EE3001 Machine Learning **Fall** 2025

Lecture 04. Convex Sets

Lecturer: Xiaojun Chang Date: Oct 9, 2025

Key distinction is not linear vs. nonlinear, but convex or. nonconvex.

R. Tyrrell Rockafellar

1 Introduction

Many popular machine learning models take the form of

$$\min_{\mathbf{w}} f(\mathbf{w}) + \lambda\Omega(\mathbf{w}),$$

where f is the so-called loss function that measures how well the model fits the training data, Ω is a regularization term, and $\lambda > 0$ is the regularization parameter. When f is the least squares loss and Ω is the square of the ℓ_2 norm of the model parameters, we have the well-known ridge regression:

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_2^2. \tag{1}$$

If we replace the regularization term in (1) by the ℓ_1 norm, we have another popular model, that is, Lasso, as follows.

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1. \tag{2}$$

We have seen that, the ridge regression admits a closed form solution if the data matrix $\mathbf{X}$ has full column rank, while the computational cost can be expensive as it involves finding the inverse of a large-scale matrix. Noticing that the objective function in (1) is differentiable, we can use the classical gradient descent method to iteratively find a solution up to a given accuracy. However, this approach does not work for the Lasso problem in (2), as the regularizer is not differentiable.

Problems like (2) involving nondifferentiable terms are the so-called nonsmooth problems, which consist of a major research topic—called sparse learning—in machine learning. To deal with the nonsmooth problems, we need to equip us with a suite of new tools. In the next couple of lectures, we study a type of optimization problems—that is, convex optimization problems—which includes many popular sparse learning models as special cases.

2 Affine Sets

Definition 1. A set $C \subseteq \mathbb{R}^n$ is *affine* if the line through any two distinct points in C lies in C, i.e., if for any $\mathbf{x}_1, \mathbf{x}_2 \in C$, where $\mathbf{x}_1 \neq \mathbf{x}_2$, and $\theta \in \mathbb{R}$, we have $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in C$.

Definition 2. A point x is called an *affine combination* of points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ if there exists $\theta_1, \theta_2, \ldots, \theta_m \in \mathbb{R}$ such that

$$\mathbf{x} = \theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 + \ldots + \theta_m\mathbf{x}_m$$

and

$$\theta_1 + \theta_2 + \ldots + \theta_m = 1.$$

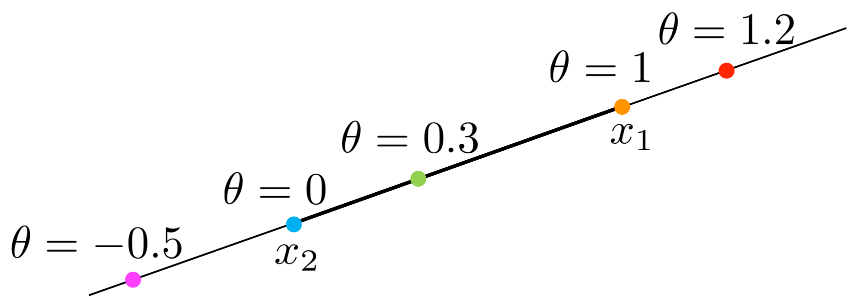

Figure 1: The line passing through x_1 and x_2 is described parametrically by $\theta x_1 + (x - \theta)x_2$, where θ goes over the real line.

If C is an affine set and $\mathbf{x}_0 \in C$, then the set

$$V = C - \mathbf{x}_0 = \{\mathbf{x} - \mathbf{x}_0 : \mathbf{x} \in C\}$$

is a subspace. Thus, we can also describe the affine set C by

$$C = V + \mathbf{x}_0 = \{\mathbf{v} + \mathbf{x}_0 : \mathbf{v} \in V\}.$$

The dimension of an affine set C is the dimension of the subspace $V = C - \mathbf{x}_0$, where $\mathbf{x}_0$ is an arbitrary point in C.

Example 1 (Solution set of linear equations)**.** Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. The solution set $C = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\}$ is an affine set.

Definition 3. The *affine hull* of a set C is the set of all affine combinations of points in C, which is denoted $\mathbf{aff}\ C$:

$$\mathbf{aff}\ C = \{\theta_1 \mathbf{x}_1 + \cdots + \theta_k \mathbf{x}_k : \mathbf{x}_1, \ldots, \mathbf{x}_k \in C, \theta_1 + \cdots + \theta_k = 1\}.$$

The *affine dimension* of a set C is the dimension of its affine hull.

Proposition 1. *The affine hull of set C is the smallest affine set that contains C.*

Definition 4. The *relative interior* of the set C, denoted $\mathbf{relint}\ C$, is its interior relative to $\mathbf{aff}\ C$:

$$\mathbf{relint}\ C = \{\mathbf{x} \in C : \exists\, r > 0, B(\mathbf{x}, r) \cap \mathbf{aff}\ C \subseteq C\},$$

where $B(\mathbf{x}, r) = \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| \leq r\}$ is the ball of radius r and centered at x. The relative boundary of C is defined as $\bar{C} \setminus \mathbf{relint}\ C$, where $\bar{C}$ is the closure of C.

3 Convex Sets

Definition 5. In $\mathbb{R}^n$, a point $\mathbf{x}$ is a convex combination of the points $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ if

$$\mathbf{x} = \theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2 + \cdots + \theta_k \mathbf{x}_k,$$

where $\theta_i \geq 0$ for $i = 1, \ldots, k$ and

$$\theta_1 + \theta_2 + \ldots + \theta_k = 1.$$

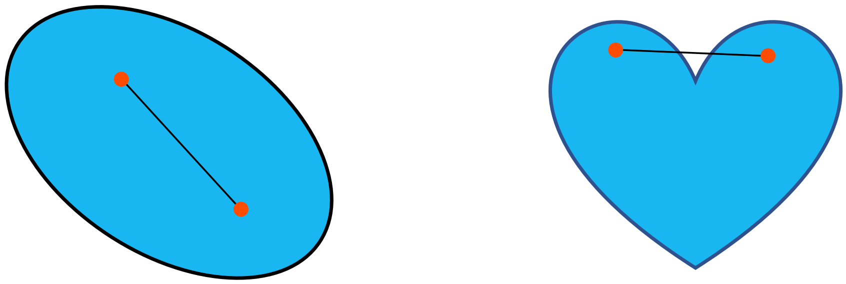

Figure 2: Convex and nonconvex sets.

Definition 6. The **convex hull** of a set $C \subseteq \mathbb{R}^n$, denoted by **conv** C, is the set of all convex combinations of points in C:

$$\textbf{conv}\ C = \left\{ \sum_{i=1}^{k} \theta_i \mathbf{x}_i : \mathbf{x}_i \in C, \theta_i \geq 0, \sum_{i=1}^{k} \theta_i = 1 \right\}.$$

The idea of convex combination can be generalized to include infinite sums, integrals, and, in the most general form, probability distributions [1] (expectation).

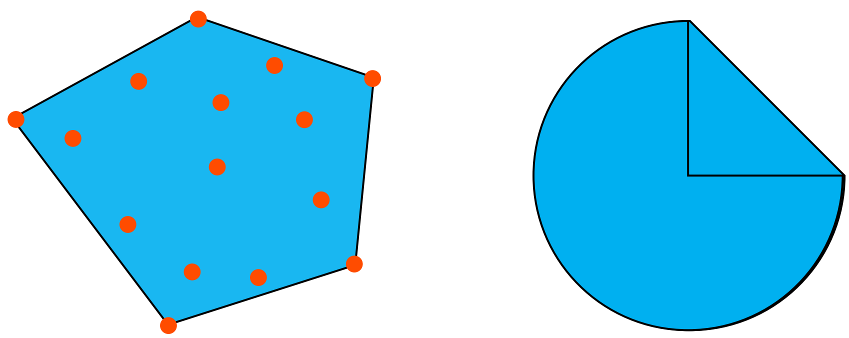

Figure 3: Convex hull.

Definition 7. A set C is **convex** if the line segment between any two points in C lies in C; that is, if $\forall\, \mathbf{x}_1, \mathbf{x}_2 \in C$ and $\forall\, \theta \in [0,1]$, we have

$$\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in C.$$

Example 2. Suppose $p : \mathbb{R}^n \to \mathbb{R}$ satisfies $p(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in C$ and $\int_C p(\mathbf{x}) d\mathbf{x} = 1$, where $C \subseteq \mathbb{R}^n$ is convex. Then

$$\int_C p(\mathbf{x})\mathbf{x} d\mathbf{x} \in C,$$

if the integral exists.

Definition 8. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine if it takes the form of:

$$f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$.

Proposition 2.

1. *The intersection $\cap_{i\in\mathcal{I}} C_i$ of any collection $\{C_i : i \in \mathcal{I}\}$ of convex sets is convex, where $\mathcal{I}$ is an index set.*
2. *The closure and the interior of a convex set are convex.*
3. *The image and the inverse image of a convex set under an affine function are convex.*

Example 3.

1. Hyperplane: $\{\mathbf{x} : \mathbf{a}^\top \mathbf{x} = b\}$, where $\mathbf{a} \neq 0$ and $b \in \mathbb{R}$.
2. Halfspace: $\{\mathbf{x} : \mathbf{a}^\top \mathbf{x} \leq b\}$, where $\mathbf{a} \neq 0$ and $b \in \mathbb{R}$.
3. Norm ball: $\{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_0\| \leq r\}$, where $r > 0$.
4. Polyhedron: $\{\mathbf{x} : \mathbf{a}_i^\top \mathbf{x} \leq b_i, i = 1, \ldots, m\}$, where $\mathbf{a}_i \neq 0$ and $b_i \in \mathbb{R}$ for $i = 1, \ldots, m$.
5. Positive definite matrices $\mathbf{S}^n_{++}$.

Definition 9. A set C is called a *cone*, or *nonnegative homogeneous*, if $\forall\, \mathbf{x} \in C$ and $\theta \in [0, \infty)$, we have $\theta\mathbf{x} \in C$. A set C is a *convex cone* if it is convex and a cone; that is, $\forall\, \mathbf{x}_1, \mathbf{x}_2 \in C$ and $\theta_1, \theta_2 \geq 0$, we have

$$\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 \in C.$$

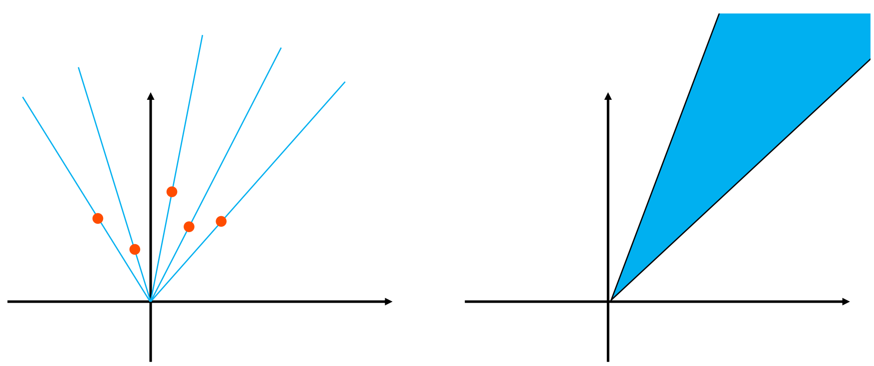

Figure 4: Cones.

- A point of the form $\theta_1\mathbf{x}_1 + \cdots + \theta_m\mathbf{x}_m$ with all nonnegative $\theta_1, \ldots, \theta_m$ is called a *conic combination* (or a *nonnegative linear combination*) of $\mathbf{x}_1, \ldots, \mathbf{x}_m$.

Definition 10. The *conic hull* of a set C is the set of all conic combinations of points in C, i.e., $\forall\, \mathbf{x}_1, \ldots, \mathbf{x}_m \in C$,

$$\{\theta_1\mathbf{x}_1 + \cdots + \theta_m\mathbf{x}_m : \theta_i \geq 0,\ i = 1, \ldots, m\},$$

which is also the smallest convex cone that contains C.

Notice that, a cone is not necessarily a convex set.

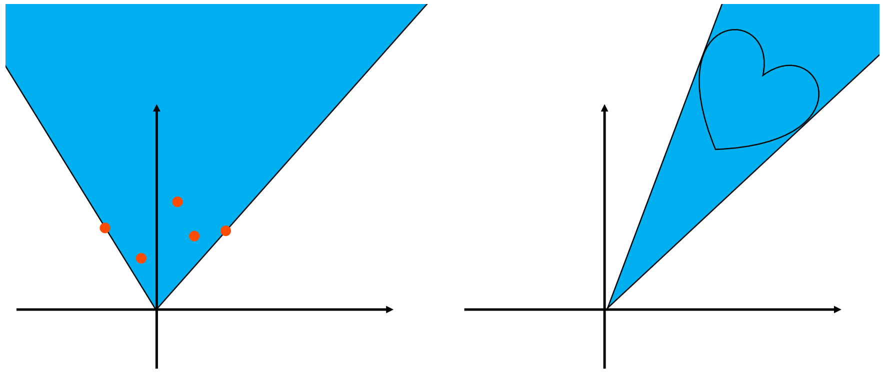

Figure 5: Conic hulls.

4 Operations that Preserve Convexity

Lemma 1. *Let $\mathcal{I}$ be an arbitrary index set. If the sets $S_i \subset \mathbb{R}^n$, $i \in \mathcal{I}$, are convex, then the set $S = \cap_{i\in\mathcal{I}} S_i$ is convex.*

Proof. Let $\mathbf{x}_1, \mathbf{x}_2 \in S$. Thus, $\forall\, i \in \mathcal{I}$, we have $\mathbf{x}_1, \mathbf{x}_2 \in S_i$. As S_i is convex, the line segment between $\mathbf{x}_1$ and $\mathbf{x}_2$ also lies in S_i. Since this applies to all S_i, $i \in \mathcal{I}$, the line segment also lies in their intersection. □

Definition 11. We define the product of a set S by a scalar c to get

$$cS = \{c\mathbf{x} : \mathbf{x} \in S\}.$$

The *Minkowski sum* of two sets is defined by:

$$S_1 + S_2 = \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in S_1, \mathbf{y} \in S_2\}.$$

Lemma 2. *Let S_1 and S_2 be convex sets in $\mathbb{R}^n$ and let $a, b \in \mathbb{R}$. Then, the set $S = aS_1 + bS_2$ is convex.*

Proof. Let $\mathbf{z}_1$, $\mathbf{z}_2 \in S$. The definition of the Minkowski sum implies that, there exist $\mathbf{x}_i, \mathbf{y}_i \in S_i$, $i = 1, 2$, such that

$$\mathbf{z}_1 = a\mathbf{x}_1 + b\mathbf{x}_2 \text{ and } \mathbf{z}_2 = a\mathbf{y}_1 + b\mathbf{y}_2.$$

Then, $\forall\, \theta \in [0, 1]$, we have

$$\theta\mathbf{z}_1 + (1-\theta)\mathbf{z}_2 = a(\theta\mathbf{x}_1 + (1-\theta)\mathbf{y}_1) + b(\theta\mathbf{x}_2 + (1-\theta)\mathbf{y}_2) \in S.$$

Therefore, the set S is convex. □

Lemma 3. *Let $S \subseteq \mathbb{R}^n$ be convex and $f : \mathbb{R}^n \to \mathbb{R}^m$ be an affine function. Then, the image of S under f*

$$f(S) = \{f(\mathbf{x}) : \mathbf{x} \in S\},$$

is convex.

Proof. Let $\mathbf{y}_1, \mathbf{y}_2 \in f(S)$, i.e., $\mathbf{y}_1 = A\mathbf{x}_1 + \mathbf{b}$ and $\mathbf{y}_2 = A\mathbf{x}_2 + \mathbf{b}$. Then,

$$\theta\mathbf{y}_1 + (1-\theta)\mathbf{y}_2 = A(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) + \mathbf{b} \in f(S).$$

□

Lemma 4 (Carathéodory's Lemma [2]). *Suppose that $S \subset \mathbb{R}^n$. Then, every element of* **conv** S *is a convex combination of at most $n+1$ points of S.*

Proof. Let $\mathbf{x} = \sum_{i=1}^m \theta_i \mathbf{x}_i$ be a convex combination of $m > n+1$ points of S. We shall show that m can be reduced by one. If $\theta_i = 0$ for some i, then we are done. Otherwise, assume that all $\theta_i > 0$. As $m > n+1$, we can find $\{\alpha_i\}_{i=1}^m$, not all equal 0, such that

$$\alpha_1 \begin{bmatrix} \mathbf{x}_1 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} \mathbf{x}_2 \\ 1 \end{bmatrix} + \cdots + \alpha_m \begin{bmatrix} \mathbf{x}_m \\ 1 \end{bmatrix} = 0.$$

Let $\tau = \min\{\theta_i/\alpha_i : \alpha_i > 0\}$, $k \in \mathbf{argmin}\{\theta_i/\alpha_i : \alpha_i > 0\}$ and $\theta_i' = \theta_i - \tau\alpha_i$, $i = 1, 2, \ldots, m$. Still, we have $\sum_{i=1}^m \theta_i' = 1$ and $\sum_{i=1}^m \theta_i' \mathbf{x}_i = \mathbf{x}$. The definition of τ leads to a fact that $\theta_k' = 0$ and we can delete the k^{th} point. Repeating the above procedure, we can reduce the number of points to $n+1$. $\square$

References

[1] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[2] A. Ruszczyński. *Nonlinear Optimization.* Princeton University Press, 2006.