Basic notions

Real vector spaces

The course will deal with convex optimization, i.e., minimization of convex functions defined on convex sets. A set \(X\) is convex if for every pair of points \(x, y \in X\) it contains also the segment \([x, y] = \{x + (1 - \lambda)y \mid \lambda \in [0, 1]\}\). We see that in order for convexity to make sense, we have to know how to define multiplication of points by real scalars and addition of points. This naturally leads to the notion of a real vector space.

**Definition 0.1.** A set \(V\) together with operations \(\cdot : V \times \mathbb{R} \to V\), \(+ : V \times V \to V\) is called a real vector space if

- \(+\) is commutative, associative, and possesses a neutral element 0, the null vector,
- the distributive laws \((a + b) \cdot u = a \cdot u + b \cdot u\) and \(a \cdot (u + v) = a \cdot u + a \cdot v\) for \(a, b \in \mathbb{R}, u, v \in V\) hold,
- \(0 \cdot u = 0, 1 \cdot u = u\) for every \(u \in V\),
- \((ab) \cdot u = a \cdot (b \cdot u)\) for \(a, b \in \mathbb{R}, u \in V\).

**Example:** The prime example of a real vector space is the space \(\mathbb{R}^n\) of \(n\)-tuples \((x_1, \ldots, x_n)^T\) of real scalars. Addition and multiplication on this space are defined componentwise.

**Definition 0.2.** A map \(f : U \to V\) between two real vector spaces is called linear if

\[f(a \cdot u + b \cdot v) = a \cdot f(u) + b \cdot f(v) \quad a, b \in \mathbb{R}, u, v \in U.\]

It is called an isomorphism if it is linear and bijective. An automorphism of a vector space is an isomorphism of the space onto itself.

The automorphisms of \(\mathbb{R}^n\) are the linear maps given by non-singular real \(n \times n\) matrices \(A\) (non-singular means \(\det A \neq 0\)). These maps form a group, the general linear group \(GL(n, \mathbb{R})\).

**Definition 0.3.** Let \(V\) be a vector space and \(\{u_1, \ldots, u_k\}\) a set of vectors. A vector \(v \in V\) is called a linear combination of the vectors \(u_1, \ldots, u_k\) if \(v = \sum_{i=1}^k \lambda_i u_i\) for some real scalars \(\lambda_1, \ldots, \lambda_k\).

A finite set \(\{u_1, \ldots, u_k\}\) of vectors in a vector space \(V\) is called linearly independent if the relation \(\sum_{i=1}^k \lambda_i u_i = 0\) implies \(\lambda_i = 0\) for all \(i = 1, \ldots, k\).

A (finite) basis of \(V\) is a linearly independent set \(\{u_1, \ldots, u_n\}\) such that every element of \(V\) is a linear combination of \(u_1, \ldots, u_n\).

Not every vector space possesses a finite basis. Those vector spaces which possess a finite basis are called finite-dimensional and the cardinality of the basis is their dimension. It is a fundamental fact of linear algebra that all bases of a vector space have the same cardinality. The space \(\mathbb{R}^n\) has dimension \(n\).

We will deal with optimization problems in finite-dimensional spaces only.

**Definition 0.4.** Let \(X \subset V\) be a subset of a vector space. The linear hull of \(X\) is the set of all linear combinations of elements of \(X\).

A linear subspace of \(V\) is a subset which equals its linear hull.

The linear hull of an arbitrary subset \(X\) is the smallest linear subspace of \(V\) which contains \(X\), or equivalently, the intersection of all linear subspaces containing \(X\).

Norms

Optimization problems are thus defined on real vector spaces. They rarely can be solved exactly. Most often a solution algorithm delivers a sequence of iterates that converge to an optimal solution of the problem. In order to define convergence and to measure the quality of approximation, we shall need the notions of topology and norm.
Definition 0.5. Let $V$ be a real vector space. A function $\cdot : V \times V \to \mathbb{R}$ is called a scalar product if

- $u \cdot v = v \cdot u$ for all $u, v \in V$,
- $(au + bv) \cdot w = a(u \cdot w) + b(v \cdot w)$ for all $a, b \in \mathbb{R}$, $u, v, w \in V$,
- $u \cdot u \geq 0$ for all $u \in V$ and $u \cdot u = 0$ if and only if $u = 0$.

Equivalently, a scalar product is a symmetric positive definite bilinear form on $V$.

Definition 0.6. Let $V$ be a real vector space. A function $\| \cdot \| : V \to \mathbb{R}_+$ is called a norm if

- $\| u \| = 0$ if and only if $u = 0$,
- $\| au \| = |a| \| u \|$ for all $a \in \mathbb{R}$ and $u \in V$,
- $\| u + v \| \leq \| u \| + \| v \|$ for all $u, v \in V$.

Every scalar product on $V$ defines a norm by $\| u \| = \sqrt{u \cdot u}$, but not every norm can be represented in such a way. Norms defined by scalar products are called Euclidean.

Every norm on $V$ defines a distance function, or metric, by $d(u, v) := \| u - v \|$.

Two norms $\| \cdot \|, \| \cdot \|'$ are called strongly equivalent if there exist constants $\alpha, \beta > 0$ such that

$$\alpha \| u \| \leq \| u \|' \leq \beta \| u \|$$

for all $u \in V$.

The unit ball and the open unit ball of a norm $\| \cdot \|$ are the sets

$$B_1 = \{ u \in V \mid \| u \| \leq 1 \}, \quad B_0^o = \{ u \in V \mid \| u \| < 1 \},$$

respectively.

Example: The space $\mathbb{R}^n$ can be equipped with a Euclidean distance function, which is induced by the 2-norm $\| a \|_2 = \sqrt{\sum_{k=1}^{n} a_k^2}$.

We can define also other norms on $\mathbb{R}^n$, the most important are the 1-norm $\| a \| = \sum_{k=1}^{n} | a_k |$ and the $\infty$-norm $\| a \|_\infty = \max k | a_k |$. These norms are members of the family of $p$-norms defined by $\| a \|_p := (\sum_{k=1}^{n} | a_k |^p)^{1/p}$. All these norms are strongly equivalent.

Figure 1: Unit balls of some norms

Topology

Defining a topology on a set allows one to define the notion of convergence of sequences of points to points of the set. This is necessary since usually the output of an optimization algorithm will not consist of an optimal solution of the optimization problem, but merely of a sequence of iterates which converge to an optimal solution.
Definition 0.7. Let $X$ be a set. A collection $\mathcal{T}$ of subsets $U \subset X$ is called a topology on $X$ if

- $\emptyset, X \in \mathcal{T}$,
- finite intersections of sets in $\mathcal{T}$ are again in $\mathcal{T}$,
- arbitrary unions of sets in $\mathcal{T}$ are again in $\mathcal{T}$.

The set $X$ equipped with a topology $\mathcal{T}$ is called a topological space.

The topology defines which subsets of $X$ are open and which are closed: $U \subset X$ is open if $U \in \mathcal{T}$ and it is closed if $X \setminus U \in \mathcal{T}$. If $x \in X$ is a point, then any open set containing $x$ is called a neighbourhood of $x$.

In this way, finite unions of closed sets and arbitrary intersections of closed sets are again closed.

Definition 0.8. Let $U \subset X$ be a subset of a topological space. The interior of $U$, denoted $\text{int} U$ or $U^o$, is the largest open set contained in $U$, or equivalently, the union of all open sets contained in $U$.

The closure of $U$, denoted $\text{cl} U$ or $\overline{U}$, is the smallest closed set containing $U$, or equivalently, the intersection of all closed sets containing $U$.

The boundary of $U$, denoted $\partial U$, is the difference $\overline{U} \setminus U^o$.

Example: Let $X = \mathbb{R}$ and $U = \mathbb{Q}$ the subset of rational numbers. Then $U^o = \emptyset$ and $\overline{U} = \mathbb{R}$.

Definition 0.9. Let $U, V \subset X$ be subsets of a topological space. The $U$ is called dense in $V$ if $U \subset V \subset \overline{U}$.

The topology allows to define the notion of convergence.

Definition 0.10. Let $X$ be a topological space. A sequence $x_1, x_2, \ldots$ of points converges to a point $x^*$ if for every open set $U \subset X$ containing $x^*$, there exists a number $N_U$ such that $x_k \in U$ for all $k \geq N_U$.

Thus $\{x_k\}$ converges to $x^*$ if the sequence eventually enters and no more leaves arbitrary small neighbourhoods of $x^*$.

Often it is convenient to define or describe topologies by the following simpler notion.

Definition 0.11. Let $X$ be a set. A collection $\mathcal{B}$ of subsets $U \subset X$ is called a base if

- $\bigcup_{U \in \mathcal{B}} = X$,
- for every $U_1, U_2 \in \mathcal{B}$ and every $x \in U_1 \cap U_2$ there exists $U_3 \in \mathcal{B}$ such that $x \in U_3 \subset U_1 \cap U_2$.

Every base defines a topology by taking the open subsets of $X$ to be arbitrary unions of elements of $\mathcal{B}$. On the other hand, every topology can be defined by a base, the largest such base being the topology itself.

A norm on a vector space $V$ induces a topology on $V$ by the base

$$\mathcal{B} = \{u + \varepsilon B_1 | u \in V, \varepsilon > 0\}.$$  

Thus a set $U \subset V$ is open if for every $u \in U$ there exists a constant $\varepsilon > 0$ such that $u + \varepsilon B_1 \subset U$.

The notion of convergence can then be reformulated as follows: a sequence $\{x_k\}$ converges to $x^*$ if and only if $\lim_{k \to \infty} ||x_k - x^*|| = 0$.

Strongly equivalent norms define the same topology on $V$. In finite dimension all norms are strongly equivalent, hence every norm defines the same topology on $V$.
Coordinate systems

The raw data of an optimization problem is often not adapted to processing by numerical algorithms. Different entries of the data may, p.ex., correspond to different physical quantities which have completely different orders. However, the user has the freedom to rescale the inputs or even to replace them by linear or affine combinations by introducing a tailored coordinate system. This step is called *preconditioning*.

Every $n$-dimensional vector space $V$ over $\mathbb{R}$ is isomorphic to $\mathbb{R}^n$, since an isomorphism can be constructed by sending some basis of $V$ to some basis of $\mathbb{R}^n$. Any isomorphism $\iota : V \rightarrow \mathbb{R}^n$ assigns a coordinate vector $\iota(v) = (v_1, \ldots, v_n)$ to every element $v \in V$ and in this way introduces a coordinate system on $V$. If a coordinate system on $V$ is given by some isomorphism $\iota$, then the norms and the corresponding distances on $\mathbb{R}^n$ can be carried over to $V$ by means of $\iota$, $||u|| := ||\iota(u)||$.

The isomorphism $\iota : V \rightarrow \mathbb{R}^n$ is not unique, and from a given norm on $\mathbb{R}^n$ different isomorphisms define different norms on $V$.

Affine space

We have seen that the notion of convexity relies on the ability to define the segment $[x, y] = \{\lambda x + (1 - \lambda)y \mid \lambda \in [0, 1]\}$ between points of the set under consideration. Obviously the notion of segment is invariant not only under automorphisms of the underlying vector space, but also under translations which do not preserve the zero vector and are hence not automorphisms. We therefore do not need the full structure of a vector space in order to work with convex sets. It is sufficient to keep the structure of an affine space, which can be obtained from a vector space by forgetting the location of the zero vector.

**Definition 0.12.** Let $V$ be a real vector space. An affine space with associated vector space $V$ is a set $A$ together with a map $+: A \times V \rightarrow A$ such that

- $x + 0 = x$ for all $x \in A$,
- $(x + u) + v = x + (u + v)$ for all $x \in A$, $u, v \in V$,
- $v \mapsto x + v$ is a bijection between $V$ and $A$ for all $x \in A$.

*Example:* An affine subspace $A$ of a vector space $W$ is an affine space in the sense above. The associated vector space $V$ is the linear subspace of $W$ spanned by the differences of the elements of $A$. In particular, the vector space $W$ itself is also an affine space, with associated the vector space being again $W$.

The third property of the definition allows to consider the elements of $V$ as differences between points in $A$: for $x, y \in A$ we define $x - y$ to be the unique vector $v \in V$ such that $x = y + v$. The points of the segment $[x, y]$, where $x, y \in A$, can then be written as

$$\lambda x + (1 - \lambda)y = y + \lambda(x - y).$$

Since $x - y$ is a vector, its multiple $\lambda(x - y)$ is also a vector, and hence the right-hand side of the relation is again a point of the affine space. Generally, every combination of points of affine space with coefficients summing to 0 can be interpreted as a vector in $V$, since it can be written as a linear combination of differences of elements of $A$. Therefore every combination with coefficients summing to 1 can be seen again as a point in affine space, because it can be written as an element of affine space plus a combination of points with coefficients summing to 0. We have the following analoga of Definitions 0.3 and 0.4.

**Definition 0.13.** Let $x_1, \ldots, x_k$ be points in an affine space $A$. Then $\sum_{i=1}^{k} \lambda_i x_i$ is called an affine combination of the points $x_1, \ldots, x_k$ if $\sum_{i=1}^{k} \lambda_i = 1$.

A finite set $\{x_1, \ldots, x_k\}$ of points in an affine space $A$ is called affinely independent if the relations $\sum_{i=1}^{k} \lambda_i u_i = 0$, $\sum_{i=1}^{k} \lambda_i = 0$ imply $\lambda_i = 0$ for all $i = 1, \ldots, k$.

An affine basis of an affine space $A$ is an affinely independent set $\{x_1, \ldots, x_n\}$ such that every element of $A$ is an affine combination of $x_1, \ldots, x_n$. 


If \( \{x_1, \ldots, x_n\} \) is an affine basis of \( A \), then \( \{x_2 - x_1, \ldots, x_n - x_1\} \) is a basis of the vector space associated to \( A \). Hence the dimension of \( A \) is equal to \( n - 1 \), i.e., the number of elements in any of its affine bases, minus 1.

**Definition 0.14.** The affine hull of a subset \( X \subset A \) of an affine space is the set of all affine combinations of elements of \( X \).

An affine subspace of \( A \) is a subset of \( A \) which equals its affine hull.

The affine hull of an arbitrary subset \( X \) is the smallest affine subspace of \( A \) which contains \( X \), or equivalently, the intersection of all affine subspaces containing \( X \).

Given an affine basis \( \{x_1, \ldots, x_n\} \) of \( A \), we can represent every point \( x \in A \) uniquely as an affine combination \( x = \sum_{i=1}^{n} \lambda_i x_i \). The coefficients \( \lambda_1, \ldots, \lambda_n \) are called the barycentric coordinates of \( x \) with respect to the basis \( \{x_1, \ldots, x_n\} \).

**Definition 0.15.** A map \( f : A \rightarrow B \) between affine spaces \( A, B \) with associated vector spaces \( U, V \) is called affine if the map \( \tilde{f} : U \rightarrow V \) defined by \( \tilde{f} : x - y \mapsto f(x) - f(y) \) is well-defined and linear.

An affine isomorphism is a bijective affine map.

Here "well-defined" means that if two differences \( x - y, x' - y' \) amount to the same vector, then we should also have \( f(x) - f(y) = f(x') - f(y') \).