Lecture 1 — August 28

Lecturer: Sanghavi *Scribe: Euichan Moon & Tzu-Ling Kan*

In this lecture we give the basic definition of a convex set, a convex function, and convex optimization. As an application of these ideas, we derive the characterization of an optimal solution to an unconstrained convex optimization problem.

1.1 Formulation of Convex Optimization Problems

The overarching goal of learning convex optimization is to learn how to properly formulate a convex optimization problem. After forming the convex optimization problem, deriving the solution becomes only a matter of applying well-established algorithms.

A typical **formulation of a convex optimization problem** is described by

$$\begin{aligned} \text{min}: \quad & f_0(x) \\ \text{s.t.}: \quad & f_i(x) \le 0, \qquad i = 1, \ldots, m. \end{aligned}$$

where each f_j is convex for $j = 0, \ldots, m$. We also often write the problem with an abstract (convex) set constraint,

$$\begin{aligned} \text{min}: \quad & f_0(x) \\ \text{s.t.}: \quad & x \in \mathcal{X}, \end{aligned}$$

where f_0 is a convex function and $\mathcal{X}$ is a convex set.

1.2 Convex Sets

Before we are able to formulate a convex optimization problem, we must understand what constitutes a convex set. We will begin by providing the general definition of a convex set and look at many different sets that are convex.

1.2.1 Definition of a Convex Set

Definition 1. *A* ***convex combination*** *of points $x_1, \ldots, x_k$ is described by*

$$\sum_{i=0}^{k} \theta_i x_i, \tag{1.1}$$

where $\theta_1 + \cdots + \theta_k = 1$ and $\theta_i \ge 0$.

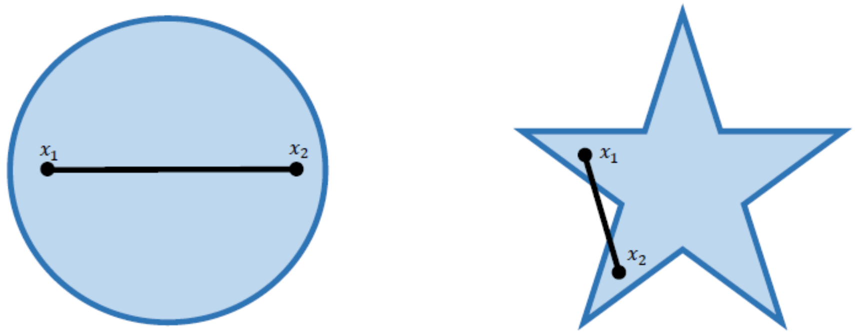

Figure 1.1. A convex set can be easily determined by examining whether the line segment between any two points in the set are in the set. Thus the figure on the left (circle) is in the set whereas the figure on the right (star) is not.

Definition 2. *A set, $\mathcal{X}$, is called a **convex set** if and only if the convex combination of any two points in the set belongs to the set, i.e.*

$$\mathcal{X} \subseteq \mathbb{R}^n \text{ is convex } \Leftrightarrow \text{ if } \forall x_1, x_2 \in \mathcal{X} \text{ and } \forall \lambda \in [0,1],\ \lambda x_1 + (1-\lambda)x_2 \in \mathcal{X}\ . \tag{1.2}$$

1.2.2 Examples of Convex Sets

Definition 3. *A **affine combination** of points $x_1, \ldots, x_k$ is described by*

$$\sum_{i=0}^{k} \theta_i x_i, \tag{1.3}$$

where $\theta_1 + \cdots + \theta_k = 1$. (Note: Affine combination lacks the nonnegative constraint on θ_i).

Definition 4. *A set, $\mathcal{A}$, is called a **affine set** if and only if the affine combination of any two points in the set belongs to the set. In symbols:*

$$\mathcal{A} \subseteq \mathbb{R}^n \text{ is affine } \Leftrightarrow \text{ if } \forall x_1, x_2 \in \mathcal{A},\ \lambda x_1 + (1-\lambda)x_2 \in \mathcal{A}\ . \tag{1.4a}$$

An equivalent definition, using the solution set of a system of linear equations is

$$\mathcal{A} \subseteq \mathbb{R}^n \text{ is affine } \Leftrightarrow \left\{x \mid Ax = b, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^n\right\}. \tag{1.4b}$$

Remark 1. *Affine sets are convex.*

Proof: *Let $x_1, x_2 \in \mathcal{A}$ and $\lambda \in [0,1]$. Taking the convex combination of x_1 and x_2,*

$$\begin{aligned} A(\lambda x_1 + (1-\lambda)x_2) &= \lambda A x_1 + (1-\lambda) A x_2 \\ &= \lambda b + (1-\lambda) b \\ &= b \end{aligned}$$

Since the convex combination of points in $\mathcal{A}$ is also in the set, $\mathcal{A}$ is a convex set. □

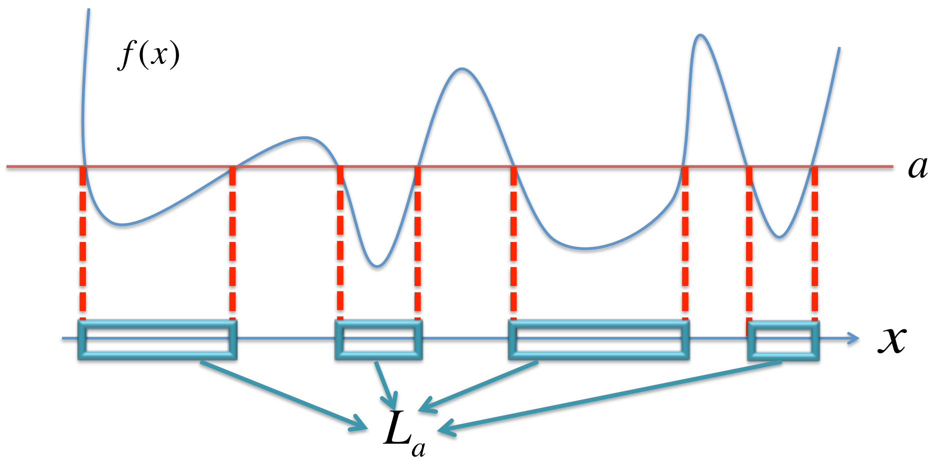

Figure 1.2. Level set L_a of function $f(x)$

Definition 5. *The* ***level set*** *of function f at a is:*

$$L_a = \{x \mid f(x) \leq a\} \tag{1.5}$$

Level set of a convex function is convex for any value of a (note that the empty set is convex by convention). Figure 1.2 shows an example of a level set.

Remark 2. *Level sets are convex.*

Proof: *Basically, we must show that if $x_1, x_2 \in \mathcal{C}$ then any convex combination $\lambda x_1 + (1-\lambda)x_2 \in \mathcal{C}$. If $x_1, x_2 \in \mathcal{C}$ then*

$$\begin{aligned} f(\lambda x_1 + (1-\lambda)x_2) &\overset{(a)}{\leq} \lambda f(x_1) + (1-\lambda)f(x_2)\ \forall \lambda \in [0,1] \\ &\overset{(b)}{\leq} \lambda a + (1-\lambda)a\ \forall \lambda \in [0,1] \\ &= a, \end{aligned}$$

where (a) follows by using the convexity of f and (b) follows from using the definition of the level set. $\square$

Definition 6. *A* ***hyperplane****, $\mathcal{H}$, is a set defined by*

$$\mathcal{H} = \{x \in \mathbb{R}^n \mid s^T x = b\}(s \neq \emptyset), \tag{1.6}$$

where b is the offset and s is the normal vector. If $b = 0$, $\mathcal{H}$ is a $(n-1)$ dimensional subspace.

Remark 3. *Hyperplanes are convex.*

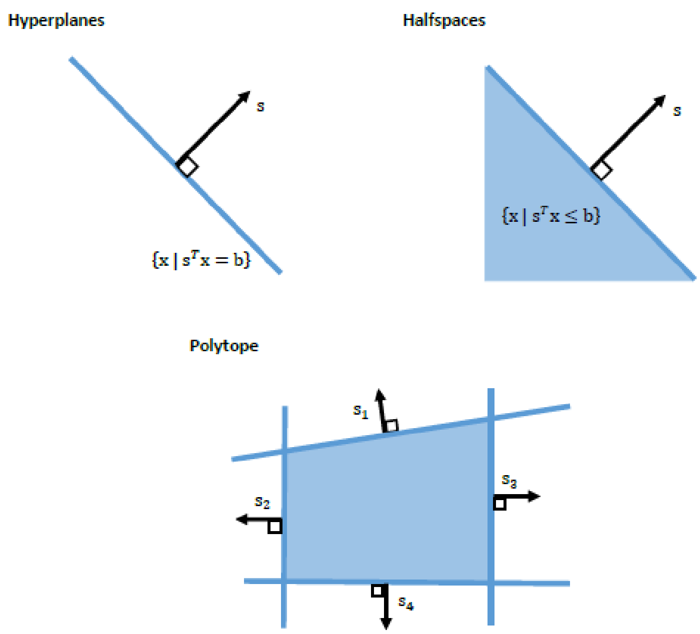

Figure 1.3. Examples of Convex Sets

Proof: *The proof follows the same line of reasoning as that of* **Remark 1**. □

Definition 7. *A* ***halfspace****,* $\mathcal{H}_+$*, is a set defined as*

$$\mathcal{H}_+ = \{x \in \mathbb{R}^n \mid s^T x \le b\}(s \neq \emptyset), \tag{1.7}$$

where b *is the offset and* s *is the normal vector.*

Remark 4. *Halfspaces are convex.*

Proof: *The proof follows the same line of reasoning as that of* **Remark 1**. □

Definition 8. *A* ***polyhedron*** *is an intersection of finite number of halfspaces and hyperplanes. A* ***polytope*** *is a polyhedron that is also bounded.*

Remark 5. *Polyhedrons are convex.*

Proof: *Covex sets are closed under intersections. This is readily seen when taking two points* $x_1, x_2 \in A \cap B$*, where both* A *and* B *are convex sets, and applying the definition of convex sets as in the proof of* **Remark 1**. *Since halfspaces and hyperplanes are convex, polyhedrons must also be convex.* □

A very interesting set that turns out to be useful for many applications, and in general comes up frequently in convex optimization, is the set $\mathcal{S}_+^n$, of symmetric $n \times n$ matrices with non-negative eigenvalues. As a simple exercise, we can show that this set is in fact convex.

Proof: We use the fact that a symmetric matrix M is in $\mathcal{S}_+^n$ iff $x^T M x \geq 0 \; \forall x \in \mathbb{R}^n$. Using this, checking convexity becomes straightforward. For a convex combination of any two matrices $M_1, M_2 \in \mathcal{S}_+^n$ we have

$$\begin{aligned} x^T(\lambda M_1 + (1-\lambda)M_2)x &= \lambda x^T M_1 x + (1-\lambda) x^T M_2 x \\ &\geq 0 \end{aligned}$$

Hence, $\lambda M_1 + (1-\lambda)M_2 \in \mathcal{S}_+^n \; \forall \, \lambda \in [0,1]$. □

Definition 9. *A* ***subspace*** $\mathcal{V} \subset \mathbb{R}^n$ *is a set which is closed under any linear combination of the constituent vectors,* v_i*, or*

$$\mathcal{V} = \{w \mid \sum_{i=1}^{K} \lambda_i v_i = w, \; \forall \lambda_i \in \mathbb{R}, \, v_i \in \mathcal{V}\} \tag{1.8}$$

Remark 6. *Subspaces are convex.*

Proof: *We showed the equivalent definition of an affine set and a system of linear equations in* **Definition 4**. *Since subspaces are affine sets and affine sets are convex, subspaces are convex.* □

Remark 7. *The* ***nullspace*** *of a* $m \times n$ *matrix* A *given by*

$$null(A) = \{v \mid Av = 0\}$$

is a subspace. This can be shown by seeing the trivial proof that nullspaces are closed under vector additions and scalar multiplications.

Remark 8. *The* ***range*** *of a* $m \times n$ *matrix* A *given by*

$$range(A) = \{b \mid Av = b\}$$

is a subspace as taking the linear combination of constituent vectors yeilds

$$\sum_{i=1}^{K} \lambda_i A v_i = A\left(\sum_{i=1}^{K} \lambda_i v_i\right) = A v_0 \in range(A),$$

where $v_0 = \sum_{i=1}^{K} \lambda_i v_i$ *is some new vector.*

Remark 9. *An* ***affine subspace*** $\mathcal{Q}$ *in* $\mathbb{R}^n$ *is a set which is closed under affine combination of the constituent vectors, or*

$$\mathcal{Q} = \{w \mid \sum_{i=1}^{K} \lambda_i q_i = w,\ \forall \lambda_i,\ s.t \sum_{i=1}^{K} \lambda_i = 1,\ q_i \in \mathcal{Q}\}. \tag{1.9}$$

It is trivial to show that affine subspaces are convex.

Definition 10. *A convex hull of a set $\mathcal{C}$ is the set of all convex combinations of points in $\mathcal{C}$. As the name implies, convex hulls are convex. See Fig. 1.4 for an illustration.*

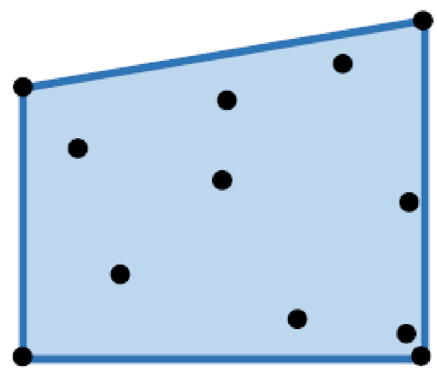

Figure 1.4. Convex hull

1.3 Convex functions

We give three definitions of convex functions: A basic definition that requires no differentiability of the function; a first-order definition of convexity that uses the gradient; and a second-order condition of convexity. We show that (for smooth functions) these three are equivalent. Later in the course, it will be important to deal with non-differentiable functions.

First, we define the set of points where a function is finite.

Definition 11. *The domain of a function $f : \mathbb{R}^n \to \mathbb{R}$ is denoted* $\mathrm{dom}(f)$, *and is defined as the set of points where a function f is finite:*

$$\mathrm{dom}(f) = \{x \in \mathbb{R}^n \,:\, f(x) < \infty\}.$$

Definition 12. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if dom(f) is convex and $\forall x_1, x_2 \in$ dom$(f) \subseteq \mathbb{R}^n$, $\lambda \in [0,1]$, we have:*

$$\lambda f(x_1) + (1-\lambda) f(x_2) \ \geq f(\lambda x_1 + (1-\lambda) x_2) . \tag{1.10}$$

This inequality is illustrated in Figure 1.5.

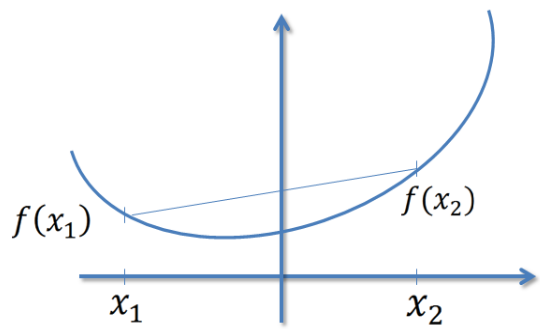

Figure 1.5. Convex functions

We now give the first-order condition for convexity.

Definition 13. *Suppose a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then it is convex if and only if dom(f) is convex and*

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) . \tag{1.11}$$

Intuitively speaking, equation (1.11) states that the first-order Taylor approximation is in fact a global underestimator of the function, as illustrated in figure 1.6.

Proposition 1. *For differentiable functions, definition 12 and definition 13 are equivalent.*

Proof: Consider first a univariate function. Suppose a function f satisfies the derivative-free definition of convexity and $dom(f)$ is convex. Then we know that $\forall x_1, x_2 \in dom(f)$, and $\lambda \in [0,1]$,

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

Rearranging terms and dividing by λ, we obtain:

$$f(x_2) \geq f(x_1) + \frac{f(x_1 + \lambda(x_2 - x_1)) - f(x_1)}{\lambda}.$$

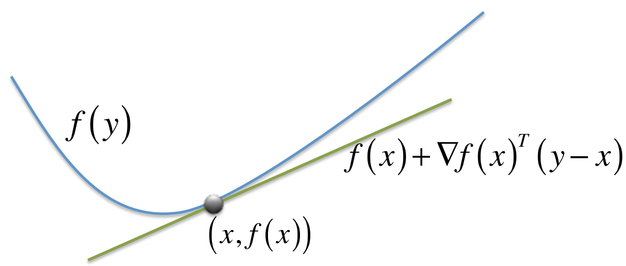

Figure 1.6. If f is convex and differentiable, then eq. (1.11) is a global underestimator of f.

Taking λ to zero, we obtain:

$$f(x_2) \geq f(x_1) + \lim_{\lambda \longrightarrow 0} \frac{f(x_1 + \lambda(x_2 - x_1)) - f(x_1)}{\lambda}. \tag{1.12}$$

$$\lim_{\lambda \longrightarrow 0} \frac{f(x_1 + \lambda(x_2 - x_1)}{\lambda} \overset{(a)}{=} \lim_{\lambda \longrightarrow 0} (x_2 - x_1) f'(x_1 + \lambda(x_2 - x_1)) = (x_2 - x_1) f'(x_1) \tag{1.13}$$

where (a) follows by using L'Hospital's rule. Substitute eq. (1.13) into eq. (1.12), we get the first-order condition for convexity.

Conversely, suppose now that function f satisfies the first-order condition. Pick points $x_1 < x_2$, and let $\bar{x} = \lambda x_1 + (1 - \lambda) x_2$ be some point in the convex hull ($\lambda \in (0, 1)$). Drawing the picture of the first-order condition from $\bar{x}$, and how both x_1 and x_2 are underestimated, should immediately convey the intuition of the equivalence. To see it algebraically: the two underestimates are:

$$\begin{aligned} f(x_1) &\geq f(\bar{x}) + f'(\bar{x}) \cdot (\bar{x} - x_1), \\ f(x_2) &\geq f(\bar{x}) + f'(\bar{x}) \cdot (\bar{x} - x_2). \end{aligned}$$

Multiplying the first inequality by λ, the second by $(1 - \lambda)$ and adding, we recover the derivative-free condition.

For the multivariate case, we use the fact that if a function is convex along all line segments, then it is convex, in which case this reduces to the above proof. $\square$

We now give the final definition of convexity.

Definition 14. *Suppose that a function $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable. Then f is convex iff its Hessian is positive semidefinite:*

$$\bigtriangledown^2 f(x) \succeq 0, \ \forall x \in dom(f) \tag{1.14}$$

As an example, consider the function:

$$f(x) = \|Ax - b\|_2^2.$$

Expanding, we have $f(x) = x^\top A^\top Ax - b^\top Ax - x^\top A^\top b + \|b\|_2^2$. The Hessian of this is $A^\top A$, which is positive semidefinite, since for any vector x, $x^\top (A^\top A) x = \|Ax\|_2^2 \geq 0$.

Proposition 2. *The definition give above is equivalent to the definition 12 and 13. That is, a twice differentiable function f is convex if*

$$\nabla^2 f(x) \in \mathcal{S}_+^n. \tag{1.15}$$

Proof: One way to prove this is via Taylor's theorem, using the Lagrange form of the remainder. To prove that a function with positive semidefinite Hessian is convex, using a second order Taylor expansion we have:

$$f(y) = f(x) + \nabla f(x)^T (y-x) + (y-x)^T \nabla^2 f(x + \alpha (y-x)) (y-x) \tag{1.16}$$

for some value of $\alpha \in [0,1]$. Now, since the Hessian is positive semidefinite,

$$(y-x)^T \nabla^2 f(x + \alpha (y-x)) (y-x) \geq 0, \tag{1.17}$$

which leads to

$$f(y) \geq f(x) + \nabla f(x)^T (y-x), \tag{1.18}$$

which is the first-order condition of convexity. Hence, this proves that $f(x)$ is convex. For the reverse direction, showing that convexity implies positive semi definiteness of the Hessian, again we can use Taylor's theorem. However, there is a slightly delicate issue because of the Hessian. For this, we need to use some properties of symmetric matrices, to claim that if for some orthonormal basis $\{x_i\}$, a matrix A satisfies $x^\top A x \geq 0$, then A is positive semidefinite. This completes the proof. □

1.3.1 Examples of Convex Functions

- ***Exponential*** $f(x) = e^{ax}, \forall a \in \mathbb{R}$ is convex on $\mathbb{R}$. To show e^{ax} is convex $\forall a \in \mathbb{R}$, we could simply see the second derivative of function f, which is $a^2 e^{ax} \geq 0, \forall a \in \mathbb{R}$

- ***Powers*** $f(x) = x^a$ is convex on $\mathbb{R}_{++}$ when $a \geq 1$ *or* $a \leq 0$, concave otherwise. Since $f''(x) = a(a-1)x^{a-2}, x \in \mathbb{R}_{++}$ is non-negative when $a \geq 1$ *or* $a \leq 0$.

- ***Negative Logarithm*** $f(x) = -logx$ is convex on its domain $\mathbb{R}_{++}$, because $f''(x) = \frac{1}{x^2} > 0, \forall x \in \mathbb{R}_{++}$

- ***Norms*** Every norms on $\mathbb{R}^n$ is convex. Using triangle inequality and positive homogenity, $f(\lambda x + (1-\lambda)y) \leq f(\lambda x) + f((1-\lambda)y) = \lambda f(x) + (1-\lambda) f(y), \forall x, y \in \mathbb{R}^n, 0 \leq \lambda \leq 1$.

- ***Max Function*** $f(x) = max\{x_1, x_2, \cdots, x_n\}$ is convex on $\mathbb{R}^n$. To show this, $f(\lambda x + (1-\lambda y)) = max\{\lambda x_1 + (1-\lambda)y_1, \cdots, \lambda x_n + (1-\lambda)y_n\} = \lambda x_i + (1-\lambda)y_i \leq \lambda x_{max} + (1-\lambda)y_{max} = \lambda f(x) + (1-\lambda) f(y), \forall x, y \in \mathbb{R}^n$.

1.4 Convex Optimization

Now let us revisit the basic convex optimization problem we saw at the beginning of the lecture:

$$\begin{aligned} \min: \quad & f_0(x) \\ \text{s.t.}: \quad & f_i(x) \leq 0, \qquad i = 1, \ldots, m. \end{aligned}$$

If all functions are convex, then the feasible set is convex. The fundamental consequence of convexity is that local optimality implies global optimality. This notion will prove extremely useful. In terms of analysis, it implies that we can characterize the optimal solution of a convex optimization problem using only local conditions. In terms of algorithms, it means that doing things that are locally optimal will result in a globally optimal solution. See also the problem in Homework One.

We now give optimality conditions for unconstrained optimization. These are the familiar first-order conditions of optimality for convex optimization. While simple, the condition is extremely useful both algorithmically and for analysis.

1.4.1 Unconstrained optimization

The formulation of an unconstrained optimization problem is as follows

$$\min : f(x)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is smooth and convex. In these problems, the necessary and sufficient condition for the optimal solution x_o is

$$\nabla f(x) = 0 \text{ at } x = x_o. \tag{1.19}$$

The intuition here is simple: if there are no local directions of descent, then a point must be locally optimal, which in turn (thanks to convexity) implies that it is globally optimal. We can see more precisely that the condition $\nabla f(x_0) = 0$ implies global optimality by using the first order conditions for convexity.

Proof: Recall that if f is convex, we have

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \;\; \forall y \in dom(f).$$

At $x = x_o$, $\nabla f(x_o) = 0$ and hence this reduces to

$$f(y) \geq f(x_o) \; \forall y \in dom(f),$$

which indeed is the statement that x_o is a global optimum. □

Note of course that this need not be unique. For uniqueness, we need a stronger version of convexity. This will be addressed in the next lecture.

1.4.2 Constrained optimization

Next we consider a general constrained convex optimization problem.

$$\begin{gathered}\min f(x)\\ \text{s.t } x \in \mathcal{X},\end{gathered}$$

where $f : \mathcal{X} \to \mathbb{R}$ is convex and smooth. The same intuition used above still applies here: a point x_o is globally optimal if it is locally optimal. Locally optimal means that there are no directions of descent that are feasible, i.e., that stay inside the feasible set $\mathcal{X}$. We make this precise in the next lecture.