CSE 840: Computational Foundations of Artificial Intelligence February 5th, 2025

Norm and Function Spaces

Instructor: Vishnu Boddeti *Scribe: Mohammad Eini, Cayden Koweck*

1 Equivalence of Norms

Theorem 1 *All norms on $\mathbb{R}^n$ are (topologically) equivalent: if $||.||_a$ and $||.||_b : \mathbb{R}^n \to \mathbb{R}$ are two norms defined on $\mathbb{R}^n$, then there exist two constants $\alpha, \beta > 0$ such that:*

$$\forall mathbfx \in \mathbb{R}^n : \alpha.||\mathbf{x}||_a \leq ||\mathbf{x}||_b \leq \beta.||\mathbf{x}||_a \tag{1}$$

Proof of Theorem 1: *Without loss of generality (W.L.O.G) we prove that if $||.||$ is any norm on $\mathbb{R}^n$, then it is equivalent to $||.||_\infty$ in $\mathbb{R}^n$. For this purpose we need to use two inequalities.*

Lemma 2 $\exists C_1 > 0,\ \forall x : ||x|| \leq C_1.||x||_\infty$
Let $\mathbf{x} = \sum_i x_i e_i$ be the representation of $\mathbf{x}$ in the standard basis of $\mathbb{R}^n$.

$$||\mathbf{x}|| = ||\sum_{i=1}^{n} x_i e_i|| \leq \sum_{i=1}^{n} ||x_i e_i|| \leq \sum_{i=1}^{n} ||x||_\infty ||e_i|| = ||x||_\infty \sum_{i=1}^{n} ||e_i|| = ||x||_\infty C_1$$

Lemma 3 $\exists C_2 > 0,\ \forall x : ||x||_\infty \leq C_2.||x||$
Let $S := \{\mathbf{x} \in \mathbb{R}^n \mid ||\mathbf{x}||_\infty = 1\}$ be the unit sphere w.r.t. $||.||_\infty$. Consider $f : S \to \mathbb{R}$, $\mathbf{x} \to ||x||$. The mapping f is continuous w.r.t. $||.||_\infty$. This follows from the fact that:

$$|f(\mathbf{x}) - f(\mathbf{y})| = |\ ||\mathbf{x}|| - ||\mathbf{y}||\ | \leq ||\mathbf{x} - \mathbf{y}|| \leq C_1.||\mathbf{x} - \mathbf{y}||_\infty$$

Which is an instance of ***Lipschitz Continuity*** *concept.The S is closed and bounded, so S is compact(from analysis). Any continuous mapping on a compact set takes its minimum and maximum. Define:*

$$\widetilde{C_2} := min\{f(\mathbf{x}) \mid \mathbf{x} \in S\}$$
$$\mathbf{x} \in S : ||\frac{\mathbf{x}}{1}|| = ||\frac{\mathbf{x}}{||\mathbf{x}||_\infty}|| = \frac{||\mathbf{x}||}{||\mathbf{x}||_\infty}$$
$$\Rightarrow \widetilde{C_2} \leq \frac{||\mathbf{x}||}{||\mathbf{x}||_\infty} \Rightarrow ||\mathbf{x}||_\infty \leq \frac{1}{\widetilde{C_2}} ||\mathbf{x}||$$

Now choose $C_2 = \frac{1}{\widetilde{C_2}}$ which proves the lemma.
It is also worth mentioning that in here, we used the ***Extreme Value theorem*** *which states that: Let $f : [a, b] \to \mathbb{R}$ be a continuous function on the closed interval $[a, b]$. Then f attains both a maximum and a minimum value on $[a, b]$, meaning there exist points $c, d \in [a, b]$ such that:*

$$f(c) \geq f(x) \quad \textit{for all } x \in [a, b]$$

$$f(d) \leq f(x) \quad \textit{for all } x \in [a, b]$$

That is, f has an absolute maximum at c and an absolute minimum at d.

□

2 Convex Sets Are Unit Balls of Norms

Definition 4 *Consider a real vector space* $\mathbf{V}$ *and* $S \subset \mathbf{V}$*.* S *is called convex if:*

$$\forall b : \ 0 \leq b \leq 1 \ and \ \forall \mathbf{x}, \mathbf{y} \in S : \ b.\mathbf{x} + (1-b).\mathbf{y} \in S$$

in the Figure 1, you can see a demonstration of this concept.

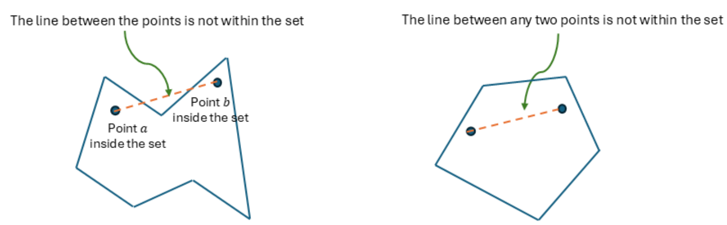


Figure 1: A demonstration of convex and concave sets

Definition 5 *A set* $C \subset \mathbf{V}$ *is considered symmetric if* $\mathbf{x} \in C \Rightarrow -\mathbf{x} \in C$*. You can see a demonstration of the symmetric concept in Figure 2*

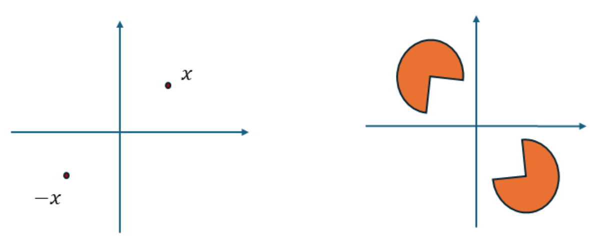


Figure 2: A demonstration of symmetric sets

Theorem 6 *(1) Let* $C \subset \mathbb{R}^d$ *to be closed, symmetric, convex and has none-empty interior. Define* $p(\mathbf{x}) := inf\{t > 0 \mid \mathbf{x} \in t.C\}$*. Then* p *is a semi-norm. If* C *is bounded, then* p *is a norm and its unitball coincides with* C*.(ie.* $C = \{\mathbf{x} \in \mathbb{R}^d \mid p(\mathbf{x}) \leq 1\}$*.An intuition of definition of the functionp*(x) *can be seen in Figure 3 (2) For any norm* $||.||$ *on* $\mathbb{R}^d$*, the set* $C := \{\mathbf{x} \in \mathbb{R}^d \mid ||x|| \leq 1\}$ *is bounded, symmetric, closed, convex and has none-empty interior.*

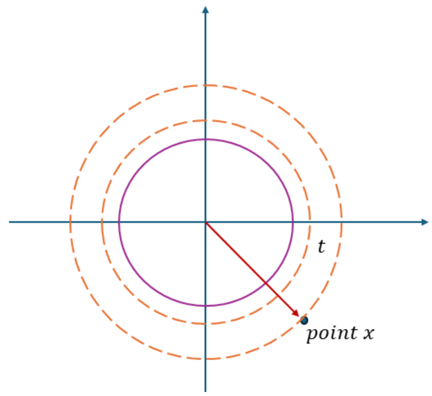


Figure 3: An intuition of $p(x)$

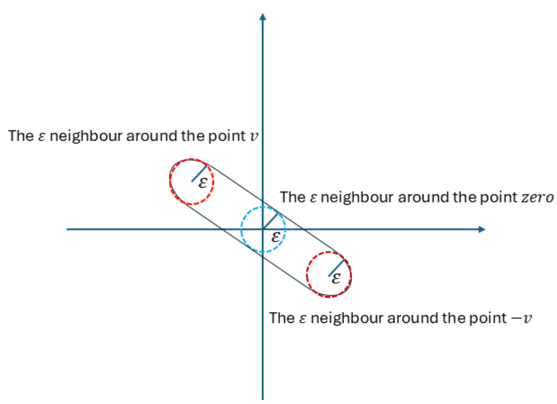


Figure 4: An intuition of the proof of existing ϵ ball around zero

Proof of Theorem 6: $p(\mathbf{x})$ is well defined.
Want to prove: $\mathbf{x} \in \mathbb{R}^d$ the set $\{t > 0 \mid \mathbf{x} \in t.C\} \neq \emptyset$. We are going to prove: $\exists \epsilon > 0$ such that $B_\epsilon(0) = \{e \in \mathbb{R}^d \mid ||e|| < \epsilon\} \subset C$
As shown in Figure 4, intuitively we are trying to push the ϵ neighborhood around the point v around the zero.

- By assumption, C has at least one interior point.

$$\mathbf{v} \in C^\circ \Rightarrow \mathbf{v} + B_\epsilon(0) = \{\mathbf{v} + e \mid e \in B_\epsilon(0)\}$$

- By symmetry, $\mathbf{v} + e \in C \Rightarrow -(\mathbf{v} + e) \in C$
- By convexity,$\frac{1}{2}(\mathbf{v} + e) + \frac{1}{2}(-\mathbf{v} + c) = e \in C$

So $B_\epsilon(0) \subset C$ and the set $\{t > 0 \mid \mathbf{x} \in t.C\}$ is not empty.The infimum of $inf\{t > 0 \mid \mathbf{x} \in t.C\}$ exist because $\{t > 0 \mid \mathbf{x} \in t.C\} \subset \mathbb{R}$ has 0 as its lower bound.

Definition 7 *Infimum: The largest lower-bound of a set.*

(P1) $p(0) = 0$

- have seen: $0 \in C$
- $\forall t > 0 \,:\, 0 \in 0.C$
- $inf\,\{t\,|\,0 \in t.C\} = 0$
- $\Rightarrow p(0) = 0$

(P2) $P(\alpha * x) = |\alpha| * P(x)$

· $\forall \alpha > 0$, we have: $P(x \cdot x)$
$= \inf\{t > 0 \mid x \cdot x \int \cdot c\}$
$= \inf\{\alpha \cdot s > 0 \mid x \in s \cdot c\}\,(s = t/d)$
$= \alpha \cdot P(x)$
$\implies P(\alpha x) = \alpha P(x)$

· By symmetry, we also get that: $P(-x) = P(x)$

· Combining the two statements gives us $P(x \cdot x) = |\alpha| \cdot P(x)$, which fulfills Homogeneity.

(P3) Triangle-Inequality

Consider $x, y \in \mathbb{R}^d, s, t > 0\, s.t\ \frac{x}{s} \in C, \frac{y}{t} \in C$
Observe: $s\frac{s}{+}t + t\frac{s}{+}t = 1$, So by convexity:
$\frac{s}{s+t} \cdot \frac{x}{s} + \frac{t}{s+t} \cdot \frac{y}{t} \in C \implies \frac{x+y}{s+t} \in C \implies \frac{x+y}{u_\circ}$
$\implies P(x + y) = inf\{u > 0 \mid x + y \in u\cdot\} \leq u_\circ \leq s + t = P(x) + P(y)$
We know s = P(x) because s was chosen s.t $x \in s \cdot C$
We know t = P(y) because t was chosen s.t $y \in t \cdot C$
Consider a sequence $(Si)_{i\in\mathbb{N}}$ s.t. $x \in s_i \cdot C$ and $s_i \to P(x)$
Similarly $(t_i)_{i\in\mathbb{N}}$ s.t. $y \in t_i \cdot C$ and $t_i \to P(y)$

$\forall i: \; P(x+y) \leq s_i + t_i = P(x) + P(y) \implies P(x+y) \leq P(x) + P(y)$

(P4) $P(x) = 0 \implies x = 0$

$P(x) = 0 \iff inf\{t > 0 \mid x \in t \cdot C\} = 0 \implies \exists\, (t_k)_{k \in \mathbb{N}}$ (A sequence) $\mid t_k \to 0, x \in t_k \cdot C \; \forall_k$

Assume: $x \neq 0$

This implies that $(\frac{x}{t_k})_{k \in \mathbb{N}}$ is unbounded, which is a contradiction since we already know C is bounded.

□

3 Normed Function Spaces

Definition 8 *Space of Continuous Functions: Left T be a metric space,* $e^b(T) := \{f : T \to \mathbb{R} \mid f$ *is continuous and bounded* $\}$

Definition 9 *Bounded: A function where* $\exists\, c \in \mathbb{R} \mid \forall t \in T, |f(t)| < c$

Then the space $e^b(t)$ we choose:

$$||f||_\infty := \underset{t \in T}{sup} \, |f(t)|$$

Definition 10 *Supremum: The smallest upper-bound of a set.*

The norm exists since we are in the space of bounded functions, bounded from above.
Then the space $e^b(t)$ with norm $|| \cdot ||_\infty$ is a Banach space.

Definition 11 *Banach Space: A normed space* $(x, || \cdot ||)$ *where* $(x, d_{|| \cdot ||})$ *is a complete metric space.*

Proof: To prove that we can convert any space $e^b(t)$ into a Banach Space via the infinity norm: $|| \cdot ||_\infty$ we can:

(i) Check Vector space axioms

(ii) Norm Axioms

(iii) Completeness: follows from the fact that $|| \cdot ||_\infty$ includes uniform convergence

□

3.1 ML Application: Tikhonov Regularization

Regularization covers various methods of improving the generalization of a solution or function. Explicit regularization involves penalizing the exploration of certain solutions to an optimization problem by adding a term. [JKC17]

This method is effective for ill-posed problems that are highly susceptibly to noise. Consider a general minimization problem.

$$\min_{x \in \mathbb{R}^n} ||Ax - b||$$

Where b is data contaminated with some noise: $b = \hat{b} + e$ Note the usage of the Euclidean Norm:

$$|| \cdot || = \sqrt{\sum_{k=1}^{n} |x_k|^2}$$

A more useful solution would be:

$$\min_{\hat{x} \in \mathbb{R}^n} ||A\hat{x} - \hat{b}||$$

$\hat{x}$ is a more desirable solution, as its what ignoring the noise in b returns. However, since $\hat{b}$ is unknown, the solution above is unattainable. Instead we can add a penalty term to the solution of a least-squares problem to get a reasonable approximation.

$$\min_{\hat{x} \in \mathbb{R}^n} \{||Ax - b||^2 + ||\mathbf{L}_\mu \mathbf{x}||^{\mathbf{2}}\} \approx \min_{\hat{x} \in \mathbb{R}^n} ||A\hat{x} - \hat{b}||$$

Where μ is the factor of regularization (a parameter we can adjust), and L is some matrix, usually the Identity Matrix, though there are alternatives provided L is a linear function of μ. This method is referred to as **Tikhonov Regularization**.[MF11]

Super Resolution involves using multiple noisy low resolution (LR) images to estimate a corresponding High Resolution image. The problem can be modeled like:

$$Y = Hf + n$$

- Y: vector of LR images
- H: Degradation operator
- f: Acquired HR image
- Gaussian white noise contamination

If the amount of LR images is insufficient, and H is often ill-conditioned. The model cant be inverted (to attain the HR image from the LR images) without losing stability. Instead, researchers applied Tikhonov Regularization to reform the problem as such:

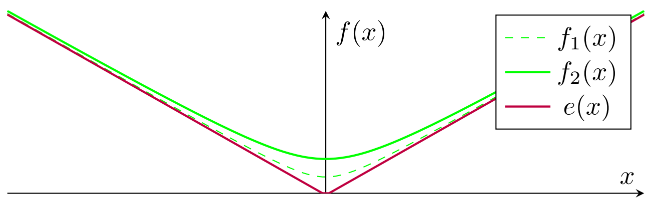
$$\min_{f} \{||Y - Hf||^2_{L_2} + \alpha ||Cf||^2_{L_2}\}$$

C in this case represents a high-pass filter to reduce the effect of the Gaussian noise of the LR images. This system proved effective at improving readability of slice-select MRI data without sacrificing the Signal-to-Noise ratio (SNR). [XZW07]

4 Space of Differentiable Functions

Definition 12 *Space of Differentiable Functions: Let* $[a,b] \subset \mathbb{R}, e`([a,b]) = \{f : [a,b] \to \mathbb{R} \mid f$ *is continuously differentiable or:* $(f \in C^1(a,b))\}$

Which Norm???:



Does a better norm exist? Of course! There are many examples:

- $||f|| := \sup_{t \in [a,b]} max\{|f(t)|, |f`(t)|\}$
- $||f|| := ||f||_\infty + ||f`||_\infty \, e`([a,b])$ with any of these two norms is a Banach space.

5 Insights into Banach Spaces in Machine Learning

Banach spaces, which are **complete normed vector spaces**, provide a powerful framework for understanding functional spaces that appear in machine learning. Their structure supports concepts like convergence, optimization, and function approximation, which are fundamental in ML.

5.1 Feature Spaces in Learning Models

Many machine learning models operate in high-dimensional function spaces, such as function approximation in regression and kernel methods. A key example is the space of continuous functions $C([a,b])$, equipped with the sup norm:

$$\|f\| = \sup |f(x)|$$

which forms a Banach space and is widely used in function approximation.

5.2 Norms and Stability in Learning

Normed spaces, including Banach spaces, help analyze the **stability** and **generalization** of learning algorithms. For instance, Lipschitz continuity ensures controlled changes in outputs relative to inputs, a property naturally defined in Banach spaces.

5.3 Optimization and Convergence

Many optimization algorithms in ML (e.g., gradient descent, convex optimization) are analyzed in Banach spaces to ensure convergence. The **Banach Fixed-Point Theorem** guarantees the convergence of certain iterative methods, which is crucial in neural network training.

5.4 Dual Spaces and Regularization

The dual space of a Banach space, which consists of all continuous linear functionals, is useful for **regularization techniques**. Examples include:

- **Lasso Regression** (L1 norm regularization)
- **Ridge Regression** (L2 norm regularization)

These techniques help control overfitting by penalizing the norm of model coefficients.

6 Applications of Banach Spaces in Machine Learning

6.1 Reproducing Kernel Banach Spaces (RKBS)

RKBS extends the concept of Reproducing Kernel Hilbert Spaces (RKHS) to Banach spaces. This is used in kernel methods such as **Support Vector Machines (SVMs)** for learning nonlinear relationships.

6.2 Neural Networks and Functional Spaces

Function spaces modeled as Banach spaces help understand the expressivity of neural networks. The **universal approximation theorem** holds in certain Banach spaces of continuous functions.

6.3 Metric Learning and Embedding Spaces

Many embedding techniques (e.g., word embeddings, manifold learning) operate in Banach spaces where distances are defined via norms.

6.4 Inverse Problems and Deep Learning

Many deep learning problems involve solving inverse problems (e.g., image reconstruction), where solutions naturally exist in Banach spaces. Techniques such as **variational methods** are formulated in this framework.

6.5 Sparse Learning and Compressed Sensing

Banach spaces, particularly ℓ_p-spaces for $0 < p \leq 1$, are crucial in sparse optimization and **compressed sensing**, which reconstructs signals from minimal measurements.

7 Conclusion

Banach spaces provide a rigorous mathematical foundation for several areas in machine learning, including **optimization, kernel methods, neural network analysis, and sparse learning**. Their completeness and norm-based structure ensure that key algorithms converge and perform reliably.

References

[JKC17] Vladimir Golkov Jan Kukačka and Daniel Cremers, *Regularization for deep learning: A taxonomy.*

[MF11] Lothar Reichel Martin Fuhry, *A new tikhonov regularization method.*

[XZW07] Ed X. Wu Xin Zhang, Edmund Y. Lam and Kenneth K.Y. Wong, *Application of tikhonov regularization to super-resolution reconstruction of brain mri images*, Medical Imaging and Informatics **2** (2007), 51–55.