Math Camp 1: Functional analysis

About the primer

Goal To briefly review concepts in functional analysis that will be used throughout the course.* The following concepts will be described

1. Function spaces

2. Metric spaces

3. Dense subsets

4. Linear spaces

5. Linear functionals

*The definitions and concepts come primarily from "Introductory Real Analysis" by Kolmogorov and Fomin (highly recommended).

6. Norms and semi-norms of linear spaces

7. Euclidean spaces

8. Orthogonality and bases

9. Separable spaces

10. Complete metric spaces

11. Hilbert spaces

12. Riesz representation theorem

13. Convex functions

14. Lagrange multipliers

Function space

A **function space** is a space made of functions. Each function in the space can be thought of as a point. Examples:

1. $C[a,b]$, the set of all real-valued *continuous* functions in the interval $[a,b]$;

2. $L_1[a,b]$, the set of all real-valued functions whose absolute value is integrable in the interval $[a,b]$;

3. $L_2[a,b]$, the set of all real-valued functions square integrable in the interval $[a,b]$

Note that the functions in 2 and 3 are not necessarily continuous!

Metric space

By a **metric space** is meant a pair (X, ρ) consisting of a space X and a distance ρ, a single-valued, nonnegative, real function $\rho(x, y)$ defined for all $x, y \in X$ which has the following three properties:

1. $\rho(x, y) = 0$ *iff* $x = y$;

2. $\rho(x, y) = \rho(y, x)$;

3. Triangle inequality: $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$

Examples

1. The set of all real numbers with distance

$$\rho(x,y) = |x-y|$$

is the metric space $\mathbb{R}^1$.

2. The set of all ordered n-tuples

$$x = (x_1, ..., x_n)$$

of real numbers with distance

$$\rho(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

is the metric space $\mathbb{R}^n$.

3. The set of all functions satisfying the criteria

$$\int f^2(x)dx < \infty$$

with distance

$$\rho(f_1(x), f_2(x)) = \sqrt{\int (f_1(x) - f_2(x))^2 dx}$$

is the metric space $L_2(\mathbb{R})$.

4. The set of all probability densities with Kullback-Leibler *divergence*

$$\rho(p_1(x), p_2(x)) = \int \ln \frac{p_1(x)}{p_2(x)} p_1(x) dx$$

is not a metric space. The divergence is not symmetric

$$\rho(p_1(x), p_2(x)) \neq \rho(p_2(x), p_1(x)).$$

Dense

A point $x \in \mathbb{R}$ is called a *contact point* of a set $A \in \mathbb{R}$ if every ball centered at x contains at least one point of A. The set of all contact points of a set A denoted by $\bar{A}$ is called the *closure* of A.

Let A and B be subspaces of a metric space $\mathbb{R}$. A is said to be **dense** in B if $B \subset \bar{A}$. In particular A is said to be *everywhere dense* in $\mathbb{R}$ if $\bar{A} = R$.

Examples

1. The set of all rational points is dense in the real line.

2. The set of all polynomials with rational coefficients is dense in $C[a, b]$.

3. The RKHS induced by the gaussian kernel on $[a, b]$ in dense in $L_2[a, b]$

Note: A hypothesis space that is dense in L_2 is a desired property of any approximation scheme.

Linear space

A set L of elements $x, y, z, ...$ is a **linear space** if the following three axioms are satisfied:

1. Any two elements $x, y \in L$ uniquely determine a third element in $x + y \in L$ called the sum of x and y such that
   (a) $x + y = y + x$ (commutativity)
   (b) $(x + y) + z = x + (y + z)$ (associativity)
   (c) An element $0 \in L$ exists for which $x + 0 = x$ for all $x \in L$
   (d) For every $x \in L$ there exists an element $-x \in L$ with the property $x + (-x) = 0$

2. Any number α and any element $x \in L$ uniquely determine an element $\alpha x \in L$ called the product such that
   (a) $\alpha(\beta x) = \beta(\alpha x)$
   (b) $1x = x$

3. Addition and multiplication follow two distributive laws
   (a)$(\alpha + \beta)x = \alpha x + \beta x$
   (b)$\alpha(x + y) = \alpha x + \alpha y$

Linear functional

A functional, $\mathcal{F}$, is a function that maps another function to a real-value

$$\mathcal{F} : f \rightarrow \mathbb{R}.$$

A linear functional defined on a linear space L, satisfies the following two properties

1. Additive: $\mathcal{F}(f+g) = \mathcal{F}(f) + \mathcal{F}(g)$ for all $f, g \in L$

2. Homogeneous: $\mathcal{F}(\alpha f) = \alpha \mathcal{F}(f)$

Examples

1. Let $\mathbb{R}^n$ be a real n-space with elements $x = (x_1, ..., x_n)$, and $a = (a_1, ..., a_n)$ be a fixed element in $\mathbb{R}^n$. Then

$$\mathcal{F}(x) = \sum_{i=1}^{n} a_i x_i$$

is a linear functional

2. The integral

$$\mathcal{F}[f(x)] = \int_a^b f(x)p(x)dx$$

is a linear functional

3. Evaluation functional: another linear functional is the

Dirac delta function

$$\delta_t[f(\cdot)] = f(t).$$

Which can be written

$$\delta_t[f(\cdot)] = \int_a^b f(x)\delta(x-t)dx.$$

4. Evaluation functional: a positive definite kernel in a RKHS

$$\mathcal{F}_t[f(\cdot)] = (K_t, f) = f(t).$$

This is simply the reproducing property of the RKHS.

Normed space

A **normed** space is a linear (vector) space N in which a norm is defined. A nonnegative function $\|\cdot\|$ is a norm *iff* $\forall f, g \in N$ and $\alpha \in \mathbb{R}$

1. $\|f\| \geq 0$ and $\|f\| = 0$ *iff* $f = 0$;

2. $\|f + g\| \leq \|f\| + \|g\|$;

3. $\|\alpha f\| = |\alpha| \; \|f\|$.

Note, if all conditions are satisfied except $\|f\| = 0$ *iff* $f = 0$ then the space has a seminorm instead of a norm.

Measuring distances in a normed space

In a normed space N, the distance ρ between f and g, or a *metric*, can be defined as

$$\rho(f, g) = \|g - f\|.$$

Note that $\forall f, g, h \in N$

1. $\rho(f, g) = 0$ *iff* $f = g$.

2. $\rho(f, g) = \rho(g, f)$.

3. $\rho(f, h) \leq \rho(f, g) + \rho(g, h)$.

Example: continuous functions

A norm in $C[a,b]$ can be established by defining

$$\|f\| = \max_{a \le t \le b} |f(t)|.$$

The distance between two functions is then measured as

$$\rho(f,g) = \max_{a \le t \le b} |g(t) - f(t)|.$$

With this metric, $C[a,b]$ is denoted as C.

Examples (cont.)

A norm in $L_1[a, b]$ can be established by defining

$$\|f\| = \int_a^b |f(t)|dt.$$

The distance between two functions is then measured as

$$\rho(f, g) = \int_a^b |g(t) - f(t)|dt.$$

With this metric, $L_1[a, b]$ is denoted as L_1.

Examples (cont.)

A norm in $C_2[a, b]$ and $L_2[a, b]$ can be established by defining

$$\|f\| = \left(\int_a^b f^2(t)dt \right)^{1/2}.$$

The distance between two functions now becomes

$$\rho(f, g) = \left(\int_a^b (g(t) - f(t))^2 dt \right)^{1/2}.$$

With this metric, $C_2[a, b]$ and $L_2[a, b]$ are denoted as C_2 and L_2 respectively.

Euclidean space

A **Euclidean** space is a linear (vector) space E in which a dot product is defined. A real valued function $(\cdot, \cdot)$ is a dot product *iff* $\forall f, g, h \in E$ and $\alpha \in \mathbb{R}$

1. $(f, g) = (g, f)$;

2. $(f + g, h) = (f, h^*) + (g, h)$ and $(\alpha f, g) = \alpha(f, g)$;

3. $(f, f) \geq 0$ and $(f, f) = 0$ *iff* $f = 0$.

A Euclidean space becomes a *normed linear space* when equipped with the norm

$$\|f\| = \sqrt{(f, f)}.$$

Orthogonal systems and bases

A set of nonzero vectors $\{x_\alpha\}$ in a Euclidean space E is said to be an *orthogonal system* if

$$(x_\alpha, x_\beta) = 0 \quad \text{f}or \;\; \alpha \neq \beta$$

and an *orthonormal system* if

$$\begin{aligned} (x_\alpha, x_\beta) &= 0 \quad \text{f}or \;\; \alpha \neq \beta \\ (x_\alpha, x_\beta) &= 1 \quad \text{f}or \;\; \alpha = \beta. \end{aligned}$$

An orthogonal system $\{x_\alpha\}$ is called an **orthogonal basis** if it is complete (the smallest closed subspace containing $\{x_\alpha\}$ is the whole space E). A complete orthonormal system is called an **orthonormal basis**.

Examples

1. $\mathbb{R}^n$ is a real n-space, the set of n-tuples $x = (x_1, ..., x_n)$, $y = (y_1, ..., y_n)$. If we define the dot product as

$$(x, y) = \sum_{i=1}^{n} x_i y_i$$

we get Euclidean n-space. The corresponding norms and distances in $\mathbb{R}^n$ are

$$\begin{aligned} \|x\| &= \sqrt{\sum_{i=1}^{n} x_i^2} \\ \rho(x, y) = \|x - y\| &= \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \end{aligned}$$

The vectors

$$\begin{aligned} e_1 &= (1, 0, 0, ..., 0) \\ e_2 &= (0, 1, 0, ..., 0) \\ &\dots\dots\dots\dots\dots\dots \\ e_n &= (0, 0, 0, ..., 1) \end{aligned}$$

form an orthonormal basis in $\mathbb{R}^n$.

2. The space l_2 with elements $x = (x_1, x_2, ..., x_n, ...)$, $y = (y_1, y_2, ..., y_n, ...)$, ..., where

$$\sum_{i=1}^{\infty} x_i^2 < \infty, \ \sum_{i=1}^{\infty} y_i^2 < \infty, \ ..., \ ...,$$

becomes an infinite-dimensional Euclidean space when equipped with the dot product

$$(x, y) = \sum_{i=1}^{\infty} x_i y_i.$$

The simplest orthonormal basis in l_2 consists of vectors

$$
\begin{aligned}
e_1 &= (1, 0, 0, 0, ...) \\
e_2 &= (0, 1, 0, 0, ...) \\
e_3 &= (0, 0, 1, 0, ...) \\
e_4 &= (0, 0, 0, 1, ...) \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}
$$

there are an infinite number of these bases.

3. The space $C_2[a, b]$ consisting of all continuous functions on $[a, b]$ equipped with the dot product

$$(f, g) = \int_a^b f(t)g(t)dt$$

is another example of Euclidean space.

An important example of orthogonal bases in this space is the following set of functions

$$1,\ \cos\frac{2\pi nt}{b-a},\ \sin\frac{2\pi nt}{b-a} \quad (n = 1, 2, \ldots).$$

Cauchy-Schwartz inequality

Let H be an Euclidean space. Then $\forall f, g \in H$, it holds

$$|(f,g)| \leq \|f\| \ \|g\|$$

Sketch of the proof. The case $f \propto g$ is trivial, hence let us assume the opposite is true. For all $x \in \mathbb{R}$,

$$0 < (f + xg, f + xg) = x^2 \ \|g\|^2 + 2x \ (f,g) + \|f\|^2,$$

since the quadratic polynomial of x above has no zeroes, the discriminant Δ must be negative

$$0 > \Delta/4 = (f,g)^2 - \|f\|^2 \ \|g\|^2.$$

Separable

A metric space is said to be **separable** if it has a countable everywhere dense subset.

Examples:

1. The spaces $\mathbb{R}^1$, $\mathbb{R}^n$, $L_2[a, b]$, and $C[a, b]$ are all separable.

2. The set of real numbers is separable since the set of rational numbers is a countable subset of the reals and the set of rationals is is everywhere dense.

Completeness

A sequence of functions f_n is *fundamental* if $\forall \epsilon > 0$ $\exists N_\epsilon$ such that

$$\forall n \text{ and } m > N_\epsilon, \quad \rho(f_n, f_m) < \epsilon.$$

A metric space is **complete** if all fundamental sequences converge to a point in the space.

C, L^1, and L^2 are complete. That C_2 is not complete, instead, can be seen through a counterexample.

Incompleteness of C_2

Consider the sequence of functions $(n = 1, 2, ...)$

$$\phi_n(t) = \begin{cases} -1 & \text{if } -1 \le t < -1/n \\ nt & \text{if } -1/n \le t < 1/n \\ 1 & \text{if } 1/n \le t \le 1 \end{cases}$$

and assume that ϕ_n converges to a continuous function ϕ in the metric of C_2. Let

$$f(t) = \begin{cases} -1 & \text{if } -1 \le t < 0 \\ 1 & \text{if } 0 \le t \le 1 \end{cases}$$

Incompleteness of C_2 (cont.)

Clearly,

$$\left(\int (f(t)-\phi(t))^2 dt\right)^{1/2} \leq \left(\int (f(t)-\phi_n(t))^2 dt\right)^{1/2} + \left(\int (\phi_n(t)-\phi(t))^2 dt\right)^{1/2}.$$

Now the l.h.s. term is strictly positive, because $f(t)$ is not continuous, while for $n \to \infty$ we have

$$\int (f(t)-\phi_n(t))^2 dt \to 0.$$

Therefore, contrary to what assumed, ϕ_n cannot converge to ϕ in the metric of C_2.

Completion of a metric space

Given a metric space $\mathbb{R}$ with closure $\bar{\mathbb{R}}$, a complete metric space $\mathbb{R}^*$ is called a **completion** of $\mathbb{R}$ if $\mathbb{R} \subset \mathbb{R}^*$ and $\bar{\mathbb{R}} = \mathbb{R}^*$.

Examples

1. The space of real numbers is the completion of the space of rational numbers.

2. L_2 is the completion of the functional space C_2.

Hilbert space

A **Hilbert space** is a Euclidean space that is *complete*, *separable*, and generally *infinite-dimensional*.

A Hilbert space is a set H of elements $f, g, ...$ for which

1. H is a Euclidean space equipped with a scalar product

2. H is complete with respect to metric $\rho(f, g) = \|f - g\|$

3. H is separable (contains a countable everywhere dense subset)

4. (generally) H is infinite-dimensional.

l_2 and L_2 are examples of Hilbert spaces.

Evaluation functionals

A linear evaluation functional is a linear functional $\mathcal{F}_t$ that *evaluates* each function in the space at the point t, or

$$\mathcal{F}_t[f] = f(t)$$

$$\mathcal{F}_t[f + g] = f(t) + g(t).$$

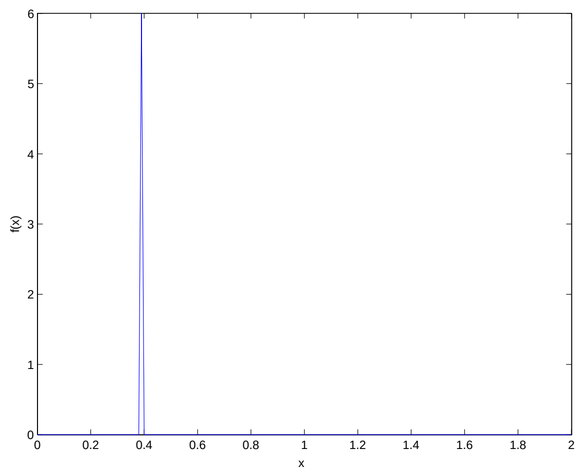
The functional is bounded if there exists a M s.t.

$$|\mathcal{F}_t[f]| = |f(t)| \leq M\|f\|_{Hil} \;\; \forall t$$

for all f where $\| \cdot \|_{Hil}$ is the norm in the Hilbert space.

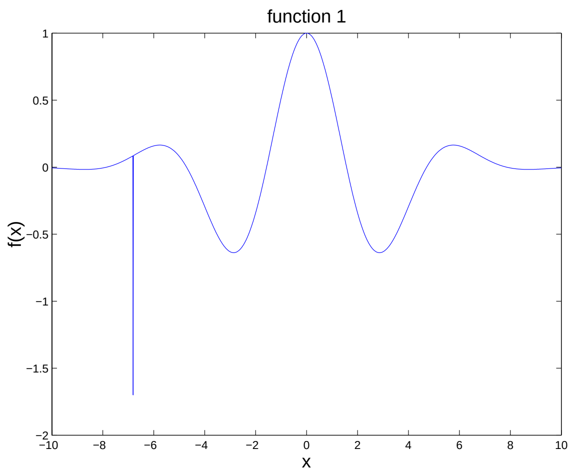
Evaluation functionals in Hilbert space

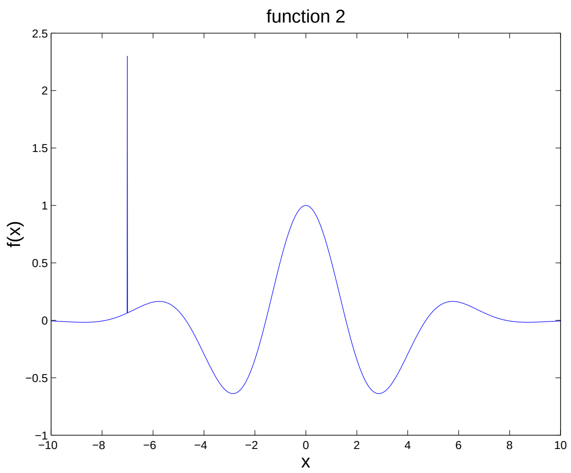
The evaluation functional is not bounded in the familiar Hilbert space $L_2([0,1])$, no such M exists and in fact elements of $L_2([0,1])$ are not even defined pointwise.



Evaluation functionals in Hilbert space

In the following pictures the two functions have the same norm but they are very different on sets of zero measure





Riesz Representation Theorem

For every *bounded* linear functional $\mathcal{F}$ on a Hilbert space $\mathcal{H}$, there is a unique $v \in \mathcal{H}$ such that

$$\mathcal{F}[x] = (x, v)_{\mathcal{H}}, \qquad \forall x \in \mathcal{H}$$

Convex sets

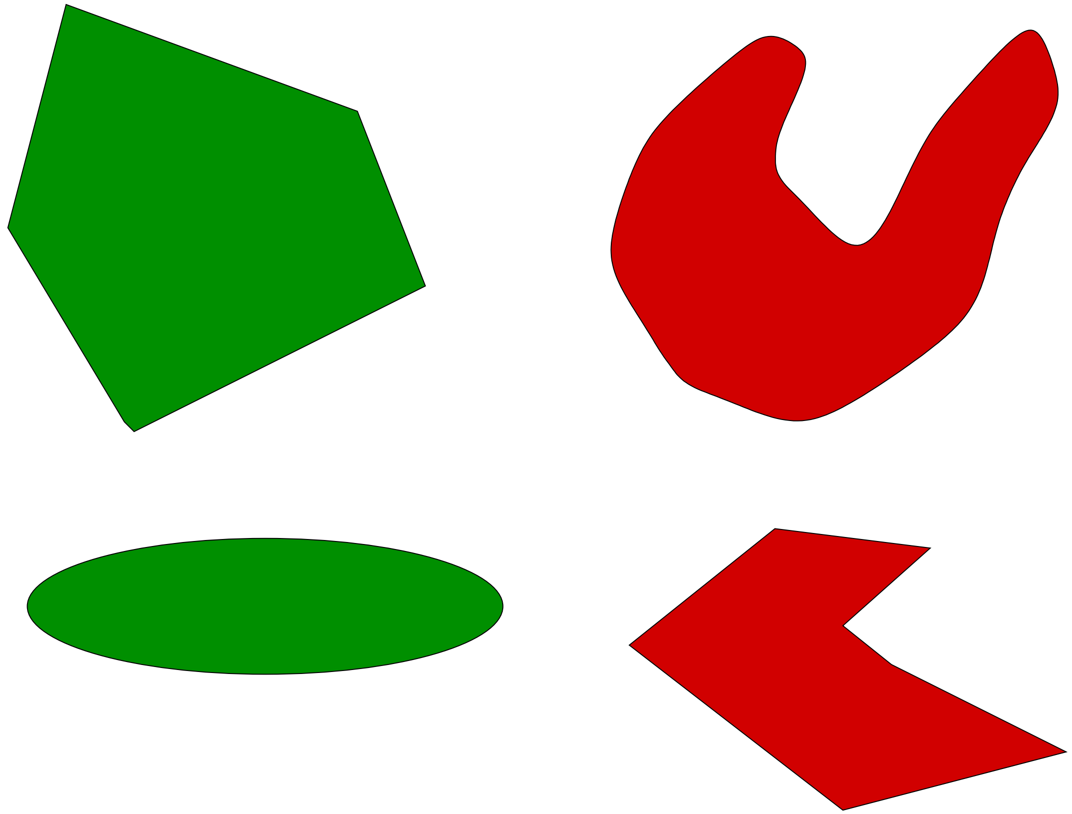
A set $X \in \mathbb{R}^n$ is **convex** if

$$\forall x_1, x_2 \in X,\ \forall \lambda \in [0,1],\ \lambda x_1 + (1-\lambda) x_2 \in X.$$

A set is convex if, given any two points in the set, the line segment connecting them lies entirely inside the set.

Convex and Non-convex sets

Convex Sets | Non-Convex Sets



Convex Functions

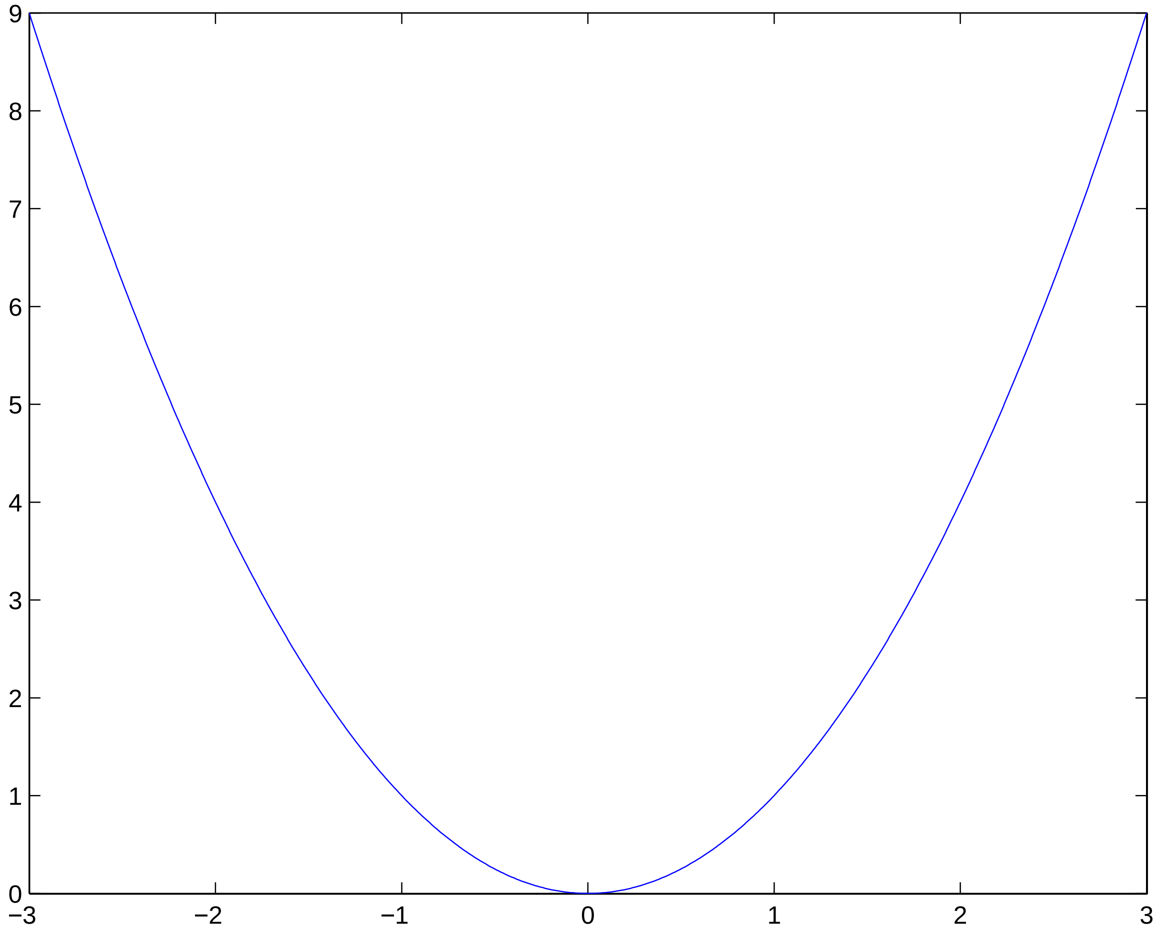
A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if:

For any x_1 and x_2 in the domain of f, for any $\lambda \in [0, 1]$,

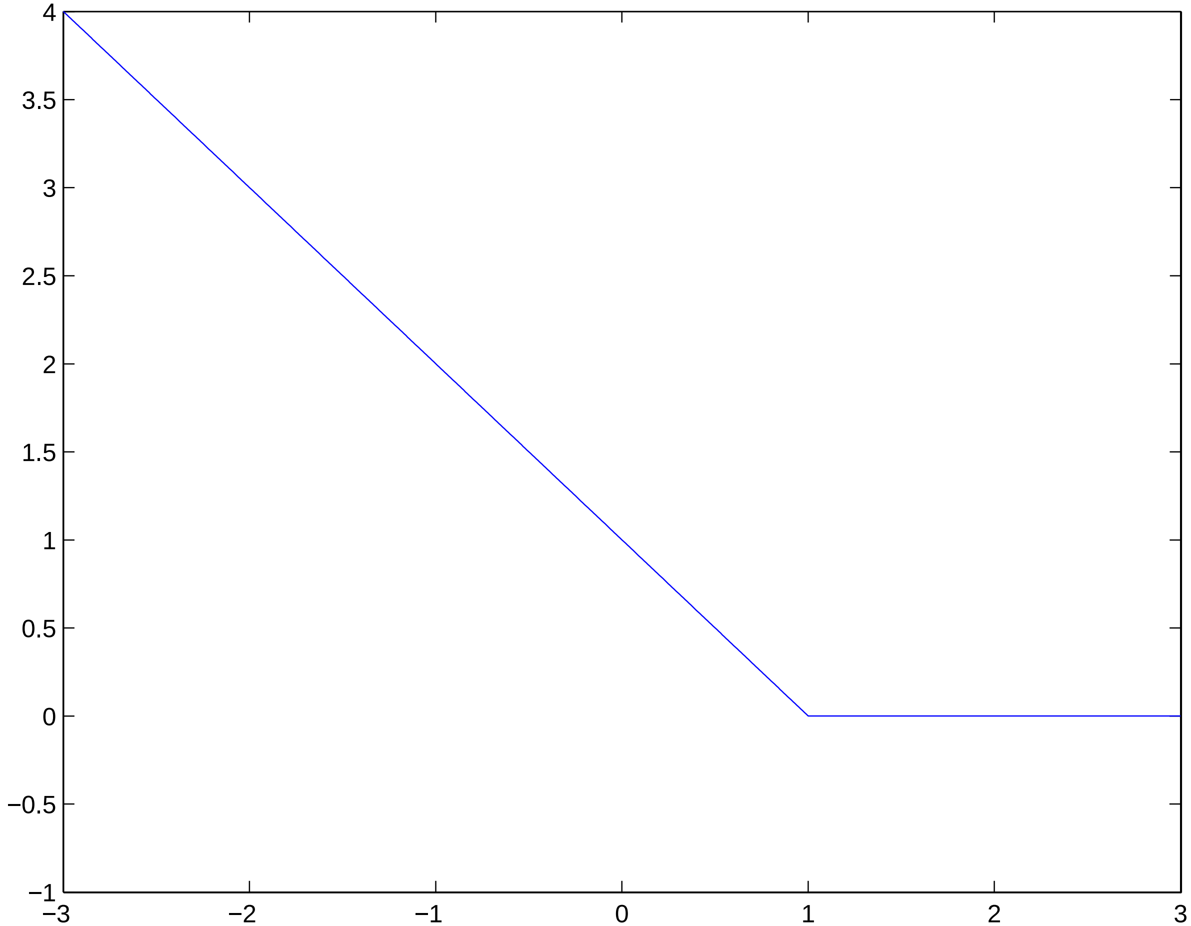
$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).$$

A function is strictly convex if we replace "$\leq$" with "$<$".

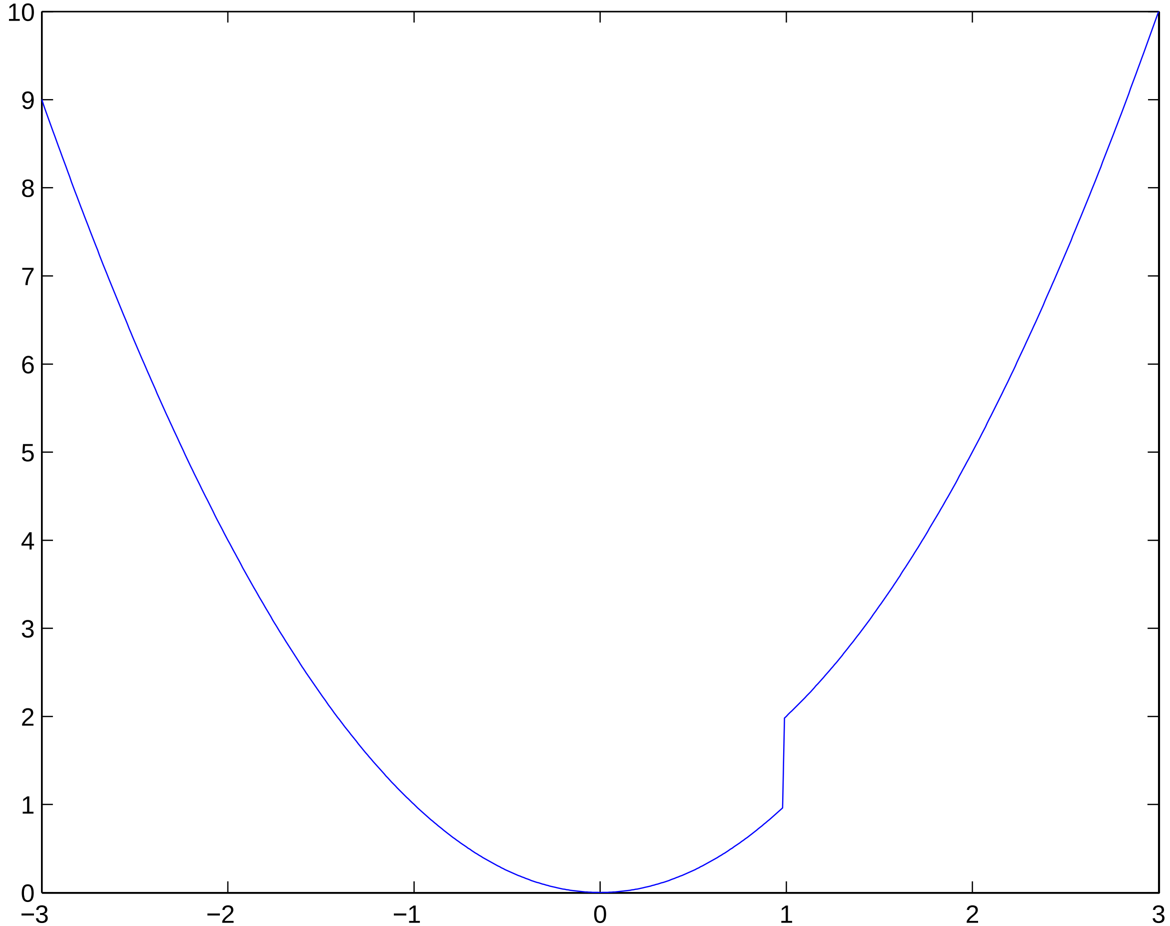
A Strictly Convex Function



A Convex Function



A Non-Convex Function



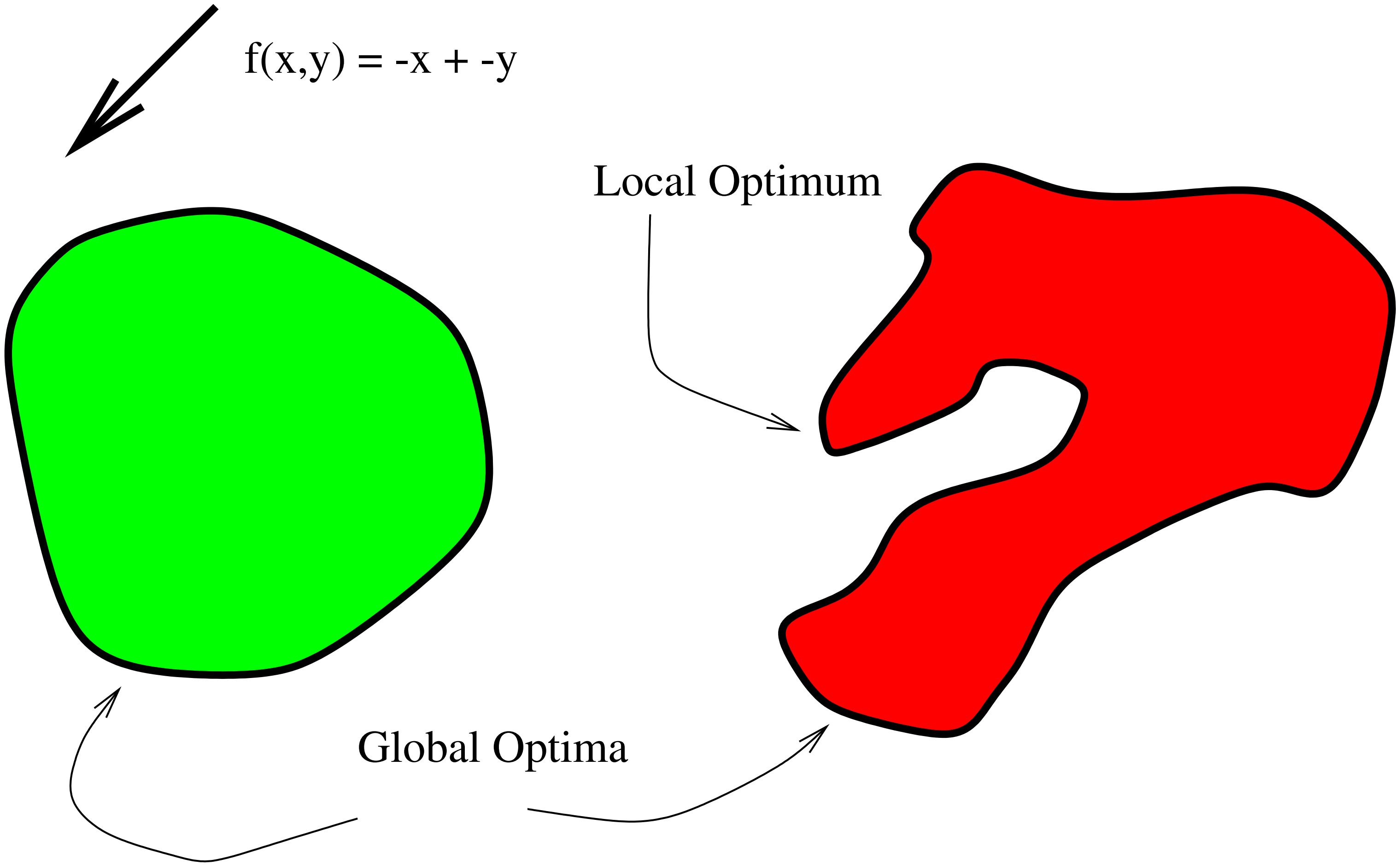
Why We Like Convex Functions

Unconstrained convex functions (convex functions where the domain is all of $\mathbb{R}^n$) are easy to minimize. Convex functions are differentiable almost everywhere. Directional derivatives always exist. If we cannot improve our solution by moving locally, we are at the optimum. If we cannot find a direction that improves our solution, we are at the optimum.

Why We Like Convex Sets

Convex functions over convex sets (a convex domain) are also easy to minimize. If the set and the functions are both convex, if we cannot find a direction which we are able to move in which decreases the function, we are done. Local optima are global optima.

Optimizing a Convex Function Over a Convex and a Non-Convex Set



Existence and uniqueness of minimum

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a *strictly convex* function.

The function f is said to be *coercive* if

$$\lim_{\|x\| \to +\infty} f(x) = +\infty.$$

Strictly convex and coercive functions have exactly one local (global) minimum.

Local condition on the minimum

If the convex function f is differentiable, its gradient ∇f is null on the minimum x_0.

Even if the gradient does not exist, the *subgradient* ∂f always exists.

The subgradient of f in x is defined by

$$\partial f(x) = \{w \in \mathbb{R}^n | \forall x' \in \mathbb{R}^n,\ f(x') \geq f(x) + w \cdot (x' - x)\},$$

On the minimum x_0, it holds

$$0 \in \partial f(x_0),$$

Lagrange multiplier's technique

Lagrange multiplier's technique allows the reduction of the constrained minimization problem

$$\begin{array}{ll} \textbf{Minimize} & I(x) \\ \textbf{subject to} & \Phi(x) \le m \qquad\qquad \text{(for some } m\text{)} \end{array}$$

to the unconstrained minimization problem

$$\textbf{Minimize} \quad J(x) = I(x) + \lambda\Phi(x) \qquad\qquad \text{(for some } \lambda \ge 0\text{)}$$

Geometric intuition

The fact that ∇I does not vanish in the interior of the domain implies that the constrained minimum $\bar{\mathbf{x}}$ must lie on the domain's boundary (the level curve $\Phi(x) = m$). Therefore, at the point $\bar{\mathbf{x}}$ the component of ∇I along the tangent to the curve $\Phi = m$ vanishes.

But since the tangent to $\Phi = m$ is orthogonal to $\nabla\Phi$, we have that at the point $\bar{\mathbf{x}}$, $\nabla\Phi$ and ∇I are parallel, or

$$\nabla I(\bar{\mathbf{x}}) \propto \nabla\Phi(\bar{\mathbf{x}}).$$

Geometric intuition (Cont)

We thus introduce a parameter $\lambda \geq 0$, called **Lagrange multiplier**, and consider the problem of finding the unconstrained minimum x_λ of

$$J(x) = I(x) + \lambda\Phi(x)$$

as a function of λ.

By setting $\nabla J = 0$, we actually look for the points where ∇I and $\nabla\Phi$ are parallel. The idea is to find all such points and then check which of them lie on the curve $\Phi = m$.