Mathematics for Machine & Deep Learning (Sample)

Dr. Yves J. Hilpisch with GPT-5

October 3, 2025

THE AI ENGINEER

Contents

Preface

What This Book Is

This book is a focused, modern mathematics backbone for machine learning, deep learning, and large language models. It collects exactly the results that practitioners reach for when they design models, read papers, and reason about training dynamics and generalization—no broader and no narrower. The aim is fluency: turning definitions into usable tools and theorems into everyday instincts.

Approach & Objects

We build from first principles and keep the line of sight to applications clear.

- **Foundations:** sets, logic, functions, inequalities.
- **Linear Algebra & Matrices:** vector spaces, eigen/SVD, projections; the workbench for representations and optimization.
- **Analysis & Calculus:** continuity/differentiability in $\mathbb{R}^d$, Taylor models, integration.
- **Convex Analysis & Optimization:** geometry of sets/functions, first/second order, constraints, and stochastic variants.
- **Probability & Random Matrices:** concentration, tail bounds, random features/spectra.
- **Information, Transforms, Kernels, OT:** information theory, Fourier/functional tools, RKHS, and transport.
- **Numerical Analysis:** floating–point, conditioning, iterative linear algebra, and optimization stability.

Throughout, statements are minimal but sharp; proofs highlight the one or two key moves (compactness, orthogonality, convexity, or a concentration inequality). Boxes labeled *At a Glance* summarize takeaways; examples and exercises focus on what you will actually do.

Why It Matters for ML/DL/LLMs

Modern ML systems are mathematical artifacts. Training stability, generalization, prompt sensitivity, and scaling behavior all rest on a small set of ideas:

- **Representation & Geometry** (linear algebra, manifolds) shape expressivity and inductive bias.
- **Optimization** (convex tools, nonconvex heuristics) governs convergence, step selection, and robustness.

- **Uncertainty** (measure, concentration, random matrices) explains regularization, averaging, and spectral effects.
- **Information & Transforms** connect compression, attention, kernels, and frequency–domain reasoning.
- **Numerics** ensures the math survives floating–point, finite precision, and iterative solvers.

LLMs amplify these needs: spectra and conditioning affect transformer stability; concentration and random matrix theory underlie initialization and scaling; convex analysis informs proximal and preconditioning ideas used inside nonconvex training; and transport distances support alignment and evaluation.

What You Are Holding

This sample includes the full Preface and Chapter 1. The chapter sets notation and core tools we use repeatedly across the book. The complete work adds the remaining chapters, part introductions, reference appendices, and a visual companion, all with the same styling and structure you see here.

How to Read

Skim the *At a Glance* callouts first, then read until an example feels routine. Treat exercises as a checklist of skills: if you can do them cleanly, you own the topic. When in doubt, draw the geometry, check the inequality's equality cases, and write down the exact assumptions used by a theorem.

Part I

Foundations

Why Formalism Pays Off

Modern ML/DL systems run on numbers, but succeed because of structure. The foundations in this part turn vague intuitions ("this should converge", "that should be small") into tools you can apply on day one of an ML project. Sets and logic fix the language; real analysis supplies limits, compactness and completeness; inequalities give you a Swiss-army knife for bounding losses, gradients, and errors. Together they let you reason about training dynamics, generalization, stability, and the tradeoffs that appear everywhere from optimization schedules to normalization layers.

Two recurring patterns appear throughout the Primer and in real systems:

- Translate an engineering question into a mathematical object (a set, a function, a sequence, an operator), then
- Control it with a bound (via an inequality), a limit (via completeness), or a decomposition (via later parts).

You will see these patterns again in optimization (smoothness bounds, Lipschitz constants), in probability (concentration and expectations), and in matrix analysis (spectral, operator-norm reasoning). This part is the on-ramp.

At a Glance

- **Ch. 1 — Sets, Logic, Functions** — the language of precise ML claims (assumptions, equivalences, images/preimages) and the habits behind correct proofs and careful API thinking.
- **Ch. 2 — Real Numbers, Sequences, Series** — limits, completeness, compactness; the substrate for convergence arguments, existence of minimizers, and continuity of training maps.
- **Ch. 3 — Inequalities Toolkit** — AM–GM/CS/Hölder/Minkowski/Jensen as everyday instruments to bound losses, gradients, and residuals; knowing when each tool bites.

1 Sets, Logic, and Functions

Motivation

Sets, logic, relations, and functions are the language of all later chapters. Clear, standard definitions and a few core proof patterns prevent subtle mistakes in linear algebra, analysis, probability, and optimization. This chapter fixes notation, reviews basic logic, formalizes functions and relations, and establishes the completeness property of $\mathbb{R}$ via suprema and infima.

At a Glance

- **Set Algebra & Logic** — operations, quantifiers, De Morgan; the grammar for proofs, events in probability, and feasible sets in optimization.
- **Functions** — image/preimage; injective/surjective/bijective; inverses; track constraints under maps, define random variables, and reason about invertibility.
- **Relations** — equivalence, partitions, partial orders; build quotients in algebra/geometry and model order structures (cones, lattices).
- **Cardinality** — countable vs. uncountable; diagonal arguments; CSB (statement); construct enumerations, use separability, and know limits of countable methods.
- **Bounds & Completeness** — sup / inf and the least upper bound axiom; justify convergence arguments and existence of extrema on compact sets.
- **Proof Patterns** — direct, contrapositive, contradiction, induction; reusable templates powering later theorems across the Primer.

1.1 Sets and Basic Logic

This section establishes the common language for reasoning throughout the Primer. Set operations and basic logic govern the algebra of events in probability, domains and ranges in functions, and feasible regions in optimization. Mastery here reduces later cognitive load when the same patterns appear inside linear algebra, measure theory, and topology.

Definition 1.1 (Set operations)**.** Given sets A, B, define union $A \cup B$, intersection $A \cap B$, difference $A \setminus B$, complement A^c (relative to an ambient universe U), and Cartesian product $A \times B = \{(a, b) : a \in A,\ b \in B\}$. Standard conventions follow [2].

Definition 1.2 (Quantifiers and implications)**.** Statements use $\forall$ ("for all") and $\exists$ ("there exists"). The implication $P \Rightarrow Q$ is logically equivalent to $\neg P \vee Q$ and contraposition $\neg Q \Rightarrow \neg P$.

Theorem 1.3 (De Morgan's laws)**.** *For subsets of a fixed universe U,*

$$(A \cup B)^c = A^c \cap B^c, \qquad (A \cap B)^c = A^c \cup B^c.$$

Proof. We show $(A \cup B)^c = A^c \cap B^c$. For any $x \in U$, $x \in (A \cup B)^c$ iff $x \notin A \cup B$ iff ($x \notin A$ and $x \notin B$) iff $x \in A^c \cap B^c$. The other identity is analogous. ☐

Proof patterns to practice

- Direct proof: manipulate definitions and known results to reach the claim.
- Contrapositive: prove $\neg Q \Rightarrow \neg P$ instead of $P \Rightarrow Q$.
- Contradiction: assume the claim false and derive an impossibility.
- Induction: establish a base case and an inductive step (weak/strong).

1.2 Functions: Image, Preimage, and Invertibility

Functions formalize mappings between sets and are the backbone of analysis, optimization, and probability (random variables are functions; optimization minimizes functions). Images and preimages are indispensable for translating constraints and for measuring sets under transformations, later used in integration and measure theory.

Definition 1.4 (Function, injective/surjective/bijective)**.** Let $f\colon X \to Y$ be a rule assigning each $x \in X$ a unique $f(x) \in Y$. f is *injective* if $f(x) = f(x') \Rightarrow x = x'$, *surjective* if $\operatorname{im} f := f(X) = Y$, and *bijective* if both injective and surjective (then f has an inverse $f^{-1}\colon Y \to X$).

Definition 1.5 (Image and preimage)**.** For $S \subseteq X$ and $T \subseteq Y$, define the image $f(S) = \{f(x) : x \in S\}$ and the preimage $f^{-1}(T) = \{x \in X : f(x) \in T\}$.

Proposition 1.6 (Image/preimage algebra)**.** *For any $A, B \subseteq X$ and $C, D \subseteq Y$:*

1. *$f(A \cup B) = f(A) \cup f(B)$ and $f(A \cap B) \subseteq f(A) \cap f(B)$ with equality if f is injective.*
2. *$f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D)$ and $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$.*
3. *$f^{-1}(Y \setminus C) = X \setminus f^{-1}(C)$.*

Proof. (a) If $y \in f(A \cup B)$ then $y = f(x)$ for some $x \in A \cup B$, hence $y \in f(A) \cup f(B)$. Conversely, $f(A) \cup f(B) \subseteq f(A \cup B)$ is immediate. For intersections, if $x \in A \cap B$ then $f(x) \in f(A) \cap f(B)$, so $f(A \cap B) \subseteq f(A) \cap f(B)$. If f is injective and $y \in f(A) \cap f(B)$ then $y = f(a) = f(b)$ with $a \in A$, $b \in B$; injectivity gives $a = b \in A \cap B$, hence $y \in f(A \cap B)$. (b)–(c) follow directly from elementwise logic. ☐

Proposition 1.7 (Left/right cancellability)**.** *Let $f\colon X \to Y$.*

1. *f is injective iff for all sets $A, B \subseteq X$, $f(A) = f(B)$ implies $A = B$.*
2. *f is surjective iff for every $T \subseteq Y$, $f(f^{-1}(T)) = T$.*

Proof. (a) If f is injective and $f(A) = f(B)$, then for any $x \in A$ we have $f(x) \in f(B)$, so $x \in B$ by injectivity applied to a preimage witness; hence $A \subseteq B$ and symmetrically $B \subseteq A$. Conversely, if f is not injective, some $x \neq x'$ satisfy $f(x) = f(x')$; taking $A = \{x\}$ and $B = \{x'\}$ yields $f(A) = f(B)$ while $A \neq B$. (b) If f is onto, every $y \in T$ has $x \in X$ with $f(x) = y$, so $y \in f(f^{-1}(T))$. The reverse inclusion always holds. If f is not onto, let $T = Y \setminus \operatorname{im} f$; then $f(f^{-1}(T)) = \varnothing \neq T$. ☐

Example 1.8 (Simple numerical intuition). Let $f\colon \mathbb{Z} \to \mathbb{N}$ be $f(n) = |n|$. Then f is not injective because $f(1) = f(-1)$, and not surjective because $0 \in \mathbb{N}$ has no preimage if $\mathbb{N} = \{1, 2, \ldots\}$; if instead $\mathbb{N} = \{0, 1, 2, \ldots\}$ then 0 is hit by $n = 0$.

1.3 Relations: Equivalence and Order

Relations generalize the idea of comparing or grouping elements. Equivalence relations produce partitions used in quotient structures (appearing in linear algebra and geometry), while partial orders model hierarchies like divisibility and set inclusion, relevant for convex cones and lattice structures later.

Definition 1.9 (Binary relation)**.** A relation $\mathcal{R}$ on a set X is a subset $\mathcal{R} \subseteq X \times X$. We write $x \,\mathcal{R}\, y$ when $(x, y) \in \mathcal{R}$.

Definition 1.10 (Equivalence relation and classes)**.** An *equivalence relation* is reflexive, symmetric, and transitive. For $x \in X$, the equivalence class is $[x] = \{y \in X : y \,\mathcal{R}\, x\}$. The set of all equivalence classes $X/\mathcal{R} = \{[x] : x \in X\}$ forms a partition of X.

Example 1.11 (Same parity on $\mathbb{Z}$). Define $m \,\mathcal{R}\, n$ iff $m - n$ is even. Then $\mathcal{R}$ is an equivalence relation with two classes: the evens and the odds.

Definition 1.12 (Partial order)**.** A *partial order* is reflexive, antisymmetric, and transitive. We write $(X, \preceq)$. A *total order* also satisfies that for any x, y, either $x \preceq y$ or $y \preceq x$.

Example 1.13 (Divisibility). On $\mathbb{N}$, define $a \preceq b$ iff a divides b. This is a partial order (not total). Minimal, maximal, least, and greatest elements differ in general.

1.4 Cardinality and Countability

Cardinality compares the "sizes" of infinite sets. Countability arguments recur when building spaces (e.g., dense subsets of $\mathbb{R}$ or separability of function spaces) and when justifying constructions such as enumerating rational approximations used in analysis and probability.

Definition 1.14 (Finite, countable, uncountable)**.** A set is finite if bijective with $\{1, \dots, n\}$ for some $n \in \mathbb{N}$. It is *countably infinite* if bijective with $\mathbb{N}$, and *countable* if finite or countably infinite. Otherwise it is *uncountable.*

Lemma 1.15 (Basic countability facts)**.** *If A and B are countable, then $A \times B$ is countable. A countable union of countable sets is countable.*

Proof. Sketch: Enumerate $A = \{a_1, a_2, \dots\}$ and $B = \{b_1, b_2, \dots\}$. Pair indices via a diagonal enumeration of $\mathbb{N} \times \mathbb{N}$ (e.g., by nondecreasing sums $i + j$). For a countable union $\bigcup_{k \geq 1} A_k$, interleave the enumerations of A_k along diagonals. □

Theorem 1.16 (Countability of $\mathbb{Q}$)**.** *The set of rational numbers $\mathbb{Q}$ is countable.*

Proof. Write each rational as p/q with $p \in \mathbb{Z}$, $q \in \mathbb{N}$, $\gcd(p, q) = 1$. The set $\mathbb{Z} \times \mathbb{N}$ is countable by Lemma 1.15. Map (p, q) to p/q; this is surjective onto $\mathbb{Q}$, hence $\mathbb{Q}$ is a countable image of a countable set and is therefore countable. □

Theorem 1.17 (Cantor's diagonal, statement)**.** *There is no bijection between $\mathbb{N}$ and $(0, 1)$; in particular, $(0, 1)$ and $\mathbb{R}$ are uncountable [1].*

Remark 1.18 (Decimal caveat). Real numbers may have two decimal expansions (e.g., $0.999 \cdots = 1$). Diagonal arguments avoid this by constructing a number differing in at least one digit from every entry.

Theorem 1.19 (Cantor–Schröder–Bernstein, statement)**.** *If there exist injections $f : A \to B$ and $g : B \to A$, then there is a bijection $h : A \to B$.* *[2, §3]*

1.5 Bounds, Supremum, and Infimum

Bounding is central to convergence, continuity, and optimization. The least upper bound (completeness) axiom distinguishes $\mathbb{R}$ from $\mathbb{Q}$ and underlies later theorems in sequences/series, integration (Monotone and Dominated Convergence), and optimization (existence of minima on compact sets).

Definition 1.20 (Bounds)**.** For $S \subseteq \mathbb{R}$, an *upper bound* is u with $s \leq u$ for all $s \in S$; similarly a *lower bound*. S is *bounded above* if it has an upper bound. The *supremum* $\sup S$ is the least upper bound; the *infimum* $\inf S$ is the greatest lower bound.

Least upper bound property

Completeness of $\mathbb{R}$: every nonempty set $S \subseteq \mathbb{R}$ that is bounded above has a supremum in $\mathbb{R}$ (and dually for inf). This fails in $\mathbb{Q}$. A classical reference is [3, Ch. 1].

Theorem 1.21 (Existence and basic properties of sup/inf)**.** *If $S \subseteq \mathbb{R}$ is nonempty and bounded above, then $\sup S$ exists and satisfies: (i) $s \leq \sup S$ for all $s \in S$; (ii) for all $\varepsilon > 0$ there exists $s \in S$ with $\sup S - \varepsilon < s \leq \sup S$. Analogous statements hold for* inf.

Proof. Existence is the least upper bound axiom. For (ii), if no $s \in S$ satisfied $\sup S - \varepsilon < s$, then $\sup S - \varepsilon$ would be an upper bound smaller than $\sup S$, contradicting minimality. □

Example 1.22 (Intuition via nested intervals)*.* Let $S = \{x \in \mathbb{R} : x^2 < 2\}$. Then S is bounded above (e.g., by 2). By Theorem 1.21, $\alpha = \sup S$ exists and equals $\sqrt{2}$. Indeed $\alpha^2 \leq 2$ else α would not bound S, and $\alpha^2 \neq 2$ with $\alpha^2 < 2$ would contradict maximality by increasing α slightly.

Remark 1.23 (Why $\mathbb{Q}$ is incomplete)*.* Consider $S = \{x \in \mathbb{Q} : x^2 < 2\}$. It is bounded above in $\mathbb{Q}$, but has no rational least upper bound since $\sqrt{2} \notin \mathbb{Q}$. This illustrates the necessity of working in $\mathbb{R}$ for analysis.

1.6 Worked Examples

Examples ground the definitions and turn formulas into usable instincts. Read them as templates: extract the algebraic pattern you need (images/preimages; bijections; equivalence classes) and reuse it in later contexts like measurable preimages and linear map images.

Example 1.24 (Images and preimages)*.* Let $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = x^2$ and $A = [-2, 1]$, $B = (1, 3)$. Then $f(A) = [0, 4]$ while $f^{-1}(B) = (-\sqrt{3}, -1) \cup (1, \sqrt{3})$. Note $f(A \cap f^{-1}(B)) = (1, 3)$ and compare with Proposition 1.6.

Example 1.25 (Constructing bijections)*.* There is a bijection $\mathbb{N} \to \mathbb{Z}$: map $0 \mapsto 0$ and $2k \mapsto k$, $2k + 1 \mapsto -(k + 1)$ for $k \geq 0$. Hence $\mathbb{Z}$ is countable.

Example 1.26 (Equivalence classes)*.* On $X = \mathbb{R} \setminus \{0\}$, define $x \sim y$ iff $x/y > 0$. The classes are the positive reals and the negative reals.

1.7 Checklist

By the end of this chapter you should be able to:

- **Prove**: set identities (e.g., De Morgan) and manipulate quantifiers correctly.
- **Classify**: whether a function is injective/surjective/bijective; compute images and preimages.

- **Organize**: sets via equivalence relations and partitions; recognize partial orders.
- **Enumerate**: classic sets and prove countability (e.g., $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{N} \times \mathbb{N}$).
- **Apply**: the least upper bound property to identify sup/inf and prove their properties.

1.8 Exercises

Short problems to solidify concepts and build fluency.

Exercise 1.27. **Set algebra warm-up**. Prove $(A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$.

Exercise 1.28. **Preimage identities**. Let $f \colon X \to Y$ and $C, D \subseteq Y$. Prove $f^{-1}(C \setminus D) = f^{-1}(C) \setminus f^{-1}(D)$.

Exercise 1.29. **Image of intersections**. Give an example where $f(A \cap B) \subsetneq f(A) \cap f(B)$. Then prove equality holds for all A, B iff f is injective.

Exercise 1.30. **Equivalence and partition**. Show that if $\{A_i\}_{i \in I}$ is a partition of X, the relation $x \sim y$ iff $x, y \in A_i$ for some i is an equivalence relation with classes A_i.

Exercise 1.31. **Countable unions (hint)**. Let $A_k = \{(k, n) : n \in \mathbb{N}\}$. Exhibit an explicit bijection $\mathbb{N} \to \bigcup_{k \geq 1} A_k$ by enumerating along diagonals of $\mathbb{N} \times \mathbb{N}$.

Exercise 1.32. **Countability of $\mathbb{Q}$ (variant)**. Construct an explicit enumeration of positive rationals by the Calkin–Wilf tree or by arranging fractions with $p + q$ increasing and skipping nonreduced ones.

Exercise 1.33. **Supremum property**. Let $S = \{\frac{n}{n+1} : n \in \mathbb{N}\}$. Prove $\sup S = 1$ but $1 \notin S$. Verify Theorem 1.21(ii).

Exercise 1.34. **Why $\mathbb{Q}$ is incomplete**. Let $T = \{x \in \mathbb{Q} : x^2 < 3\}$. Show T has no supremum in $\mathbb{Q}$.

Exercise 1.35. **Optional, challenge**. State and prove the pigeonhole principle using functions and cardinality, and apply it to show that in any set of $n + 1$ integers, two have the same remainder mod n.

Bibliography

[1] Georg Cantor. On a property of the collection of all real algebraic numbers. *Journal f"ur die reine und angewandte Mathematik*, 1874.

[2] Paul R. Halmos. *Naive Set Theory*. Springer, 1974.

[3] Walter Rudin. *Principles of Mathematical Analysis*. McGraw–Hill, 3rd edition, 1976.