Indian Institute of Technology Kanpur
CS773 Online Learning and Optimization

Scribe: Sharbatanu Chatterjee
Instructor: Purushottam Kar
Date: January 5, 2016

LECTURE
2

Preliminaries: Convex Analysis

1 Introduction

This lecture discusses a few mathematical preliminaries necessary to study convex analysis. Definitions and properties (with proofs wherever necessary) have been presented below.

2 Vector spaces

We begin with defining the ubiquitous concept of a vector space and several properties associated with them, as well as mathematical operations defined on them. The concept of a vector space is, in general, defined over an arbitrary field of *scalars*. However, we shall restrict ourselves to real vector spaces, i.e. those defined over the field of reals.

Definition 2.1 (Vector Space). A real vector space V is a set of mathematical objects which are called *vectors*, having certain properties, as mentioned below.

2.1 Properties of vector spaces

Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors in a vector space V. Then the following properties hold:

For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ then

1. *Closure under Addition*: $\mathbf{u} + \mathbf{v} \in V$
2. *Commutativity*: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
3. *Associativity*: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
4. *Identity*: $\exists\ \mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ [1]
5. *Inverse*: For every $\mathbf{v} \in V$, $\exists\ \mathbf{v}' \in V$ such that $\mathbf{v} + \mathbf{v}' = \mathbf{0}$ [2]
6. *Distributivity over real operations*: This property allows vectors to interact with real scalars (in general, elements of the underlying field). For all $a, b \in \mathbb{R}$, we have

$$a \cdot \mathbf{v} \in V \text{ (however } 0 \cdot \mathbf{v} = \mathbf{0})$$
$$(a + b) \cdot \mathbf{v} = a \cdot \mathbf{v} + b \cdot \mathbf{v}$$
$$a \cdot (\mathbf{u} + \mathbf{v}) = a \cdot \mathbf{u} + a \cdot \mathbf{v}$$

[1] $\mathbf{0}$ is called the *additive identity* of the vector space
[2] $\mathbf{v}'$ is called the *additive inverse* of the vector $\mathbf{v}$

Example 2.1. Some examples of vector spaces include:

1. $\mathbb{R}^n$, which is the set of all n-dimensional vectors with real components.
2. $B(\mathcal{X})$, which is the set of bounded functions over the domain $\mathcal{X}$.
3. $\mathcal{C}^1(\mathcal{X})$, which is the set of once differentiable functions over the domain $\mathcal{X}$.
4. $\mathcal{C}^\infty(\mathcal{X})$ which is the set of functions over the domain $\mathcal{X}$ which are infinitely differentiable (smooth functions).

3 Inner Product

The previous section introduced us to the concept of vector spaces. It behooves us to now define the concept of an inner product as a natural extension to describe the interaction of two vectors in a vector space as defined in Definition 2.1.

Definition 2.2 (Inner Product)**.** An inner product is a real valued bivariate function on vector spaces (mapping a pair of vectors to the set of reals) having the following properties.

3.1 Properties of Inner Products

An inner product, denoted as $\langle \cdot, \cdot \rangle := (\mathbf{u}, \mathbf{v}) \mapsto \mathbb{R}$, has the following properties:

1. *Distributivity*: $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
2. *Positive Definiteness*: $\forall \mathbf{u} \neq \mathbf{0}, \langle \mathbf{u}, \mathbf{u} \rangle > 0$
3. *Commutativity*: $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
4. *Linearity*: For a scalar a, $\langle a \cdot \mathbf{u}, \mathbf{v} \rangle = a \langle \mathbf{u}, \mathbf{v} \rangle$

Example 2.2. Some examples of inner products include:

1. $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^d x_i y_i$ for vectors in the d-dimensional space.[3]
2. *(Weighted inner products)* For any vector $\mathbf{w} > \mathbf{0}$[4], we define $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{w}} = \sum_{i=1}^d w_i x_i y_i$
3. *(Inner product induced by a matrix)* For any real positive definite (PD) matrix[5] A, $(A \succ 0)$, we define $\langle \mathbf{x}, \mathbf{y} \rangle_A := \mathbf{x}^\top A \mathbf{y}$

4 Norm

Inner products gave us an idea about the interaction of two vectors. This section introduces us to the concept of *norm* which corresponds to abstract notions of *lengths* of vectors in a vector space.

Definition 2.3 (Norm)**.** A norm is a real valued univariate function on vector spaces (mapping from a vector to the set of real numbers) having the following properties.

[3] This is the notion of *dot product* we generally have for vectors, also denoted frequently as $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y}$

[4] For 2 d-dimensional vectors $\mathbf{u}, \mathbf{v}$, the notation $\mathbf{u} > \mathbf{v}$ implies a coordinate-wise relation i.e. for all $i \in [d]$, $u_i > v_i$. We define the relation $\mathbf{u} \geq \mathbf{v}$ similarly.

[5] A PD matrix is a square symmetric matrix with all eigenvalues > 0

4.1 Properties of Norms

A norm, defined as $\|.\| := (\mathbf{v}) \mapsto \mathbb{R}$, has the following properties:

1. *Positivity*: $\forall \mathbf{u} \neq \mathbf{0}, \|\mathbf{u}\| > 0$.[6]
2. *Identity of the indiscernables* $\|\mathbf{0}\| = 0$ [7]
3. *Positive Homogeneity*: For any scalar $a \in \mathbb{R}, \|a \cdot \mathbf{v}\| = |a| \cdot \|\mathbf{v}\|$
4. *Traingle Inequality/Subadditivity*: $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.[8]

4.2 Induced Norms

Corresponding to every inner product defined over a vector space, we can define a norm as follows:

$$\|\mathbf{u}\| := \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

Such a norm is said to be *induced* by the corresponding inner product.

Theorem 2.1. An induced norm is also a norm.

Proof. Left as an exercise. □

Example 2.3. Some examples of commonly used norms are:

1. ℓ_p norms, where $1 \leq p < \infty$ defined as $\|\mathbf{x}\|_p := \left(\sum_{i=1}^{d} |x_i|^p\right)^{1/p}$. The ℓ_2 norm is often referred to as the *Euclidean norm*. The ℓ_1 norm is often referred to as the *Manhattan norm*.
2. l_∞ norm, defined as $\|\mathbf{x}\|_\infty := \max_i |x_i|$
3. Norms induced by a matrix, defined as $\|\mathbf{x}\|_A = \sqrt{\mathbf{x}^\top A \mathbf{x}}$ [9]

4.3 Dual Norms

Before discussing the concept of a dual norm, we first define the notion of the dual of a normed space.

Definition 2.4 (Normed Space)**.** A normed space, $(V, \|.\|)$, is a vector space over which a norm has been established.

Definition 2.5 (Linear Functional)**.** A real valued univariate function over a normed space $f : V \to \mathbb{R}$ is said to be a *linear functional* if it satisfies the following properties:

- $f(\mathbf{u}) + f(\mathbf{v}) = f(\mathbf{u} + \mathbf{v}), \forall \mathbf{u}, \mathbf{v} \in V$.
- $f(a \cdot \mathbf{u}) = a \cdot f(\mathbf{u}), \forall \mathbf{u} \in V$ and $a \in \mathbb{R}$

Definition 2.6 (Dual Space)**.** The vector space of all linear functionals over a vector space V, which we shall denote as $\mathcal{F}_{\text{lin}}(V)$, is referred to as the *dual* space of V.

[6] Norms for which this property does not hold i.e. $\|\mathbf{u}\| = 0$ for some $\mathbf{u} \neq \mathbf{0}$ are called *degenerate norms*. By definition, norms are non-degenerate.

[7] The concept behind this property is that if two vectors are an additive identity away from each other, the "distance" between them should be 0. If $\mathbf{u}, \mathbf{v}$ are two vectors, and $\mathbf{u} = \mathbf{v}$, then $\|\mathbf{u} - \mathbf{v}\| = 0$.

[8] The intuition behind this is that a detour should increase the "distance" travelled.

[9] Such a norm is non-degenerate iff A is a positive definite matrix, i.e. all its eigenvalues are strictly positive.

It is an easy exercise to show that the set of all linear functionals indeed forms a vectors space. For sake of brevity, we will often refer to the dual space by the notation $\mathcal{F}_{\text{lin}}$ alone. We now define the dual norm of a vector space as follows:

Definition 2.7 (Dual Norm)**.** Given a linear functional space $\mathcal{F}_{\text{lin}}$, associated with a normed vector space $(V, \|\cdot\|)$, the dual norm defined on $\mathcal{F}_{\text{lin}}$ is given by

$$\|f\|_* := \sup \{f(x), \|x\| \leq 1\}$$

Example 2.4. Examples of dual norms are:

1. $\|.\|_2$ is the dual norm of itself.
2. $\|.\|_1$ is the dual norm of $\|.\|_\infty$
3. The dual norm of $\|.\|_p$ is $\|.\|_q$ such that $1/p + 1/q = 1$

5 Cauchy-Schwartz Inequality

The Cauchy-Schwartz inequality gives a relation between the inner product of two vectors and the product of their ℓ_2 or Euclidean norms.

Theorem 2.2. *(Cauchy-Schwartz Inequality)* $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|_2 \|\mathbf{v}\|_2$

Proof. We will consider two separate cases for the proof.

1. *Case 1* ($\|\mathbf{u}\| = 0$ or $\|\mathbf{v}\| = 0$) Since norms are non-degenerate, this implies that either $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$. Thus, both the left and the right hand sides of the inequality are zero and the inequality holds.

2. *Case 2* ($\mathbf{u} \neq 0, \mathbf{v} \neq 0$) In this case we can assume, w.l.o.g. [10], that $\|\mathbf{u}\| = 1, \|\mathbf{v}\| = 1$. Now we only need to prove that
$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq 1$$
Consider the vector $\mathbf{u}_\perp = \mathbf{u} - \langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v}$. It is easy to see that $\langle \mathbf{u}_\perp, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{v} \rangle \langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{v} \rangle \|\mathbf{v}\|^2 = \langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{v} \rangle = 0$. Thus, $\mathbf{u}_\perp$ is the component of $\mathbf{u}$ orthogonal to $\mathbf{v}$.

   By the positivity property of norms and also by Theorem 2.1, we have $\|\mathbf{u}_\perp\|_2^2 \geq 0$ which gives us
   $\implies \langle \mathbf{u} - \langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v}, \mathbf{u} - \langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v} \rangle \geq 0$
   $\implies \|\mathbf{u}\|^2 + \|\langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v}\|^2 - 2 \langle \mathbf{u}, \langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v} \rangle \geq 0$
   $\implies 1 + \langle \mathbf{u}, \mathbf{v} \rangle^2 - 2 \langle \mathbf{u}, \mathbf{v} \rangle^2 \geq 0$
   $\implies \langle \mathbf{u}, \mathbf{v} \rangle^2 \leq 1$
   $\implies |\langle \mathbf{u}, \mathbf{v} \rangle| \leq 1$,

which concludes the proof. We note that the Cauchy-Schwartz inequality also holds for degenerate norms (also known as *semi-norms*). We establish this in Appendix A. $\square$

[10] Without loss of generality

6 Some miscellaneous definitions

6.1 Hyperplanes and Halfspaces

The next mathematical object we shall deal with are *hyperplanes* defined as follows.

Definition 2.8 (Hyperplane). For every $f \in \mathcal{F}_{\text{lin}}$ (i.e. a linear functional), we define a corresponding hyperplane, $\mathcal{H}_f$ as

$$\mathcal{H}_f := \{\mathbf{v} : f(\mathbf{v}) = 0\}$$

As we can see, in vectors over $\mathbb{R}^2$, all straight lines are hyperplanes. In vectors over the Euclidean space $\mathbb{R}^3$, all planes are hyperplanes, given by $\mathcal{H}_{\mathbf{a}} = \left\{\mathbf{v} : \mathbf{a}^\top \mathbf{v} = 0\right\}$

Corresponding to every hyperplane, there exists two *halfspaces* defined as the set of vectors on either "side" of the hyperplane.

Definition 2.9 (Halfspace). For every $f \in \mathcal{F}_{\text{lin}}$, we define the halfspace $\mathcal{E}_f$ as

$$\mathcal{E}_f = \{\mathbf{v} : f(\mathbf{v}) \geq 0\}.$$

Halfspaces may be defined using either a strict ($>$) or a non-strict ($\geq$) inequality. For example, corresponding to linear form $\mathbf{v} \mapsto \mathbf{a}^\top \mathbf{v}$, we have the halfspace $\mathcal{E}_{\mathbf{a}} = \left\{\mathbf{v} : \mathbf{a}^\top \mathbf{v} \geq 0\right\}$.

6.2 Balls and Ellipsoids

Definition 2.10 (Ball). A ball, corresponding to a norm, is defined as $\mathcal{B}_{\|\|}(\mathbf{v}, r) := \{\mathbf{x} : \|\mathbf{x} - \mathbf{v}\| \leq r\}$. The vector $\mathbf{v}$ is called the *center* of the ball and the scalar r is called the radius of the ball. The notation $\mathcal{B}_{\|\|}(r)$ is often used to refer to $\mathcal{B}_{\|\|}(\mathbf{0}, r)$.

Example 2.5. $\mathcal{B}_2(\mathbf{0}, r) = \{\mathbf{x} : \|\mathbf{x}\|_2 \leq r\}$

Definition 2.11 (Ellipsoid). An ellipsoid, similar to a ball, is defined as $\mathcal{E}_A(\mathbf{v}, r) := \{\mathbf{x} : \|\mathbf{x} - \mathbf{v}\|_A \leq r\}$. An ellipsoid is induced by a PD matrix A.

Remark 2.1. All ellipsoids are balls as they are balls corresponding to norms induced by PD matrices, but not all balls are ellipsoids.

7 Linear Combinations of Vectors

This section deals with several types of combinations of vectors. Note that combinations are important in our analysis, as we shall often define 'sets' generated by combinations of vectors. Given two vectors $\mathbf{u}, \mathbf{v} \in V$, a *linear combination* of these vectors is a vector of the form $\lambda \cdot \mathbf{u} + \mu \cdot \mathbf{v}$, where the combination coefficients λ and μ are scalars i.e reals.

In the following we will look at special types of linear combinations.

1. *Convex*: A convex combination of the vectors $\mathbf{u}$ and $\mathbf{v}$ is a linear combination where the coefficients are non-negative and add up to one i.e. of the form $\lambda \cdot \mathbf{u} + (1 - \lambda) \cdot \mathbf{v}$, for any $\lambda \in [0, 1]$.

2. *Conic*: A conic combination of the vectors $\mathbf{u}$ and $\mathbf{v}$ is a linear combination where the coefficients are non-negative i.e of the form $\lambda \cdot \mathbf{u} + \mu \cdot \mathbf{v}$, for any $\lambda \geq 0, \mu \geq 0$

3. *Affine*: An affine combination of the vectors $\mathbf{u}$ and $\mathbf{v}$ is a linear combination here the coefficients add up to one i.e. of the form $\lambda \cdot \mathbf{u} + \mu \cdot \mathbf{v}$, where λ, μ are such that $\lambda + \mu = 1$

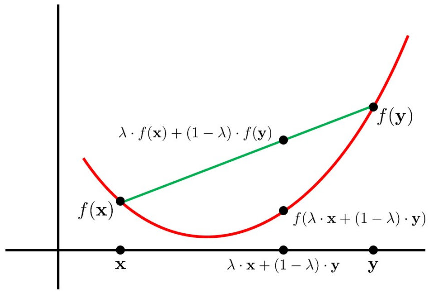

Figure 1: A convex function on the real line.

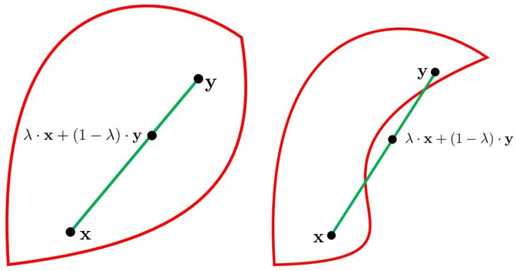

Figure 2: A convex and a non-convex set.

Definition 2.12 (Convex Set)**.** A set that is closed under all possible convex combinations is called a convex set. That is to say a set $\mathcal{C} \subset V$ is called convex if for all $\mathbf{u}, \mathbf{v} \in \mathcal{C}$ and $\lambda \in [0, 1]$, we have $\lambda \cdot \mathbf{u} + (1 - \lambda) \cdot \mathbf{v} \in \mathcal{C}$.

An illustrative example of a convex and a non-convex set is given in Figure 2. Similarly, sets closed under conic and affine combinations are, respectively, called conic and affine sets. Examples of convex sets include polytopes in Euclidean spaces, intervals over the real line, balls induced by norms. Examples of conic sets include the set of all PSD matrices. Examples of affine sets include those in Euclidean spaces

$$\mathcal{H}_{(f,b)} := \{\mathbf{v} : f(\mathbf{v}) = b\}$$

8 Convex functions

So we finally arrive at the juncture of defining convex functions.

Definition 2.13 (Convex Function)**.** A function $f : V \to \mathbb{R}$, is called convex if $\forall\ \mathbf{u}, \mathbf{v} \in V, \lambda \in [0, 1]$,

$$f(\lambda \cdot \mathbf{u} + (1 - \lambda) \cdot \mathbf{v}) \leq \lambda f(\mathbf{u}) + (1 - \lambda) f(\mathbf{v}).$$

See Figure 1 for an illustration.

This fundamental definition is however cumbersome to work with. A weaker but more workable definition is that of *mid-point convexity*:

Definition 2.14 (Midpoint convexity)**.** A function is called mid-point convex if $\forall\ \mathbf{u}, \mathbf{v} \in V$,

$$f\left(\frac{\mathbf{u} + \mathbf{v}}{2}\right) \leq \frac{f(\mathbf{u}) + f(\mathbf{v})}{2}.$$

Theorem 2.3. A continuous function is convex iff it is mid-point convex. (For a proof, originally by Jensen in 1905, refer to Sra et al. (2014))

Example 2.6. Examples of convex functions are *norms*. The summation and supremum operations preserve convexity i.e. the sum and supremum of convex functions are convex as well.

8.1 Derivatives of functions

Definition 2.15 (Fréchet derivatives). A real-valued function $f : V \to \mathbb{R}$ defined over a normed space is said to possess a Fréchet derivative $g \in \mathcal{F}_{\text{lin}}$ at a point $\mathbf{x} \in V$ if the following limiting behavior exists

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - g(\mathbf{h})|}{\|h\|} = 0$$

The derivative g is often referred to as the *gradient* of f.

Lemma 2.4. A differentiable function is convex iff

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle ,$$

where ∇f is the gradient of the function f.

Proof. (Adapted from Boyd and Vandenberghe (2004)). Since f is continuous, we will use the midpoint convexity definition to prove the result.

(If) Consider $\mathbf{z} = \frac{\mathbf{x}+\mathbf{y}}{2}$. The gradient condition gives us

$$\begin{aligned} f(\mathbf{x}) &\geq f(\mathbf{z}) + \langle \nabla f(\mathbf{z}), \mathbf{x} - \mathbf{z} \rangle \\ f(\mathbf{y}) &\geq f(\mathbf{z}) + \langle \nabla f(\mathbf{z}), \mathbf{y} - \mathbf{z} \rangle \end{aligned}$$

Adding the equations and dividing by two gives us the required result

$$\frac{f(\mathbf{x}) + f(\mathbf{y})}{2} \geq f(\mathbf{z}) = f\left(\frac{\mathbf{x}+\mathbf{y}}{2}\right)$$

(Only if) For sake of simplicity, assume that f is defined and differentiable over the entire vector space. Then for $\lambda \in (0, 1]$, we have, by the convexity property

$$f(\lambda \cdot \mathbf{x} + (1-\lambda) \cdot \mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

which, upon rearranging, gives us

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \frac{f(\mathbf{y} + \lambda \cdot (\mathbf{x} - \mathbf{y})) - f(\mathbf{y})}{\lambda}.$$

Now, we have

$$\begin{aligned} \frac{f(\mathbf{y} + \lambda \cdot (\mathbf{x} - \mathbf{y})) - f(\mathbf{y})}{\lambda} &= \frac{f(\mathbf{y} + \lambda \cdot (\mathbf{x} - \mathbf{y})) - f(\mathbf{y})}{\lambda} \\ &= \underbrace{\frac{f(\mathbf{y} + \lambda \cdot (\mathbf{x} - \mathbf{y})) - f(\mathbf{y}) - \langle \nabla f(\mathbf{y}), \lambda \cdot (\mathbf{x} - \mathbf{y}) \rangle}{\lambda}}_{(A)} + \langle \nabla f(\mathbf{y}), (\mathbf{x} - \mathbf{y}) \rangle \end{aligned}$$

Now we have

$$\begin{aligned} (A) &= \frac{f(\mathbf{y} + \lambda \cdot (\mathbf{x} - \mathbf{y})) - f(\mathbf{y}) - \langle \nabla f(\mathbf{y}), \lambda \cdot (\mathbf{x} - \mathbf{y}) \rangle}{\lambda} \\ &= \frac{f(\mathbf{y} + \lambda \cdot (\mathbf{x} - \mathbf{y})) - f(\mathbf{y}) - \langle \nabla f(\mathbf{y}), \lambda \cdot (\mathbf{x} - \mathbf{y}) \rangle}{\lambda \cdot \|\mathbf{x} - \mathbf{y}\|} \cdot \|\mathbf{x} - \mathbf{y}\| \end{aligned}$$

Taking $\lambda \to 0$ and using Definition 2.15 gives us $(A) \to 0$ which proves the result. $\square$

References

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Suvrit Sra, Avinava Dubey, and Ahmed Hefny. Scribed notes for Lecture 2, 10-801: Advanced Optimization and Convex functions, 2014. URL `http://www.cs.cmu.edu/~suvrit/teach/10801_lecture2.pdf`.

A The Cauchy-Schwartz Inequality for Semi-norms

To establish this, we need only analyze the case when $\|\mathbf{u}\| = 0$ which the following claim does.

Claim 2.5. If $\|\mathbf{u}\| = 0$ then $\langle \mathbf{v}, \mathbf{u} \rangle = 0$ for all $\mathbf{v} \in V$.

Proof. Consider any vector $\mathbf{v}$ and look at the vector $\mathbf{z} = \mathbf{v} - r \cdot \mathbf{u}$, where $r = t \cdot \langle \mathbf{u}, \mathbf{v} \rangle$ for some real value t (note that in the proof of the Cauchy-Schwartz inequality for norms, we had taken $r = \langle \mathbf{u}, \mathbf{v} \rangle$ i.e. $t = 1$). This gives us

$$\begin{aligned}\|\mathbf{z}\|_2^2 &= \|\mathbf{v}\|_2^2 + r^2 \|\mathbf{u}\|_2^2 - 2r \cdot \langle \mathbf{u}, \mathbf{v} \rangle \\ &= \|\mathbf{v}\|_2^2 - 2r \cdot \langle \mathbf{u}, \mathbf{v} \rangle \\ &= \|\mathbf{v}\|_2^2 - 2t \cdot \langle \mathbf{u}, \mathbf{v} \rangle^2 .\end{aligned}$$

Since, $\|\mathbf{z}\|_2 \geq 0$, this means $\langle \mathbf{u}, \mathbf{v} \rangle^2 \leq 1/(2t) \|\mathbf{v}\|_2^2$. Since this holds for all values of t, taking $t \to \infty$ tells us that $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. This proves the Cauchy-Schwartz inequality for semi-inner products as well. $\square$