n$ be a convex set.

If $f: X \rightarrow R$ is convex then the first order condition

$$\nabla f(x^*) = 0$$

is necessary and sufficient for x^* to be a global minimum

Optimality conditions – unconstrained convex

Let $X \subseteq R^n$ be a convex set.

If $f: X \to R$ is convex then the first order condition

$$\nabla f(x^*) = 0$$

is necessary and sufficient for x^* to be a global minimum

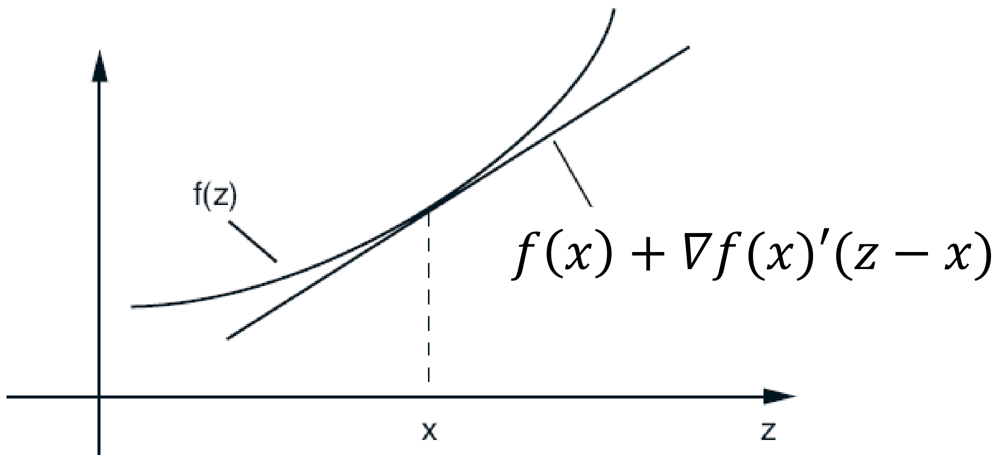
⇩

all stationary points of a convex function are global minima

Optimality conditions – unconstrained convex

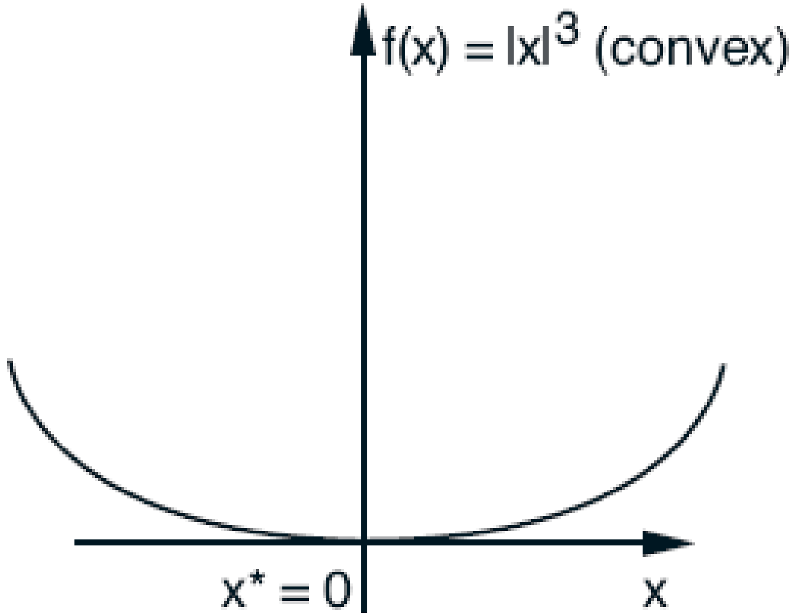
$f: X \rightarrow R$ is a convex function if and only if

$$f(z) \geq f(x) + \nabla f(x)'(z - x), \forall z \in X, \forall x \in X$$



the linear approximation of f at a point x based on its gradient underestimates f → the first order Taylor expansion at any point is a global under-estimator of the function

Optimality conditions – unconstrained convex



Optimality conditions

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \rightarrow R$ is a continuously differentiable function over X

X is a non-empty closed convex subset of R^n

Characterization of local minima through necessary and/or sufficient conditions in:

- the unconstrained case ($X = R^n$)
- the constrained case ($X \subset R^n$)

Optimality conditions – constrained opt

Necessary condition for x^* to be a local minimum of f over the convex set X

$$\nabla f(x^*)'(x - x^*) \geq 0, \forall x \in X$$

Optimality conditions – constrained opt

Necessary condition for x^* to be a local minimum of f over the convex set X

$$\nabla f(x^*)'(x - x^*) \geq 0, \forall x \in X$$

Def. x* is a **stationary point** if it satisfies this necessary condition.

Optimality conditions – constrained opt

Proof:

First order cost variation

$$f(x^* + \Delta x) \cong f(x^*) + \nabla f(x^*)'\Delta x$$

Since x^* is a local minimum

$$\nabla f(x^*)'\Delta x \geq 0$$

for any feasible (small) variation Δx, i.e., any Δx such that

$$x^* + \Delta x \in X$$

Since X is convex, feasible variations are of the form $x - x^*$ with $x \in X$, which leads to the optimality condition

$$\nabla f(x^*)'(x - x^*) \geq 0, \forall x \in X$$

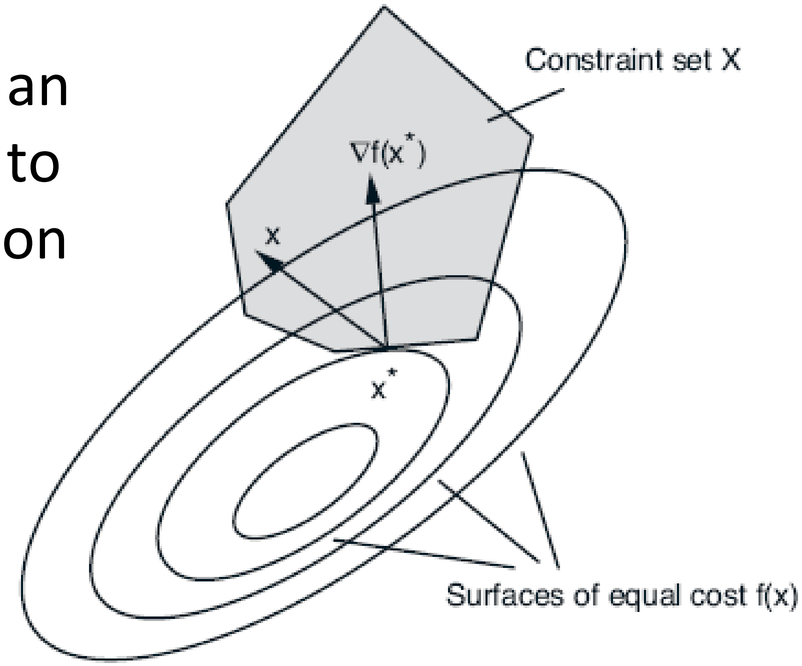
Optimality conditions – constrained opt

Necessary condition for x^* to be a local minimum of f over the convex set X

$$\nabla f(x^*)'(x - x^*) \geq 0, \forall x \in X$$

Geometric interpretation:

the gradient $\nabla f(x^*)$ makes an angle smaller than or equal to 90° with any feasible variation $x - x^*$

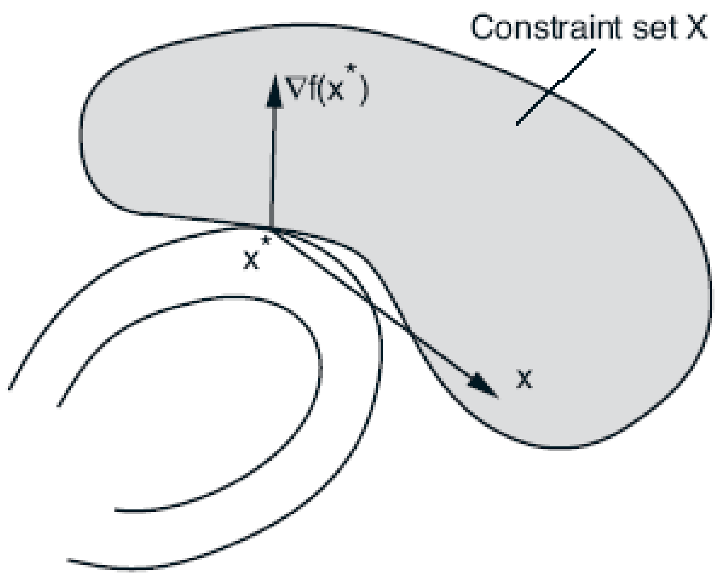


Optimality conditions – constrained opt

Necessary condition for x^* to be a local minimum of f over the convex set X

$$\nabla f(x^*)'(x - x^*) \geq 0, \forall x \in X$$

Example of failure if X non convex

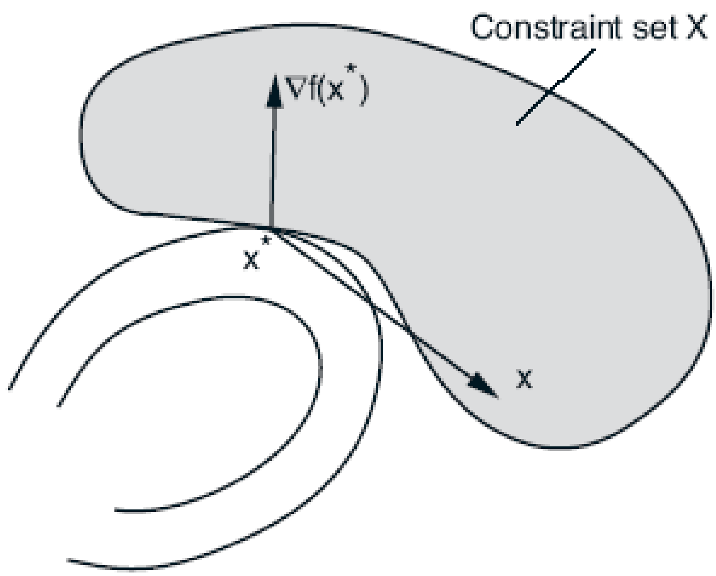


Optimality conditions – constrained opt

Necessary condition for x^* to be a local minimum of f over the convex set X

$$\nabla f(x^*)'(x - x^*) \geq 0, \forall x \in X$$

Example of failure if X non convex



$x - x^*$ *with* $x \in X$ *is not a feasible direction in this case*

Optimality conditions – constrained opt

Necessary condition for x^* to be a local minimum of f over the convex set X

$$\nabla f(x^*)'(x - x^*) \geq 0, \forall x \in X$$

If f is convex over X, then, this condition is also sufficient for x^* to be a (global) minimum of f

→ in the convex case stationary points are all global minima

Optimality conditions – constrained opt

Necessary condition for x^* to be a local minimum of f over the convex set X

$$\nabla f(x^*)'(x - x^*) \geq 0, \forall x \in X$$

If f is convex over X then this condition is also sufficient for x^* to be a (global) minimum of f

Proof [convex f]:

$$\forall x \in X, \qquad f(x) \geq f(x^*) + \nabla f(x^*)'(x - x^*) \text{ [}f\text{ convex]}$$
$$\geq f(x^*) \text{ [necessary condition above]}$$

→ x^* is a global minimum

Example of constrained convex optimization

Projection over a convex set

Let $z \in R^n$ and $X \subseteq R^n$ be a non-empty closed convex set.

Then, problem

$$\text{minimize } f(x) = \|z - x\|^2$$
$$\text{subject to } x \in X$$

has a unique solution $\mathrm{P}_X[z]$, which is the projection of z on the convex set X according to the Euclidean norm

Projection over a convex set

Proof [existence and uniqueness]:

Existence

it satisfies the sufficient condition for the existence of a minimum:

f continuous, X closed, and f coercive ($\lim_{\|x\| \to \infty} f(x) = +\infty$)

Uniqueness

$f(x) = \|z - x\|^2$ is strictly convex

Projection over a convex set

Projection Theorem

Let $z \in R^n$ and $X \subseteq R^n$ be a non-empty closed convex set.

Then, we have that:

$x^* \in X$ is the projection of z on X, i.e., $x^* = \mathrm{P}_X[z]$, if and only if

$$(z - x^*)'(x - x^*) \leq 0, \forall x \in X$$

Projection over a convex set

Projection Theorem

Let $z \in R^n$ and $X \subseteq R^n$ be a non-empty closed convex set.

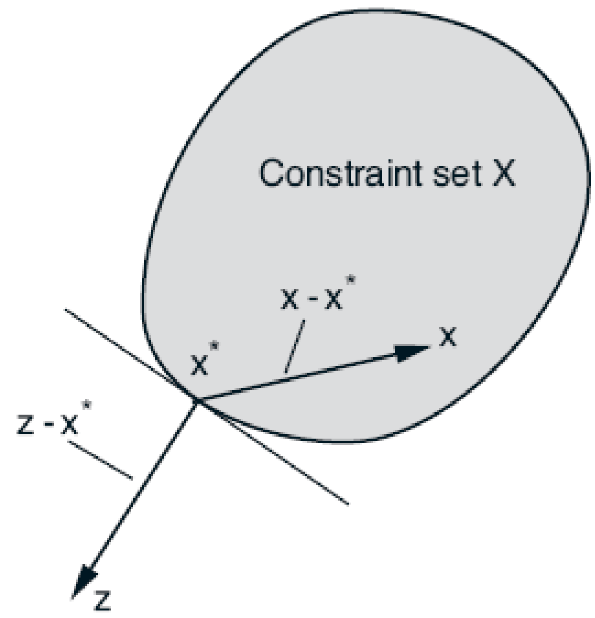
Then, we have that:

$x^* \in X$ is the projection of z on X, i.e., $x^* = \mathrm{P}_X[z]$, if and only if

$$(z - x^*)'(x - x^*) \leq 0, \forall x \in X$$

Geometric interpretation:

x^* is the projection of z on X
if and only if the angle between
$\mathrm{z} - x^*$ and every feasible variation
$x - x^*$ is larger than or equal to 90°



Projection over a convex set

Projection Theorem

Let $z \in R^n$ and $X \subseteq R^n$ be a non-empty closed convex set.

Then, we have that:

$x^* \in X$ is the projection of z on X, i.e., $x^* = \mathrm{P}_X[z]$, if and only if

$$(z - x^*)'(x - x^*) \leq 0, \forall x \in X$$

Proof:

$$\text{minimize } f(x) = \|z - x\|^2$$
$$\text{subject to } x \in X$$

is a convex optimization problem. This implies that

$x^* = \mathrm{P}_X[z]$ if and only if

$$\nabla f(x^*)'(x - x^*) \geq 0, \forall x \in X$$

where $\nabla f(x^*) = -2(\mathrm{z} - x^*)$

Projection over a convex set

Projection Theorem

Let $X \subseteq R^n$ be a non-empty closed convex set.

The projection map $T: R^n \to X$ defined by

$$T(z) = \mathrm{P}_X[z]$$

is continuous and nonexpansive, i.e.,

$$\|T(z) - T(y)\| \leq \|z - y\|, \forall z, y \in R^n$$

Outline

- Constrained and convex optimization ☺
- Optimality conditions ☺
- Descent iterative methods: gradient algorithms
- Convergence results

How to determine a minimum?

Direct use of the optimality conditions to obtain a stationary (possible a minimum) point is not a viable approach except for special cases.

Optimality conditions are useful in the design and analysis of **iterative algorithms for determining a minimum**.

Termination conditions are typically based on checking if the optimality conditions are satisfied for the current candidate solution.

Iterative descent methods

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \rightarrow R$ is a continuously differentiable function over X

X is a non-empty closed convex subset of R^n

Idea: iteratively update a tentative solution by *moving along a descent direction* so as to converge to a minimum

- Gradient methods

Gradient methods

At each k, the (feasible) tentative solution is updated as follows

$$x_{k+1} = x_k + \alpha_k d_k, k = 0,1, \dots$$

where α_k is a positive stepsize and d_k must be a feasible direction satisfying the descent condition

$$\nabla f(x_k)' d_k < 0$$

In this way, the first order cost variation

$$f(x_{k+1}) - f(x_k) \cong \nabla f(x_k)' \alpha_k d_k$$

is negative and, for sufficiently small α_k, x_{k+1} is feasible and the cost f decreases

Gradient methods

At each k, the (feasible) tentative solution is updated as follows

$$x_{k+1} = x_k + \alpha_k d_k, k = 0,1, \dots$$

where α_k is a positive stepsize and d_k must be a feasible direction satisfying the descent condition

$$\nabla f(x_k)' d_k < 0$$

Stopping criterion:
optimality conditions satisfied at the current iterate, i.e.,

$\nabla f(x_{k+1}) = 0$ for the unconstrained case

$\nabla f(x_{k+1})'(x - x_{k+1}) \geq 0, \forall x \in X$ for the constrained case

Gradient method: iterative scheme

- Starts with a feasible $x_0 \in X$
- Generates a sequence $\{x_k\}$ according to

$$x_{k+1} = x_k + \alpha_k d_k, k = 0,1, \ldots$$

  where if x_k is not stationary, d_k is a feasible direction at x_k which is also a *descent direction*, i.e.,

$$\nabla f(x_k)' d_k < 0$$

  and the stepsize α_k is chosen to be positive and such that

$$x_k + \alpha_k d_k \in X$$

- If x_{k+1} is stationary, i.e., it satisfies the optimality conditions then, the method stops.

Iterative descent methods

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \to R$ is a continuously differentiable function over X

X is a non-empty closed convex subset of R^n

Gradient methods

- the unconstrained case ($X = R^n$)
- the constrained case ($X \subset R^n$)

Gradient methods – unconstrained opt

Many gradient methods take the form:

$$x_{k+1} = x_k + \alpha_k d_k, k = 0,1, \ldots$$

with

$$d_k = -D_k \nabla f(x_k)$$

where D_k is a positive definite symmetric matrix.

Gradient methods – unconstrained opt

Many gradient methods take the form:

$$x_{k+1} = x_k + \alpha_k d_k, k = 0,1, \dots$$

with

$$d_k = -D_k \nabla f(x_k)$$

where D_k is a positive definite symmetric matrix.

Indeed, if $\nabla f(x_k) \neq 0$,

$$\nabla f(x_k)' d_k = -\nabla f(x_k)' D_k \nabla f(x_k) < 0$$

Gradient methods – unconstrained opt

Many gradient methods take the form:

$$x_{k+1} = x_k + \alpha_k d_k, k = 0,1, \dots$$

with

$$d_k = -D_k \nabla f(x_k)$$

where D_k is a positive definite symmetric matrix.

Steepest descent: $d_k = -\nabla f(x_k)$

Gradient methods – unconstrained opt

Many gradient methods take the form:

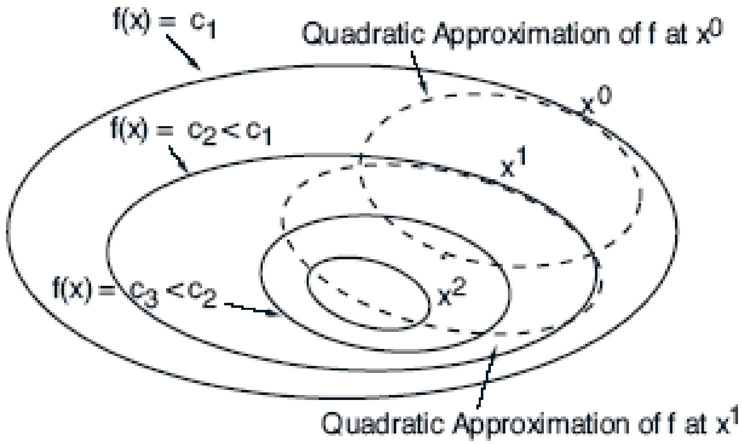
$$x_{k+1} = x_k + \alpha_k d_k, k = 0,1,\ldots$$

with

$$d_k = -D_k \nabla f(x_k)$$

where D_k is a positive definite symmetric matrix.

Newton's method: $d_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$



$\alpha_k = 1, k = 1,2,\ldots$

if f convex, $\nabla^2 f(x_k)$ positive semidefinite

Choice of the stepsize

- Minimization rule

$$\alpha_k = \operatorname*{argmin}_{\alpha \geq 0} f(x_k + \alpha d_k)$$

Choice of the stepsize

- Minimization rule

$$\alpha_k = \operatorname*{argmin}_{\alpha \geq 0} f(x_k + \alpha d_k)$$

- Constant stepsize

$$\alpha_k = c, k = 0,1, \ldots$$

- Diminishing stepsize

$$\lim_{k \to \infty} \alpha_k = 0$$

with $\sum_{k=0}^{\infty} \alpha_k = \infty$ and $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$

Iterative descent methods

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \to R$ is a continuously differentiable function over X

X is a non-empty closed convex subset of R^n

Gradient methods

- the unconstrained case ($X = R^n$)
- the constrained case ($X \subset R^n$)

Gradient methods

At each k, the (feasible) tentative solution is updated as follows

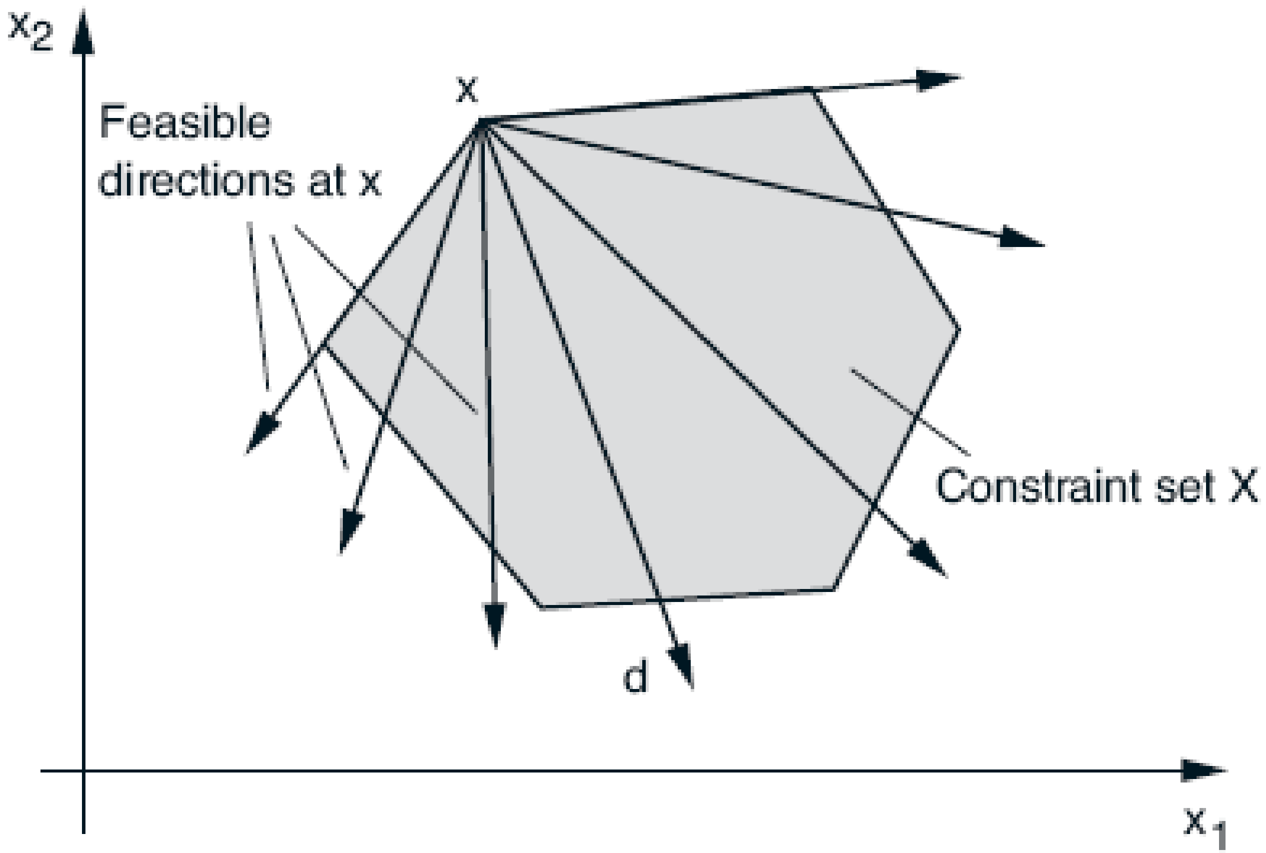
$$x_{k+1} = x_k + \alpha_k d_k, k = 0,1, \dots$$

where α_k is a positive stepsize and d_k must be a feasible direction satisfying the descent condition

$$\nabla f(x_k)' d_k < 0$$

Gradient methods – constrained opt

the **descent directions have to be feasible** so as to maintain feasibility of the iterates



Gradient methods – constrained opt

How to easily obtain a feasible direction d_k at x_k?

Gradient methods – constrained opt

How to easily obtain a feasible direction d_k at x_k?

Since X is convex, all feasible directions can be expressed as

$$d_k = \bar{x}_k - x_k$$

with $\bar{x}_k \in X$

How to easily obtain a feasible direction d_k at x_k?

Since X is convex, all feasible directions can be expressed as

$$d_k = \bar{x}_k - x_k$$

with $\bar{x}_k \in X$

We then get

$$x_{k+1} = x_k + \alpha_k d_k = x_k + \alpha_k(\bar{x}_k - x_k)$$

which belongs to X for any $\alpha_k \in (0,1]$.

Gradient methods – constrained opt

How to easily obtain a feasible direction d_k at x_k?

Since X is convex, all feasible directions can be expressed as

$$d_k = \bar{x}_k - x_k$$

with $\bar{x}_k \in X$

We then get

$$x_{k+1} = x_k + \alpha_k d_k = x_k + \alpha_k(\bar{x}_k - x_k)$$

which belongs to X for any $\alpha_k \in (0,1]$.

→ need to choose $\bar{x}_k \in X$ such that

$$\nabla f(x_k)'(\bar{x}_k - x_k) < 0$$

[descent condition]

Gradient methods – constrained opt

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k)$$

Gradient methods – constrained opt

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k)$$

Conditional gradient method

$$\bar{x}_k = \operatorname*{argmin}_{x \in X} \nabla f(x_k)'(x - x_k)$$

Gradient methods – constrained opt

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k)$$

Conditional gradient method

$$\bar{x}_k = \underset{x \in X}{\operatorname{argmin}} \nabla f(x_k)'(x - x_k)$$

Note that $\bar{x}_k$ satisfies the descent condition

$$\nabla f(x_k)'(\bar{x}_k - x_k) < 0$$

unless x_k is a stationary point

Gradient methods – constrained opt

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k)$$

Gradient projection method

$$\bar{x}_k = \mathrm{P}_X[x_k - c_k \nabla f(x_k)]$$

Gradient methods – constrained opt

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k)$$

Gradient projection method

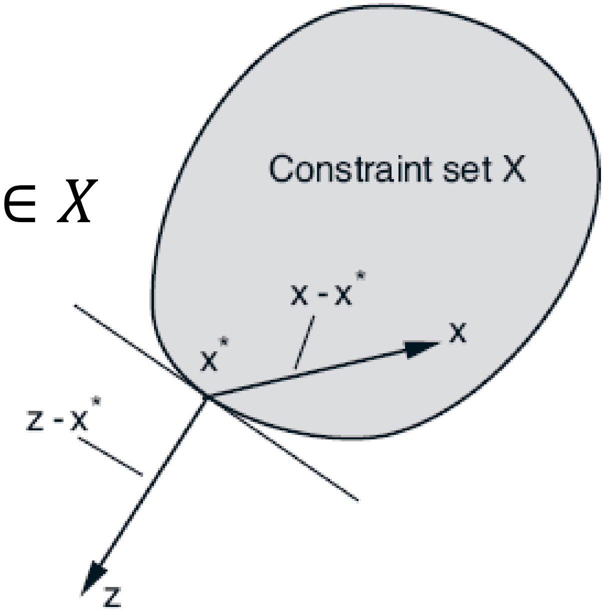
$$\bar{x}_k = \mathrm{P}_X[x_k - c_k \nabla f(x_k)]$$

Note that $\bar{x}_k$ satisfies the descent condition

$$\nabla f(x_k)'(\bar{x}_k - x_k) < 0$$

since by the projection theorem

$$(x_k - c_k \nabla f(x_k) - \bar{x}_k)'(x - \bar{x}_k) \leq 0, \forall x \in X$$



Gradient methods – constrained opt

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k)$$

Gradient projection method

$$\bar{x}_k = \mathrm{P}_X[x_k - c_k \nabla f(x_k)]$$

Note that $\bar{x}_k$ satisfies the descent condition

$$\nabla f(x_k)'(\bar{x}_k - x_k) < 0$$

since by the projection theorem

$$(x_k - c_k \nabla f(x_k) - \bar{x}_k)'(x - \bar{x}_k) \le 0, \forall x \in X$$

and if we set $x = x_k$, we obtain

$$\nabla f(x_k)'\,(\bar{x}_k - x_k) \le -\frac{1}{c_k}\|\bar{x}_k - x_k\|^2$$

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k)$$

Gradient projection method

$$\bar{x}_k = \mathrm{P}_X[x_k - c_k \nabla f(x_k)]$$

If $\alpha_k = 1$, then,

$$x_{k+1} = \bar{x}_k = \mathrm{P}_X[x_k - c_k \nabla f(x_k)]$$

Gradient projection reduces to a steepest descent step when

$$x_k - c_k \nabla f(x_k) \in X$$

Gradient methods – constrained opt

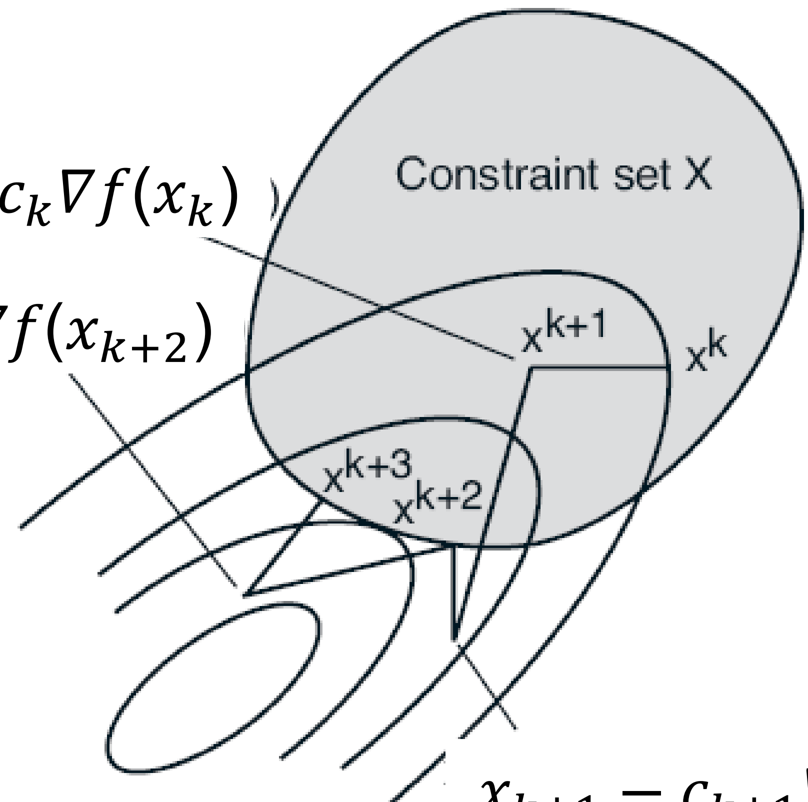
$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k)$$

Gradient projection method

$$x_{k+1} = \mathrm{P}_X[x_k - c_k \nabla f(x_k)]$$

$$x_{k+1} = x_k - c_k \nabla f(x_k)$$

$$x_{k+2} - c_{k+2} \nabla f(x_{k+2})$$



$$x_{k+1} - c_{k+1} \nabla f(x_{k+1})$$

Gradient methods – constrained opt

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k)$$

Gradient projection method: $x_{k+1} = \mathrm{P}_X[x_k - c_k \nabla f(x_k)]$

The algorithm stops when

$$x_{k+1} = \mathrm{P}_X[x_k - c_k \nabla f(x_k)] = x_k$$

and this occurs if and only if $x_{k+1} = x_k = x^*$ is a stationary point.

Gradient methods – constrained opt

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k)$$

Gradient projection method: $x_{k+1} = \mathrm{P}_X[x_k - c_k \nabla f(x_k)]$

The algorithm stops when

$$x_{k+1} = \mathrm{P}_X[x_k - c_k \nabla f(x_k)] = x_k$$

and this occurs if and only if $x_{k+1} = x_k = x^*$ is a stationary point.

Proof: a stationary point has to satisfy

$$\nabla f(x^*)'(x - x^*) \geq 0, \forall x \in X$$

which is equivalent to

$$((x^* - \gamma \nabla f(x^*)) - x^*)'(x - x^*) \leq 0, \forall x \in X, \forall \gamma > 0,$$

Projection over a convex set

Projection Theorem

Let $z \in R^n$ and $X \subseteq R^n$ be a non-empty closed convex set.

Then, we have that:

$x^* \in X$ is the projection of z on X, i.e., $x^* = \mathrm{P}_X[z]$, if and only if

$$(z - x^*)'(x - x^*) \leq 0, \forall x \in X$$

Gradient methods – constrained opt

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k)$$

Gradient projection method: $x_{k+1} = \mathrm{P}_X[x_k - c_k \nabla f(x_k)]$

The algorithm stops when

$$x_{k+1} = \mathrm{P}_X[x_k - c_k \nabla f(x_k)] = x_k$$

and this occurs if and only if $x_{k+1} = x_k = x^*$ is a stationary point.

Proof: a stationary point has to satisfy

$$\nabla f(x^*)'(x - x^*) \geq 0, \forall x \in X$$

which is equivalent to

$$((x^* - \gamma \nabla f(x^*)) - x^*)'(x - x^*) \leq 0, \forall x \in X, \forall \gamma > 0,$$

that is satisfied if and only if x^* is the projection of $\mathrm{z} = x^* - \gamma \nabla f(x^*)$ on X

→$\mathrm{P}_X[x^* - c_k \nabla f(x^*)] = x^*$ is stationary

Choice of the stepsize

- Minimization rule

$$\alpha_k = \operatorname*{argmin}_{\alpha \in [0,1]} f(x_k + \alpha d_k)$$

- Constant stepsize

$$\alpha_k = c, k = 0,1, \ldots$$

- Diminishing stepsize

$$\lim_{k \to \infty} \alpha_k = 0$$

with $\sum_{k=0}^{\infty} \alpha_k = \infty$ and $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$

Outline

- Constrained and convex optimization ☺
- Optimality conditions ☺
- Descent iterative methods: gradient algorithms ☺
- Convergence results

Gradient methods: convergence

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \to R$ is a continuously differentiable function over X

X is a non-empty closed convex subset of R^n

Only convergence to stationary points can be guaranteed.

⬇

In the convex case, convergence to a global minimum can be guaranteed, since each stationary point is a global minimum

Convex optimization

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \to R$ is a continuously differentiable **convex function** over X

X is a non-empty closed convex subset of R^n

Properties:

(a) A local minimum of f over X is also a global minimum. If f is strictly convex, then, there exists at most one global minimum

(b) The optimality conditions are necessary and sufficient for a point to be a global minimum of f over X or, equivalently, all stationary points are global minima

(c) Convergence to a stationary point means convergence to a global minimum

Gradient methods: convergence

Since in gradient methods

$$x_{k+1} = x_k + \alpha_k d_k, k = 0,1, \ldots$$

where d_k satisfies the descent condition

$$\nabla f(x_k)' d_k < 0$$

if d_k tends to be orthogonal to the gradient $\nabla f(x_k)$ when x_k approaches a nonstationary point, then, there is the risk of getting stuck near such a point

Gradient methods: convergence

Since in gradient methods

$$x_{k+1} = x_k + \alpha_k d_k, k = 0,1, \dots$$

where d_k satisfies the descent condition

$$\nabla f(x_k)' d_k < 0$$

if d_k tends to be orthogonal to the gradient $\nabla f(x_k)$ when x_k approaches a nonstationary point, then, there is the risk of getting stuck near such a point

technical conditions are considered on d_k for this not to happen. They are naturally satisfied or enforced in the algorithm

Gradient methods: convergence

Gradient related condition:

For any subsequence $\{x_k\}_{k \in K}$ that converges to a nonstationary point, the corresponding subsequence $\{d_k\}_{k \in K}$ is bounded and satisfies

$$\limsup_{k \to \infty,\, k \in K} \nabla f(x_k)' d_k < 0$$

Gradient methods: convergence

Gradient related condition:

For any subsequence $\{x_k\}_{k\in K}$ that converges to a nonstationary point, the corresponding subsequence $\{d_k\}_{k\in K}$ is bounded and satisfies

$$\limsup_{k\to\infty,\, k\in K} \nabla f(x_k)' d_k < 0$$

This rules out the possibility of converging to a nonstationary point through a sequence characterized by directions d_k orthogonal to the gradient $\nabla f(x_k)$

Convergence results – unconstrained opt

Proposition [stationarity of limit points for gradient methods]

Let $\{x_k\}$ be a sequence generated by a gradient method according to

$$x_{k+1} = x_k + \alpha_k d_k.$$

Assume that $\{d_k\}$ satisfies the gradient related condition and α_k *is chosen by the minimization rule*

$$\alpha_k = \operatorname*{argmin}_{\alpha \geq 0} f(x_k + \alpha d_k).$$

Then, every limit point of $\{x_k\}$ is a stationary point.

Convergence results – unconstrained opt

Proposition [stationarity of limit points for gradient methods]

Let $\{x_k\}$ be a sequence generated by a gradient method according to

$$x_{k+1} = x_k + \alpha_k d_k.$$

Assume that $\{d_k\}$ satisfies the gradient related condition and α_k *is chosen by the minimization rule*

$$\alpha_k = \underset{\alpha \geq 0}{\operatorname{argmin}} f(x_k + \alpha d_k).$$

Then, every limit point of $\{x_k\}$ is a stationary point.

Remark:

$d_k = -D_k \nabla f(x_k)$ with D_k positive definite with bounded eigenvalues, i.e., $c_1 \|z\|^2 \leq z' D_k z \leq c_2 \|z\|^2$, satisfies the gradient related condition

Convergence results – constrained opt

Proposition [stationarity of limit points for gradient methods]

Let $\{x_k\}$ be a sequence generated by a gradient method according to

$$x_{k+1} = x_k + \alpha_k d_k.$$

Assume that $\{d_k\}$ satisfies the gradient related condition and α_k *is chosen by the minimization rule*

$$\alpha_k = \operatorname*{argmin}_{\alpha \in [0,1]} f(\,x_k + \alpha d_k).$$

Then, every limit point of $\{x_k\}$ is a stationary point.

Convergence results – constrained opt

Proposition [stationarity of limit points for gradient methods]

Let $\{x_k\}$ be a sequence generated by a gradient method according to

$$x_{k+1} = x_k + \alpha_k d_k.$$

Assume that $\{d_k\}$ satisfies the gradient related condition and α_k *is chosen by the minimization rule*

$$\alpha_k = \underset{\alpha\in[0,1]}{\operatorname{argmin}} f(x_k + \alpha d_k).$$

Then, every limit point of $\{x_k\}$ is a stationary point.

Remark:

Conditional gradient and gradient projection (with c_k constant) methods satisfy the gradient related condition

Convergence results

What about the constant and diminishing stepize rules?

Convergence results

What about the constant and diminishing stepize rules?

Some onvergence results have been proven under some regularity assumption on the gradient (Lipschitz continuity):

i) f continuously differentiable

ii) there exists $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \forall x, y \in R^n$$

Convergence results – unconstrained opt

Proposition [convergence for a constant stepsize]

Let $\{x_k\}$ be a sequence generated by a gradient method

$$x_{k+1} = x_k + \alpha_k d_k.$$

Assume that $\{d_k\}$ satisfies the gradient related condition and that the gradient is Lipschitz continuous with constant $L > 0$. If there exists ε such that for all k

$$0 < \varepsilon \leq \alpha_k \leq \frac{(2-\varepsilon)|\nabla f(x_k)' d_k|}{L\|d_k\|^2}$$

Then, every limit point of $\{x_k\}$ is a stationary point.

Convergence results – unconstrained opt

Proposition [convergence for a constant stepsize]

Let $\{x_k\}$ be a sequence generated by a gradient method

$$x_{k+1} = x_k + \alpha_k d_k.$$

Assume that $\{d_k\}$ satisfies the gradient related condition and that the gradient is Lipschitz continuous with constant $L > 0$. If there exists ε such that for all k

$$0 < \varepsilon \le \alpha_k \le \frac{(2-\varepsilon)|\nabla f(x_k)' d_k|}{L\|d_k\|^2}$$

Then, every limit point of $\{x_k\}$ is a stationary point.

Remark:

$d_k = -D_k \nabla f(x_k)$ with D_k positive definite with bounded eigenvalues, i.e., $c_1\|z\|^2 \le z' D_k z \le c_2\|z\|^2$, $c_2 \ge c_1 > 0$, satisfies the gradient related condition and $0 < \varepsilon \le \alpha_k \le \frac{(2-\varepsilon)c_2}{L{c_1}^2}$

Convergence results – constrained opt

Convergence results for the constant stepsize case are specific to the considered method.

Here, we consider the gradient projection method and provide statement and proof.

We first need to show an instrumental lemma.

Convergence results – gradient projection

Descent Lemma

If the gradient ∇f is Lipschitz continuous with constant $L > 0$, then,

$$f(x+y) - f(x) \leq y'\nabla f(x) + \frac{L}{2}\|y\|^2, \forall x, y$$

Convergence results – gradient projection

Descent Lemma

If the gradient ∇f is Lipschitz continuous with constant $L > 0$, then,

$$f(x+y) - f(x) \leq y'\nabla f(x) + \frac{L}{2}\|y\|^2, \forall x, y$$

Proof. Set $\mathrm{g}(\alpha) = f(x + \alpha y)$. Then,

$$f(x+y) - f(x) = g(1) - g(0) = \int_0^1 \frac{dg}{d\alpha}(\alpha) d\alpha$$

$$= \int_0^1 y'\nabla f(x + \alpha y) d\alpha$$

Convergence results – gradient projection

Descent Lemma

If the gradient ∇f is Lipschitz continuous with constant $L > 0$, then,

$$f(x+y) - f(x) \leq y'\nabla f(x) + \frac{L}{2}\|y\|^2, \forall x, y$$

Proof. Set $\mathrm{g}(\alpha) = f(x + \alpha y)$. Then,

$$f(x+y) - f(x) = g(1) - g(0) = \int_0^1 \frac{dg}{d\alpha}(\alpha) d\alpha$$

$$= \int_0^1 y'\nabla f(x + \alpha y) d\alpha$$

$$\leq \int_0^1 y'\nabla f(x) d\alpha + \left| \int_0^1 y'\big(\nabla f(x + \alpha y) - \nabla f(x)\big) d\alpha \right|$$

Descent Lemma

If the gradient ∇f is Lipschitz continuous with constant $L > 0$, then,

$$f(x+y) - f(x) \leq y'\nabla f(x) + \frac{L}{2}\|y\|^2, \forall x, y$$

Proof. Set $g(\alpha) = f(x + \alpha y)$. Then,

$$f(x+y) - f(x) = g(1) - g(0) = \int_0^1 \frac{dg}{d\alpha}(\alpha) d\alpha$$

$$= \int_0^1 y'\nabla f(x + \alpha y) d\alpha$$

$$\leq \int_0^1 y'\nabla f(x) d\alpha + \left| \int_0^1 y'\big(\nabla f(x + \alpha y) - \nabla f(x)\big) d\alpha \right|$$

$$\leq y'\nabla f(x) + \int_0^1 \|y\|\, L\, \alpha \|y\| d\alpha = y'\nabla f(x) + \frac{L}{2}\|y\|^2$$

Convergence results – gradient projection

Proposition [convergence for a constant stepsize]

Let $\{x_k\}$ be a sequence generated by the gradient projection method

$$x_{k+1} = \mathrm{P}_X[x_k - c\nabla f(x_k)]$$

Suppose that the gradient ∇f is Lipschitz continuous over X with constant $L > 0$. Then, if

$$0 < c < \frac{2}{L}$$

every limit point of $\{x_k\}$ is stationary.

Convergence results – gradient projection

Proposition [convergence for a constant stepsize]

Let $\{x_k\}$ be a sequence generated by the gradient projection method

$$x_{k+1} = \mathrm{P}_X[x_k - c\nabla f(x_k)]$$

Suppose that the gradient ∇f is Lipschitz continuous over X with constant $L > 0$. Then, if

$$0 < c < \frac{2}{L}$$

every limit point of $\{x_k\}$ is stationary.

Remark: if f is continuously differentiable and X is compact, then, Lipschitz continuity of the gradient is guaranteed.

Convergence results – gradient projection

Proof.

By the descent lemma

$$f(x+y) - f(x) \leq y'\nabla f(x) + \frac{L}{2}\|y\|^2, \forall x, y$$

if we set $x = x_k$ and $y = x_{k+1} - x_k$, we get

$$f(x_{k+1}) - f(x_k) \leq (x_{k+1} - x_k)'\nabla f(x_k) + \frac{L}{2}\|x_{k+1} - x_k\|^2$$

Convergence results – gradient projection

Proof:

Observe now that $x_{k+1} = \mathrm{P}_X[x_k - c\nabla f(x_k)]$ so that by the projection theorem

$$(x_k - c\nabla f(x_k) - x_{k+1})'(x - x_{k+1}) \leq 0, \forall x \in X$$

If we set $x = x_k$, we obtain

$$\nabla f(x_k)'\,(x_{k+1} - x_k) \leq -\frac{1}{c}\|x_{k+1} - x_k\|^2$$

Convergence results – gradient projection

Proof:

By combining the two inequalities that we have just proven, i.e.,

$$f(x_{k+1}) - f(x_k) \leq (x_{k+1} - x_k)' \nabla f(x_k) + \frac{L}{2} \|x_{k+1} - x_k\|^2$$

$$\nabla f(x_k)' \, (x_{k+1} - x_k) \leq -\frac{1}{c} \|x_{k+1} - x_k\|^2$$

we get

$$f(x_{k+1}) \leq f(x_k) - \left(\frac{1}{c} - \frac{L}{2}\right) \|x_{k+1} - x_k\|^2$$

Convergence results – gradient projection

Proof:

By combining the two inequalities that we have just proven, i.e.,

$$f(x_{k+1}) - f(x_k) \le (x_{k+1} - x_k)' \nabla f(x_k) + \frac{L}{2}\|x_{k+1} - x_k\|^2$$

$$\nabla f(x_k)' \, (x_{k+1} - x_k) \le -\frac{1}{c}\|x_{k+1} - x_k\|^2$$

we get

$$f(x_{k+1}) \le f(x_k) - \left(\frac{1}{c} - \frac{L}{2}\right)\|x_{k+1} - x_k\|^2$$

If $0 < c < \frac{2}{L}$, then, if x^* is the limit point of a subsequence $\{x_k\}_{k\in\mathcal{K}}$ we have that $f(x_k) \downarrow f(x^*)$ and then

$$\lim_{k\to\infty}\|x_{k+1} - x_k\|^2 = 0$$

Convergence results – gradient projection

Proof:

$$\lim_{k\to\infty} \|x_{k+1} - x_k\|^2 = 0$$

where $x_{k+1} = \mathrm{P}_X[x_k - c\nabla f(x_k)] = T(x_k)$.

By the continuity of the projection map, it then follows that x^* satisfies $T(x^*) = x^*$ and, hence, it is stationary.

Convergence results – unconstrained opt

Proposition [convergence for a diminishing stepsize]

Let $\{x_k\}$ be a sequence generated by a gradient method

$$x_{k+1} = x_k + \alpha_k d_k.$$

Assume that the gradient ∇f is Lipschitz continuous with constant $L > 0$ and that there exist positive scalars c_1 and c_2 such If there exists ε such that for all k

$$c_1 \, \|\nabla f(x_k)\|^2 \leq -\nabla f(x_k)' d_k, \|d_k\|^2 \leq c_2 \|\nabla f(x_k)\|^2$$

Then, if a diminishing stepsize is adopted, every limit point of $\{x_k\}$ is a stationary point.

Convergence results – unconstrained opt

Proposition [convergence for a diminishing stepsize]

Let $\{x_k\}$ be a sequence generated by a gradient method

$$x_{k+1} = x_k + \alpha_k d_k.$$

Assume that the gradient ∇f is Lipschitz continuous with constant $L > 0$ and that there exist positive scalars c_1 and c_2 such If there exists ε such that for all k

$$c_1 \|\nabla f(x_k)\|^2 \leq -\nabla f(x_k)' d_k, \|d_k\|^2 \leq c_2 \|\nabla f(x_k)\|^2$$

Then, if a diminishing stepsize is adopted, every limit point of $\{x_k\}$ is a stationary point.

Remarks:

$d_k = -D_k \nabla f(x_k)$ with D_k positive definite with bounded eigenvalues satisfies the conditions above.

Similar results hold for the constrained optimization case.

Outline

- Constrained and convex optimization ☺
- Optimality conditions ☺
- Descent iterative methods: gradient algorithms ☺
- Convergence results ☺