Real Vector Spaces, the Cauchy-Schwarz Inequality, & Convex Functions in ACL2(r)

Carl Kwan
Mark R. Greenstreet
University of British Columbia

15th International Workshop on the ACL2 Theorem Prover and Its Applications
Introduction

Outline:

- Framework for reasoning about real vector spaces and convex functions
  - The Cauchy-Schwarz inequality
- Proof “engineering”
  - Design proofs such that theorem statements are clear and concise
  - Avoid fundamental logical limitations

Motivation:

- Reasoning about convex optimisation algorithms
- Cauchy-Schwarz is useful and elegant
  - Top 100 Theorems / Formalising 100 Theorems

\[^{1}\text{cs.ru.nl/\sim freek/100}\]
Vector Spaces

$\left(\mathbb{R}^n, \mathbb{R}, \cdot, +\right)$ such that

- $+: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is associative and commutative
- Identity elements: $0 + v = v$ and $1v = v$
- Inverse elements: $v + (-v) = 0$
- Compatibility: $a(bv) = (ab)v$
- Distributivity (two ways): $a(u + v) = au + uv$ and $(a + b)v = av + bv$
Inner Product Spaces

Inner Product Space = Vector Space + Inner Product

\[ \langle -, - \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \]

- Positive-definiteness: \( \langle u, u \rangle \geq 0 \) and \( \langle u, u \rangle = 0 \iff u = 0 \)
- Symmetry\(^2\): \( \langle u, v \rangle = \langle v, u \rangle \)
- Linearity of the first coordinate:
  \[
  \langle au + v, w \rangle = a \langle u, w \rangle + \langle v, w \rangle
  \]

For \( \mathbb{R}^n \) and \( u = (u_i)_{i=1}^n, \ v = (v_i)_{i=1}^n \), use the dot product:

\[
\langle u, v \rangle = \sum_{i=1}^{n} u_i v_i
\]

\(^2\) when over \( \mathbb{R} \)
Theorem 1 (The Cauchy-Schwarz Inequality)
Let \( u, v \in \mathbb{R}^n \). Then

\[
|\langle u, v \rangle|^2 \leq \langle u, u \rangle \langle v, v \rangle \tag{CS1}
\]

or, equivalently,

\[
|\langle u, v \rangle| \leq \|u\| \cdot \|v\| \tag{CS2}
\]

with equality iff \( u, v \) are linearly dependent. Here \( \|u\| := \sqrt{\langle u, u \rangle} \).

How to prove it? Clever set-up + basic algebraic manipulations
Proof of $|\langle u, v \rangle|^2 \leq \langle u, u \rangle \langle v, v \rangle$

How to prove it? Clever set-up + basic algebraic manipulations:

**Proof (sketch).**

From positive-definiteness:

$$0 \leq \langle u - av, u - av \rangle = \langle u, u \rangle - 2a\langle u, v \rangle + a^2 \langle v, v \rangle.$$ 

Set $a = \frac{\langle u, v \rangle}{\langle v, v \rangle}$ and rearrange (a bunch) to get

$$0 \leq \cdots = \langle u, u \rangle + \langle u, v \rangle \left( -2 \frac{\langle u, v \rangle}{\langle v, v \rangle} + \frac{\langle u, v \rangle}{\langle v, v \rangle} \right) = \langle u, u \rangle - \frac{\langle u, v \rangle^2}{\langle v, v \rangle}.$$ 

How to formalise it? Follow (mostly) from the classical proofs.
Structure of Cauchy-Schwarz

Approach:

CS1: $|\langle u, v \rangle|^2 \leq \langle u, u \rangle \langle v, v \rangle$

CS2: $|\langle u, v \rangle| \leq \|u\| \|v\|$

EQ: $|\langle u, v \rangle|^2 = \langle u, u \rangle \langle v, v \rangle \iff \exists a \in \mathbb{R}, u = av$
Cauchy-Schwarz: Conditions for Equality

\[ \exists a \in \mathbb{R}, \quad u = av \]

\[ |\langle u, v \rangle|^2 = \langle u, u \rangle \langle v, v \rangle \]

In ACL2, just reverse and use positive-definiteness

\[ 0 \leq \langle u - av, u - av \rangle = \cdots = \langle u, u \rangle - \frac{\langle u, v \rangle^2}{\langle v, v \rangle}. \]

How to express “\( \exists a \)”?

1. Explicitly compute \( a \) from \( |\langle u, v \rangle|^2 = \langle u, u \rangle \langle v, v \rangle \)
   - hard & annoying

2. Use Skolem functions - much easier
Skolem functions have bodies with outermost quantifiers\textsuperscript{3}:

\texttt{(defun-sk linear-dependence (u v))}
\texttt{(exists a (equal u (scalar-* a v)))}

Requires a witness:

\[ 0 = \langle u - av, u - av \rangle \iff u - av = 0 \iff u = av \]

where \( a = \frac{\langle u,v \rangle}{\langle v,v \rangle} \) from before.

\textsuperscript{3}\texttt{scalar-*} is scalar-vector multiplication
Results:

- Reason about real vector & inner product spaces
- Formalised Cauchy-Schwarz inequality

Proof design issues:

- Exhibiting linear dependence in Cauchy-Schwarz
  - Use Skolem functions
  - Explicitly computing coefficients is hard
    - why compute when you don’t need to?
Metric Spaces

\[ \langle u - v, u - v \rangle = \| u - v \|^2 = d^2(u, v) \]

inner products $\rightarrow$ norms $\rightarrow$ metrics

\((M, d)\) where \(d : M \times M \to \mathbb{R}\) such that

1. Indiscernibility: \(d(x, y) = 0 \iff x = y\)
2. Symmetry: \(d(x, y) = d(y, x)\)
3. Triangle inequality: \(d(x, y) \leq d(x, z) + d(z, y)\)

Let \(M = \mathbb{R}^n\) and \(d(x, y) = \| x - y \|:\)

1. & 2. Immediate

3. Use Cauchy-Schwarz: let \(x = x' - z, y = z - y'\)

\[
\| x + y \|^2 = \| x \|^2 + 2 \langle x, y \rangle + \| y \|^2 \\
\leq \| x \|^2 + 2\| x \|\| y \| + \| y \|^2 = (\| x \| + \| y \|)^2
\]
A number $x$ is *standard* if it satisfies our usual definition of real. A number $x > 0$ is *$i$-small* if it is less than any positive standard.

**Continuity:** A function $f$ is *continuous* at a standard $x$ if for any $y$

$$d(x, y) \text{ i-small} \implies d(f(x), f(y)) \text{ i-small}$$

Univariate

$$f : \mathbb{R} \rightarrow \mathbb{R}, \quad d = | \cdot |$$

Multivariate

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad d = \| \cdot \|$$

**Differentiability:** The *derivative* of $f$ is a function $f'$ satisfying the conditions below for “i-small” $h$

Univariate

$$f'(x) = \frac{f(x+h) - f(x)}{h}$$

Multivariate

$$\|f(x+h) - f(x) - \langle f'(x), h \rangle\| \|h\| = 0$$

What does “i-small” mean for a vector in $\mathbb{R}^n$?  

---

*informal*
Recognizing “i-small” Vectors

Want:

(defun i-small-vecp (vec)
  (if (null vec) t (and (i-small (car vec))
                         (i-small-vecp (cdr vec))))))

NO! Non-classical\(^5\) recursive functions are prohibited! Instead,

\[ \|x\| = \sqrt{\sum_{i=1}^{n} z_i^2} \geq \max_i |x_i| \geq |x_i| \]

so

\[ \|x\| \text{ i-small} \implies |x_i| \text{ i-small } \forall i \in [1, n] \]

\(^5\text{functions defined only in ACL2(r)}\)
Recognizing “i-small” Vectors

$$\| x \| \text{i-small} \Rightarrow |x_i| \text{i-small} \ \forall i \in [1, n]$$

Avoid recursion by reasoning over $i$:

```lisp
(defun eu-norm-i-small-implies-elements-i-small
  (implies (and (real-listp x)
               (i-small (eu-norm x))
               (natp i)
               (< i (len x)))
           (i-small (nth i x))))
```

eu-norm is the Euclidean norm
Real Vector & Metric Spaces - Summary

Results:

▶ Reason about real vector spaces
▶ Reason about real metric spaces
  ▶ Multivariate continuity & differentiabilty

Proof design issues:

▶ Exhibiting linear dependence in Cauchy-Schwarz
▶ Defining continuity
  ▶ Non-classical recursive functions are prohibited
  ▶ Show the largest entry in the vector is i-small
  ▶ Reason about the index of arbitrary entries in the vector to avoid recursion
Convex Functions

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if for all $\alpha \in [0, 1] \subset \mathbb{R}$, $x, y \in \mathbb{R}^n$

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

Theorem 2
Let $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R} \rightarrow \mathbb{R}$ be convex. Then
1. $a \cdot f$ is convex for all $a \in \mathbb{R}_{\geq 0}$,
2. $f + g$ is convex,
3. $h \circ f$ is convex.

But how do we reason about functions?
Encapsulating Convex Functions

Encapsulate and suppress function definitions after proving hypotheses:

(encapsulate

... (local (defun cvfn-1 (x) ... 1337)) ... (defthm cvfn-1-convex (implies ... ;; hypotheses

(<= (cvfn-1 (vec+- (scalar-* a x)

(scalar-* (- 1 a) y)))

(+ (* a (cvfn-1 x))

(* (- 1 a) (cvfn-1 y)))))) ...) ... ;; prove theorems about cvfn-1)

(local (in-theory (disable cvfn-1)))

How do we reason about the convexity of a function?
Nesterov’s Theorem

Theorem 3 (Nesterov)

“All the conditions below, holding for all \( x, y \in \mathbb{R}^n \) and \( \alpha \) from \([0, 1]\), are equivalent to inclusion \( f \in \mathcal{F}_{L, 1}^{1, 1}(\mathbb{R}^n)\):”

\[
\begin{align*}
f(y) & \leq f(x) + \langle f'(x), y - x \rangle + \frac{L}{2} \|x - y\|^2 \quad (N1) \\
f(x) + \langle f'(x), y - x \rangle + \frac{1}{2L} \|f'(x) - f'(y)\|^2 & \leq f(y) \quad (N2) \\
\frac{1}{L} \|f'(x) - f'(y)\|^2 & \leq \langle f'(x) - f'(y), x - y \rangle \quad (N3) \\
\langle f'(x) - f'(y), x - y \rangle & \leq L \|x - y\|^2 \quad (N4) \\
f(\alpha x + (1 - \alpha) y) + \frac{\alpha(1 - \alpha)}{2L} \|f'(x) - f'(y)\|^2 & \leq \alpha f(x) + (1 - \alpha) f(y) \quad (N5) \\
\alpha f(x) + (1 - \alpha) f(y) & \leq f(\alpha x + (1 - \alpha) y) + \alpha(1 - \alpha) \frac{L}{2} \|x - y\|^2 \quad (N6)
\end{align*}
\]

---

\(^6\)Yuri Nesterov’s *Introductory Lectures on Convex Optimization*
Lipschitz Continuity

What is $\mathcal{F}^{1,1}_L(\mathbb{R}^n)$?

A function $f$ belongs to the class $\mathcal{F}^{p,q}_L(\mathbb{R}^n)$, $p \geq q$, if

- $f$ is $p$-times continuously differentiable on $\mathbb{R}^n$, i.e. in $C^p(\mathbb{R}^n)$
- $f$ is convex, i.e. in $\mathcal{F}(\mathbb{R}^n)$
- the $q$-th derivative of $f$ is $L$-Lipschitz continuous on $\mathbb{R}^n$, i.e. $f^{(q)} \in C_L(\mathbb{R}^n)$

A derivative (gradient) $f'$ of a function $f$ is \textit{L-Lipschitz continuous} if

$$\|f'(x) - f'(y)\| \leq L\|x - y\|$$
Ambiguities in Nesterov’s Theorem

“All the conditions below, holding for all \( x, y \in \mathbb{R}^n \) and \( \alpha \) from \([0, 1]\), are equivalent to inclusion \( f \in \mathcal{F}_{L}^{1,1}(\mathbb{R}^n)\): … [N1 - N6]”

What does Nesterov mean?

| \( \forall f : \mathbb{R}^n \rightarrow \mathbb{R}, \ f \in \mathcal{F}_{L}^{1,1} \) | \( \iff \) N1 | \( \iff \) \( \cdots \) | \( \iff \) N6 | \\
| \( \forall f \in C, \ f \in \mathcal{F}_{L}^{1,1} \) | \( \iff \) N1 | \( \iff \) \( \cdots \) | \( \iff \) N6 | False \\
| \( \forall f \in C^1, \ f \in \mathcal{F}_{L}^{1,1} \) | \( \iff \) N1 | \( \iff \) \( \cdots \) | \( \iff \) N6 | False \\
| \( \forall f \in C^{1,1}, \ f \in \mathcal{F}_{L}^{1,1} \) | \( \iff \) N1 | \( \iff \) \( \cdots \) | \( \iff \) N6 | False \\
| \( \forall f \in C_{L}^{1,1}, \ f \in \mathcal{F}_{L}^{1,1} \) | \( \iff \) N1 | \( \iff \) \( \cdots \) | \( \iff \) N6 | False \\
| \( \forall f \in \mathcal{F}^{1,1}, \ f \in \mathcal{F}_{L}^{1,1} \) | \( \iff \) N1 | \( \iff \) \( \cdots \) | \( \iff \) N6 | Almost True \\
| \( \forall f \in \mathcal{F}^{1,1}, \ f \in \mathcal{F}_{L}^{1,1} \) | \( \iff \) N1 | \( \iff \) \( \cdots \) | \( \iff \) N6 | True
Nesterov’s Theorem in ACL2(r)

Nesterov’s approach

Formalisation approach

CS: Cauchy-Schwarz

\[ \int \]

x2: instantiating inequalities twice

N0: Lipschitz Continuity

Instantiating Inequalities

Sometimes we need to add two “copies” of an inequality, eg. two copies of N2 with variables swapped give N3

\[ f(x) + \langle f'(x), y - x \rangle + \frac{1}{2L} \| f'(x) - f'(y) \|^2 \leq f(y), \]

\[ f(y) + \langle f'(y), x - y \rangle + \frac{1}{2L} \| f'(y) - f'(x) \|^2 \leq f(x), \]

\[ \implies \frac{1}{L} \| f'(x) - f'(y) \|^2 \leq \langle f'(x) - f'(y), x - y \rangle \]

Usually,

\[(\text{defthm ineq-N2-implies-ineq-N3} \]

\[(\text{implies (and (real-listp x) (real-listp y) ... (ineq-N2 x y)) (ineq-N3 x y))} \]

How do we instantiate N2 with swapped variables?
Instantiating Inequalities

Maybe:

\[(\text{implies } (\text{ineq-N2 } x \ y) \ (\text{ineq-N2 } y \ x))\]

But this is not (necessarily) true:

\[\forall x, y, (P(x, y) \implies P(y, x))\]

What Nesterov means is:

\[(\forall x, y, P(x, y)) \implies (\forall x, y, P(y, x)) \quad (\ast)\]

Maybe:

\[(\text{implies } (\text{and } \ldots \ (\text{ineq-N2 } x \ y) \ (\text{ineq-N2 } y \ x))\]

\[(\text{ineq-N3}))\]

But then N1 \implies N2 would need two copies of N1, too! Etc.

Stronger than (\ast) but messy.
Instantiating Inequalities

Use Skolem functions (again), eg.\(^7\)

\[
\text{(defun-sk ineq-N2-sk ...)
  (forall (x y) (ineq-N2 x y)))
\]

Instantiate as needed, eg.

\[
\text{(implies (ineq-N2-sk ...)
  (and (ineq-N2 x y) (ineq-N2 y x)))}
\]

\(^7\)slightly more complicated in reality
Nesterov’s Final Form

\[ N_0 \iff N_1 \iff \cdots \iff N_6 \]

means

\[ (N_0 \lor N_1 \lor \cdots \lor N_6) \implies (N_0 \land N_1 \land \cdots \land N_6) \]

If any one is true, we get the rest for free, eg.

\[
\text{(defthm nesterov}
\begin{align*}
(\text{implies} & \ (\text{or} \ (\text{ineq}-N_0 \ \ldots) \ (\text{ineq}-N_1 \ \ldots) \ \ldots) \\
& \ (\text{and} \ (\text{ineq}-N_0 \ \ldots) \ (\text{ineq}-N_1 \ \ldots) \ \ldots))
\end{align*}
\text{)}
\]
Conclusion

We saw

- A new framework for reasoning about real vector spaces and convex functions
  - A formal first-order proof of the Cauchy-Schwarz inequality
- Proof “engineering”: design proofs so that
  - theorem statements are clean and unambiguous
  - fundamental logical limitations are avoided

Future:

- Convex optimisation and machine learning algorithms
  - eg. Stochastic gradient descent, perceptron, etc.
- Multivariate analysis
- Generalisations of vector/metric spaces
  - eg. Abstract inner product spaces, Hilbert spaces, etc.
Conclusion

We saw

- A new framework for reasoning about real vector spaces and convex functions
  - A formal first-order proof of the Cauchy-Schwarz inequality
- Proof “engineering”: design proofs so that
  - theorem statements are clean and unambiguous
  - fundamental logical limitations are avoided

Future:

- Convex optimisation and machine learning algorithms
  - eg. Stochastic gradient descent, perceptron, etc.
- Multivariate analysis
- Generalisations of vector/metric spaces
  - eg. Abstract inner product spaces, Hilbert spaces, etc.

Thank You