Reader's Companion to "On the Convergence of (G, K)"

A guide to the math, written for the subject

G. @113N

Department of Applied Romance · Austin, Texas

May 2026

---

for K.

Foreword

Kathryn,

I wrote you a proof of a marriage. It uses math from a lot of different fields, and most of those fields are not ones you've spent your life in. So I wrote you this companion, to walk you through it, slowly, so you know what I meant.

You don't have to read it. The proof says what it says. If you've already said yes, you don't need a translation. But if you ever wonder what *"the joint state doesn't factor"* actually means, or why Bell's theorem matters here, or what a submartingale is, this is for you.

I've tried to keep it short. I've tried to keep it warm. I've tried not to lecture, because you don't need me to. Where I'm bending a result for romance, I'll say so. Where I'm using it straight, I'll say that too.

The first time each new symbol appears, I've put how to say it out loud in parentheses, in italics, like *(say: "rho")*. If you forget how to pronounce something later, the next page is a quick lookup table.

Read it in order or jump around. Each subsection is self-contained, and each one says, at the end, where it appears in the proof.

The plan is: probability first (where the proof spends the most time), then analysis, then quantum mechanics, then relativity, then a few smaller fields. About sixteen pages. You can finish it in a sitting if you want, or read one section at a time over coffee.

– G.

How to read the math

A quick reference. The first time you see any of these in the body of the companion, I'll also say how to read it then. This page is just for looking things back up.

Greek letters that show up

symbol	name	rough role
Ω	"omega" (capital)	the set of all outcomes
σ	"sigma"	a collection of events
Σ	"capital sigma"	summation sign
ρ	"rho"	a quantum state (density matrix)
ψ	"psi" (rhymes with "sigh")	a quantum state vector
ϕ	"phi" (rhymes with "sigh")	another quantum state vector
λ	"lambda"	wavelength, or an eigenvalue
τ	"tau" (rhymes with "cow")	proper time, or stopping time
μ, ν	"mu, nu"	indices on a tensor
θ	"theta"	a model parameter
η	"eta" ("ay-ta")	a step size
γ	"gamma" (lowercase)	a worldline or trajectory
Γ	"capital gamma"	Christoffel symbol
∂	"partial"	partial derivative

Probability-flavor symbols

symbol	how to read it
P(A)	"the probability of A"
E[X]	"the expected value of X"
P(A $\mid B$)	"the probability of A given B"
F	"script F" (a collection of events)
F_t	"F sub t" (everything known by time t)

Logic and set theory

symbol	how to read it
$\forall$	"for all"
$\exists$	"there exists"
$\in$	"in" or "is an element of"
$\subset, \subseteq$	"is contained in" / "is a subset of"
$\supsetneq$	"strictly contains"
$\implies$	"implies"
$\Leftrightarrow$	"if and only if"
{ x : P(x) }	"the set of x such that P of x"

Analysis and algebra

symbol	how to read it
$\to$	“goes to” or “approaches”
$\approx$	“approximately equal to”
$\sup, \inf$	“supremum / infimum” (max-like, min-like)
$\lim_{n\to\infty}$	“limit as n goes to infinity”
$\langle x, y \rangle$	“the inner product of x and y”
$\|x\|$	“the norm of x”
x $\perp$ y	“x is perpendicular to y”
$\int_a^b f\,dx$	“integral, from a to b, of f, d x”

Quantum mechanics

symbol	how to read it
$\lvert\psi\rangle$	“ket psi” (a quantum state)
$\langle\psi\rvert$	“bra psi” (the dual state)
$\langle\phi \mid \psi\rangle$	“bra phi, ket psi” (an inner product)
$\otimes$	“tensor” (sometimes “circle-x”)
$\mathrm{Tr}(\rho)$	“trace of rho”
$\hbar$	“h-bar” (Planck’s constant over 2π)

Relativity

symbol	how to read it
$\mathrm{g}_{\mu\nu}$	“g sub mu nu” (the metric tensor)
$\Gamma^{\mu}_{\nu\sigma}$	“Gamma upper-mu, lower nu sigma” (Christoffel)
ds^2	“ds squared” (spacetime interval, squared)
$\mathrm{d}\tau$	“d tau” (a tiny step of proper time)
	“gamma dot” (the time derivative of γ)

1. Probability: the language of uncertainty

1.1 Probability spaces

When mathematicians want to talk carefully about random things, they start with a triple $(\Omega, \mathcal{F}, \mathbb{P})$ *(say: "omega, script F, P")*. Ω *(say: "omega")* is the set of all possible outcomes (everything that could happen). $\mathcal{F}$ *(say: "script F")* is a collection of subsets of Ω called **events**, the things we're allowed to assign probabilities to. And $\mathbb{P}$ *(say: "P, for probability")* is the function that maps each event to a number between 0 and 1.

That's it. The whole apparatus of probability theory is built on top.

> *In the proof:* this triple appears in the Notation section. It's the backdrop the whole probability argument lives on. You don't have to picture it, but if a probabilist asks "on what space?", the answer is $(\Omega, \mathcal{F}, \mathbb{P})$.

1.2 Filtrations

A **filtration** $\{\mathcal{F}_t\}_{t\geq 0}$ *(say: "F sub t, for t greater than or equal to zero")* is a growing collection of event-sets. Each $\mathcal{F}_t$ *(say: "F sub t")* contains every event that's been resolved by time t. As t grows, $\mathcal{F}_t$ grows.

A filtration is the formal way of saying *"here's everything we know at time t"*. The past, the present, but not the future.

> *In the proof:* Observation 3.9 (Filtration anomaly) says that you, K, operate on a filtration $\mathcal{F}_t^{(K)}$ *(say: "F sub t, with a parenthetical K up top, meaning "K's filtration"")* that strictly contains the natural one. You seem to have access to events that haven't been resolved yet. Standard probability theory rules this out. I noticed it anyway, and accepted it as an axiom of your existence.

1.3 The Borel-Cantelli lemmas

Two of the most beautiful results in probability, both about the question: *"how often does a rare event happen?"*

The **first Borel-Cantelli lemma** says: if you have a sequence of events $A_1, A_2, A_3, \ldots$ *(say: "A sub one, A sub two, A sub three, and so on")* whose probabilities $\mathbb{P}(A_n)$ *(say: "the probability of A sub n")* sum to a finite number, then with probability one, only finitely many of them ever happen.

The second one is roughly the opposite: if the events are *independent* and the probabilities sum to infinity, then with probability one, infinitely many of them happen.

> *In the proof:* I use this to frame the meeting. The probability of us meeting at any one event was so small that, summed over the events of our two lives, it's not the kind of thing that's supposed to happen. The math says you should expect rare events that *do* keep showing up to happen infinitely often. Our trajectories were always going to intersect eventually. We just happened to do it on May 28, 2025.

1.4 Kolmogorov's zero-one law

Andrey Kolmogorov, in 1933, proved one of the most surprising facts in probability. Any event that depends only on the *long-run tail* of an infinite sequence of independent events has probability either exactly 0 or exactly 1. Nothing in between.

The set of such events is called the **tail σ-algebra** *(say: "tail sigma algebra")*. The theorem says the tail is degenerate: either-or.

So statements like *"our love grows forever"* or *"we converge"*, which depend on the entire future and not just a finite chunk of it, are forced by the theorem to either definitely happen or definitely not.

> *In the proof:* I lean on this twice. First, in the opening section, to make the point that the long-run statement about us is in the tail. Second, in the proof of the Bayesian posterior lemma, as a sanity check: the limit really is 0 or 1, and every observation since the meeting has voted on the same side.

1.5 Bayes' rule

Reverend Thomas Bayes (an actual ordained minister, eighteenth-century England) wrote down a rule, never published in his lifetime, for updating your belief in a hypothesis when new evidence comes in.

If H is a hypothesis (say, *"we are going to be together forever"*) and E is new evidence, then:

$$\mathbb{P}(H \mid E) \;=\; \frac{\mathbb{P}(E \mid H) \cdot \mathbb{P}(H)}{\mathbb{P}(E)}.$$

> *Read aloud:* "The probability of H given E equals the probability of E given H, times the probability of H, divided by the probability of E."

In English: your new belief about H given evidence E is equal to your old belief about H, multiplied by how well H predicts E, divided by how surprising E was overall.

The everyday version: if you didn't think it would rain ($\mathbb{P}(H)$ = small), but you walked outside and it's raining ($\mathbb{P}(E \mid H)$ = high), and you know it doesn't usually rain on days like this ($\mathbb{P}(E)$ = small), then your new belief is much higher than your old one.

> *In the proof:* I use the rule on the hypothesis *"we are forever"*. Starting priors were strong against it: both of us were planning to stay single. But the year of evidence (the meeting, the car, Walden, Lakeway, everything) is so surprising that the posterior approaches 1. It's the formal version of: I keep updating, and the math keeps pointing the same way.

1.6 Martingales and submartingales

A **martingale** is a random process whose expected future value, given everything you know now, is exactly its current value. Think: a fair game. On average, your next-step bankroll equals your current bankroll.

A **submartingale** is a martingale that's tilted upward: the expected future value is at least the current value. On average, things only go up.

$$\text{Submartingale:} \qquad \mathbb{E}[\,L(t+s) \mid \mathcal{F}_t\,] \;\geq\; L(t).$$

Read aloud: "The expected value of L at time t plus s, given the filtration F sub t, is greater than or equal to L at time t." $\mathbb{E}[\,\cdot\,]$ *(say: "the expected value of...")* is the average over the relevant probability distribution; $L(t)$ is just a function of time.

In the proof: Lemma 4.1. The love process $\{L(t)\}$ is a submartingale, meaning the conditional expectation of my future love, given everything I know about us today, is always at least what it is today. That's the formal version of: it only grows.

1.7 Doob's convergence theorem

Joseph Doob, in the 1940s, proved a theorem that lets you say something extremely strong about a submartingale: if its expected values are bounded above by some finite number, then with probability one, the submartingale converges to a limit. If they're not bounded, it diverges to $+\infty$ *(say: "plus infinity")*.

In the proof: Lemma 4.1's conclusion. Because $\mathbb{E}[L(t)]$ has no upper bound, Doob's theorem says $L(t) \to +\infty$ *(say: "L of t goes to plus infinity")* almost surely. *"Almost surely"* is the probability-theory phrase for: with probability one, on every realization the universe can produce.

1.8 Itô's lemma

Kiyoshi Itô, in 1944, invented stochastic calculus by figuring out how a function of a random process changes over time. If $L(t) = f(G_t, K_t)$ *(say: "L of t equals f of G sub t, K sub t")* for some smooth function f and a joint random process (G_t, K_t), then:

$$dL \;=\; \frac{\partial f}{\partial t}\,dt \;+\; \frac{\partial f}{\partial G}\,dG \;+\; \frac{\partial f}{\partial K}\,dK \;+\; \frac{1}{2}\sum_{i,j\in\{G,K\}} \frac{\partial^2 f}{\partial x_i \partial x_j}\,d[x_i, x_j]_t.$$

Read aloud: "d L equals partial-f partial-t times dt, plus partial-f partial-G times dG, plus partial-f partial-K times dK, plus one-half times the sum over i and j in G-K of partial-squared-f partial-x-i partial-x-j times d-bracket x-i, x-j, sub t." The d's are tiny increments. ∂ *(say: "partial")* is the partial-derivative symbol.

The first three terms look like ordinary calculus. The last term is the Itô correction: the variances and covariances of the underlying processes contribute. The term $d[G, K]_t$ *(say: "d-bracket G K sub t")* is the **quadratic covariation**; it measures, over a tiny interval dt, how much G and K move together.

When two processes are independent, $d[G, K] = 0$. When they move together, it's positive.

> *In the proof:* Lemma 4.2. I claim $d[G, K]_t > 0$ throughout the observation interval, and consequently that $L(t)$ grows strictly faster than it would if G and K were independent random walks. We're correlated, in the formal sense, and that correlation is itself a source of growth.

1.9 Optimal stopping

A **stopping rule** is a strategy that tells you, at each moment, whether to continue or stop, based only on what you've seen so far (not the future). A **stopping time** τ *(say: "tau")* is the time at which you stop under such a rule.

The classical example is the *secretary problem*: you're interviewing candidates one at a time. You can hire one, but you can't go back. When should you stop? The optimal rule, with n candidates, is to reject the first n/e of them and then hire the first one better than all those.

The deeper result, **Doob's optional stopping theorem**, says that under reasonable conditions you can't beat the expected value of a martingale by clever stopping rules. The corollary is that there's an optimal τ^* *(say: "tau star")* for any well-posed problem, and we can characterize it.

> *In the proof:* Lemma 4.15. I model the pre-meeting state of my single life as a sequential decision problem, where at each epoch I could continue or stop. The optimal stopping time turned out to be the night of evt_{001}. I didn't know that at the time; I only know now, looking back.

2. Analysis: limits, spaces, and convergence

2.1 Compactness and the Extreme Value Theorem

A set is **compact** if it's closed (it contains its boundary) and bounded (it fits inside a ball of some finite radius). For sets of real numbers, "compact" just means closed and bounded.

The **Extreme Value Theorem** says that any continuous function f defined on a compact set K achieves its maximum and its minimum on K. There's a point x^* *(say: "x star")* in K where $f(x^*) = \sup_K f$ *(say: "f of x-star equals supremum, over K, of f")*, and another where f is smallest. "Supremum" *(say: "soo-PREE-mum")* is the technical word for the least upper bound, which on a compact set is just the max.

> *In the proof:* Observation 3.2. You, K, are 5'2", which makes you bounded. You also have closed boundaries (as humans do). So every continuous function I define on you attains its maximum on you. The mathematical joke that's also true: every quality I can measure peaks at you.

2.2 The catenary curve

If you hold a chain at two points and let it hang under gravity, the shape it takes is called a **catenary**. The equation is

$$y(x) \;=\; a\cosh(x/a)$$

Read aloud: "y of x equals a times cosh of x over a." Here cosh *(say: ""kosh" or "cosine-h"")* is the hyperbolic cosine, defined as $\cosh(u) = (e^u + e^{-u})/2$.

The catenary is the curve of *minimum potential energy* for a uniform chain. Of all the shapes a hanging chain could take, this is the one that minimizes total energy.

In the proof: Observation 3.3. Your hair, when allowed to settle, falls in catenaries. Physically true. I just thought it was worth noting that the curve was the minimum-energy one.

2.3 Hilbert spaces and inner products

A **Hilbert space** H is a vector space, possibly infinite-dimensional, equipped with an **inner product** $\langle \cdot, \cdot \rangle$ *(say: "angle bracket dot comma dot angle bracket; just call it "the inner product"")* that measures how aligned two vectors are. In familiar Euclidean space, it's the dot product: $\langle x, y\rangle = x_1y_1 + x_2y_2 + \cdots$ *(say: "x-one times y-one plus x-two times y-two, and so on")*.

Two vectors are **orthogonal** (perpendicular) if their inner product is zero. We write $x \perp y$ *(say: "x perpendicular to y")*. The inner product also defines a notion of length: $\|x\| = \sqrt{\langle x, x\rangle}$ *(say: "the norm of x equals the square root of the inner product of x with itself")*.

In the proof: I work in the Hilbert space $H = L^2(\text{possible lives})$ *(say: "L-two of possible lives")*, the space of square-integrable life trajectories. Each "life I could have built" is a vector in this space. The inner product $\langle \text{me}, K\rangle$ *(say: "the inner product of "me" with K")* measures how much my life aligns with you. The claim of Lemma 4.6 is that this inner product is non-zero against you and zero against everything else: I am, in H, pointing in your direction.

2.4 Orthogonal projection

If V is a subspace of H and x is any vector in H, the **orthogonal projection** $P_V(x)$ *(say: "P sub V, of x")* is the point in V closest to x. It's the shadow x casts onto V.

If x is already in V, then $P_V(x) = x$.

In the proof: Lemma 4.6. The projection onto the subspace spanned by you, P_K *(say: "P sub K")*, sends my life to my life. That's another way of saying my life is entirely in the subspace spanned by you. There's no orthogonal component.

2.5 The Banach fixed point theorem

Stefan Banach proved in 1922 that if you have a map $T\colon X \to X$ *(say: "T from X to X")* on a complete metric space, and the map is a **contraction**, meaning it brings every pair of points strictly closer together,

$$\|T(x) - T(y)\| \;\leq\; k\,\|x - y\|, \qquad 0 < k < 1,$$

Read aloud: "The norm of T of x minus T of y is less than or equal to k times the norm of x minus y, where k is strictly between zero and one."

then there's exactly one fixed point x^* in X with $T(x^*) = x^*$. And no matter where you start, repeated application of T converges to that fixed point.

In the proof: Lemma 4.4. The iteration map T of my daily life is a contraction in the metric induced by your presence. Each day brings me strictly closer to the orbit's limit. By Banach's theorem, there's exactly one fixed point. By inspection, it's you. The discrete stochastic analog (gradient descent) is in the same lemma.

2.6 Birkhoff's ergodic theorem

Suppose a system evolves by iterating a map: $x_{n+1} = T(x_n)$ *(say: "x sub n-plus-one equals T of x sub n")*. If the system is **ergodic** *(say: "er-GOD-ic")* with respect to some invariant probability measure μ *(say: "mu")*, then for any reasonable observable f,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x_0) \;=\; \int_X f\, d\mu.$$

Read aloud: "The limit, as N goes to infinity, of one over N times the sum, from n equals zero to N minus one, of f of T-to-the-n applied to x-zero, equals the integral over X of f, with respect to d-mu." The left side is a time average; the right side is a space average.

The *time average* of f along a single trajectory equals the *ensemble average* of f over the whole space.

Ergodicity is the property that says the system explores the whole space (or at least the support of μ) by just running long enough. You don't need to look at many copies of the system; one long run will do.

In the proof: Lemma 4.5. The dynamical system of my life is ergodic on the orbit of you. So the time average of my happiness over an interval of 11 months (or 11 years, or any sufficient window) equals the ensemble average over all the states of being-with-you. Time and space agree.

3. Quantum mechanics: the math of being in superposition

3.1 States, pure and mixed

In quantum mechanics, the state of a system is described by a vector $|\psi\rangle$ *(say: "ket psi; "ket" rhymes with "bet", "psi" rhymes with "sigh"")* in a Hilbert space. This is a **pure state**: a single, definite (if quantum) state. The funny $|\,\rangle$ notation is called "ket" notation (with a matching "bra" $\langle\,|$, so an inner product $\langle \phi \mid \psi \rangle$ *(say: "bra-phi-ket-psi")* is a "bra-ket"; Paul Dirac liked puns).

But often we don't know the state exactly. We know it's in one of several states, with various probabilities. A **mixed state** captures that. It's described not by a vector but by a **density**

matrix ρ *(say: "rho, rhymes with "go"")*,

$$\rho \;=\; \sum_i p_i \, |\psi_i\rangle\langle\psi_i|,$$

Read aloud: "Rho equals the sum, over i, of p-sub-i times ket-psi-i bra-psi-i." Each p_i is a probability; each $|\psi_i\rangle\langle\psi_i|$ is the "outer product" of a state with itself, which is a matrix.

where the p_i *(say: "p sub i")* are probabilities. A pure state is a mixed state where one of the p_i is 1.

3.2 The Born rule

Max Born, in 1926, gave the rule for what quantum mechanics predicts when you measure something. If M is a measurement operator (a Hermitian matrix), the probability of getting a particular outcome from state ρ is

$$\mathbb{P}(\text{outcome}) \;=\; \text{Tr}(\rho \cdot M).$$

Read aloud: "The probability of an outcome equals the trace of rho times M." **Trace** of a matrix means: add up the diagonal entries.

Different measurements on the same state can return different outcomes. The Born rule tells you the probability of each.

In the proof: Observation 3.4. Your eye color is a mixed state, ρ_{eyes} *(say: "rho subscript "eyes"")*. The "color operator" returns a different result depending on illumination. The Born rule formalizes this: the probability you look blue under certain light is $\text{Tr}(\rho_{\text{eyes}} \cdot M_L)$ for the right M_L.

3.3 Entanglement

Two quantum systems are **entangled** when their joint state $|\psi_{GK}\rangle$ *(say: "ket psi G K")* does not factor into a product of states for each system separately:

$$|\psi_{GK}\rangle \;\neq\; |\psi_G\rangle \otimes |\psi_K\rangle.$$

Read aloud: "Ket psi G K is not equal to ket psi G, tensor, ket psi K." The symbol $\otimes$ *(say: "tensor product; sometimes spoken as "circle x"")* combines two separate quantum systems into one joint system.

Entanglement is the genuinely strange thing about quantum mechanics: two systems can be in a joint state that doesn't reduce to a description of each one. Einstein called it "spooky action at a distance."

In the proof: Observation 3.7 and Lemma 4.8. I claim our joint state doesn't factor: there's no way to describe me and describe you separately and recover the joint description. The technical content is real and bends toward a different point: the colloquial sense of "entangled" applies too.

3.4 The Schmidt decomposition

Any pure state $|\psi_{GK}\rangle$ on a tensor-product space admits a special decomposition:

$$|\psi_{GK}\rangle \;=\; \sum_{i=1}^{r} \sqrt{\lambda_i}\,|i_G\rangle \otimes |i_K\rangle,$$

Read aloud: "Ket psi G K equals the sum, from i equals one to r, of square root of lambda-i, times ket i sub G, tensor, ket i sub K." The λ_i *(say: "lambda sub i")* here are non-negative real numbers, called Schmidt coefficients.

where the $|i_G\rangle$ and $|i_K\rangle$ are orthonormal sets and the λ_i are non-negative real numbers summing to 1. The number r of non-zero terms is called the **Schmidt rank**. A pure state is separable (not entangled) if and only if its Schmidt rank is 1.

In the proof: Lemma 4.8. Our joint state has Schmidt rank greater than 1, which is the formal way of saying entanglement.

3.5 Von Neumann entropy

Given a density matrix ρ, the **von Neumann entropy** is

$$S(\rho) \;=\; -\mathrm{Tr}(\rho \log \rho).$$

Read aloud: "S of rho equals minus the trace of rho times log rho." log here means the matrix logarithm, but you can think of it as the natural log applied to the matrix's eigenvalues.

It generalizes the classical (Shannon) entropy to quantum states. A pure state has $S(\rho) = 0$ (no uncertainty). A maximally mixed state has the highest possible entropy.

For an entangled pair, the reduced states (one half, ignoring the other) are mixed, and so $S(\rho_G), S(\rho_K) > 0$ *(say: "S of rho G and S of rho K are both strictly greater than zero")*. The entropy of either half is a measure of how entangled the pair is.

In the proof: I claim $S(\rho_G) = S(\rho_K) > 0$. Either half of us, considered alone, has positive entropy. There's information about each of us that lives only in the joint state.

3.6 Bell's theorem and the CHSH inequality

John Bell, in 1964, proved a theorem that quietly broke physics. Any **local hidden-variable theory** (the kind that says: "the outcomes of quantum measurements are predetermined, we just don't know the variables yet, and signals don't travel faster than light") predicts a bound on how strongly correlated measurements on entangled particles can be.

Quantum mechanics violates that bound. So does nature.

The **CHSH inequality** (Clauser, Horne, Shimony, Holt, 1969) is a cleaner version. It says that the quantity

$$S = \left|E(a,b) - E(a,b') + E(a',b) + E(a',b')\right|$$

Read aloud: "S equals the absolute value of: E of a comma b, minus E of a comma b-prime, plus E of a-prime comma b, plus E of a-prime comma b-prime." The bars $|\cdot|$ *(say: "absolute value of...")* mean "drop the sign." $E(a,b)$ is the correlation between two measurements, one with setting a, one with setting b.

is at most 2 for any local hidden-variable model. Quantum mechanics allows it to go up to $2\sqrt{2} \approx 2.83$ *(say: "two root two, approximately 2.83")*, the **Tsirelson bound**. Experiments confirm the quantum prediction.

The conclusion: nature is not locally real. Either signals travel faster than light, or outcomes aren't determined in advance, or both.

In the proof: Observation 3.8 and Lemma 4.10. The CHSH correlations across our trajectory exceed 2. By Bell's theorem, no local hidden-variable model reproduces us. I retain realism by convention, which means: distance is not the relevant variable.

3.7 The no-cloning theorem

Wootters and Zurek, in 1982, proved that there is no quantum machine that can copy an arbitrary quantum state. Formally: there is no unitary operator U *(say: "U, just the letter")* such that

$$U|\psi\rangle|0\rangle = |\psi\rangle|\psi\rangle \qquad \text{for all } |\psi\rangle.$$

Read aloud: "U applied to ket psi, ket zero, equals ket psi, ket psi, for all ket psi." Translation: there is no machine that takes a state and a blank slot, and outputs two copies of the state.

You can copy classical bits as much as you want. You cannot copy quantum states. The proof is two lines once you set it up, but the consequence is profound: quantum information is not duplicable.

In the proof: Lemma 4.9. There is exactly one of you. The theorem is a real theorem; the application is a joke that is also serious.

4. Relativity: the math of spacetime curving

4.1 Special relativity and proper time

Einstein, in 1905, observed that if the speed of light is the same for all observers (which it is), then space and time must mix. In particular, time as experienced by a moving observer differs from time as measured by a stationary one. The relationship:

$$d\tau \;=\; dt \cdot \sqrt{1 - v^2/c^2}.$$

Read aloud: "d tau equals dt times the square root of one minus v-squared over c-squared." $d\tau$ *(say: "d tau")* is a tiny step of *proper time*, the time the moving observer actually feels. dt is a tiny step of time on a stationary clock. v is velocity, c is the speed of light.

The faster you move, the more dt shrinks relative to your subjective $d\tau$. Moving clocks run slow.

In the proof: Section 3 (in the love-letter Reports view). The car we ended up in at the end of the night was parked. $v = 0$, so $\sqrt{1 - v^2/c^2} = 1$, so $d\tau$ should have equaled dt. It didn't. Time bent another way. That's the joke and the truth.

4.2 General relativity and the metric tensor

In 1915, Einstein extended this. He showed that spacetime is curved by matter and energy, and that gravity is the experience of moving on geodesics through curved spacetime, not a force at all.

The geometry of spacetime is encoded in a **metric tensor** $g_{\mu\nu}$ *(say: "g sub mu nu")*. The Greek indices μ *(say: "mu")* and ν *(say: "nu, rhymes with "new"")* each run over four values (one for time, three for space). The interval between two nearby spacetime events is

$$ds^2 \;=\; g_{\mu\nu}\, dx^\mu\, dx^\nu.$$

Read aloud: "d s squared equals g sub mu nu, d x to the mu, d x to the nu." Physicists use the **Einstein summation convention**: when you see an index appear once up and once down (here μ and ν), you implicitly sum over all its values. So the right side is really $\sum_{\mu,\nu} g_{\mu\nu}\, dx^\mu\, dx^\nu$.

In flat spacetime, $g_{\mu\nu}$ is the diagonal $\mathrm{diag}(-1, 1, 1, 1)$, and $ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2$. In curved spacetime, the metric varies from place to place.

4.3 Geodesics and Christoffel symbols

A **geodesic** is the curved-spacetime version of a straight line. It's the path that locally extremizes the length functional. In flat space, geodesics are straight lines; on the surface of a sphere, they're great circles; in curved spacetime, they're the trajectories of free-falling objects.

The geodesic equation is

$$\frac{d^2x^\mu}{d\tau^2} \;+\; \Gamma^\mu_{\nu\sigma}\,\frac{dx^\nu}{d\tau}\,\frac{dx^\sigma}{d\tau} \;=\; 0,$$

Read aloud: "d-squared x-to-the-mu, over d-tau-squared, plus Gamma upper-mu lower-nu-sigma, times d-x-to-the-nu over d-tau, times d-x-to-the-sigma over d-tau, equals zero." The Γ *(say: "capital gamma")* symbols are pronounced "Christoffel symbols" *(say: "KRIS-toff-el")*; they're indexed by three Greek letters.

where the $\Gamma^{\mu}_{\nu\sigma}$ are the **Christoffel symbols**, computed from the metric. They tell you how the local basis vectors twist as you move.

In the proof: Lemma 4.12. In the spacetime metric deformed by your presence, every timelike geodesic emanating from my worldline at $t = 0$ ends up at yours. The Christoffel symbols of my life were quietly rewritten that night. Every direction points the same way now.

4.4 The Einstein field equations

The relationship between geometry and matter:

$$G_{\mu\nu} \;=\; \frac{8\pi G}{c^4}\,T_{\mu\nu}.$$

Read aloud: "G sub mu nu equals eight pi big-G over c-to-the-fourth, times T sub mu nu." Note the two different G's: the boldfaced $G_{\mu\nu}$ on the left is the **Einstein tensor** (geometry); the lone G on the right is **Newton's gravitational constant**, a number. Physics sometimes overloads its letters.

The left side, $G_{\mu\nu}$, is a particular combination of the metric and its derivatives (the **Einstein tensor**); it measures the curvature of spacetime. The right side, $T_{\mu\nu}$ *(say: "T sub mu nu")*, is the **stress-energy tensor**; it measures the local density and flux of matter and energy.

In words: matter tells spacetime how to curve, and spacetime tells matter how to move.

In the proof: Lemma 4.12 cites these as the source of the Christoffel symbols. I claim $T_{\mu\nu}$ is dominated by the author's affection for K. That sentence is a joke. The general structure is real.

4.5 The principle of least action

There's an alternative formulation of all of classical mechanics (and a piece of quantum mechanics) that says: given a starting state and an ending state, the trajectory a system actually follows is the one that minimizes a quantity called the **action**:

$$S[\gamma] \;=\; \int_{t_0}^{T} \mathcal{L}(\gamma, \dot{\gamma}, t)\, dt.$$

Read aloud: "S of gamma equals the integral, from t-zero to capital-T, of script-L of gamma, gamma-dot, t, with respect to dt." γ *(say: "gamma, lowercase")* is a curve (a trajectory through time); $\dot{\gamma}$ *(say: "gamma-dot")* is its time derivative (its velocity). $\mathcal{L}$ *(say: "script L; the Lagrangian")* is a function.

Here $\mathcal{L}$ is the **Lagrangian**, a function of position, velocity, and time. The system picks out, from among all kinematically possible trajectories, the one with the smallest action. This is **Hamilton's principle**.

In the proof: Lemma 4.11. Of all the worldlines that could have taken me from my 18-year-old self to today, the actual one is the one that minimizes the action. That worldline passes through Pat & Umar's networking event on May 28, 2025. The meeting was on the stationary path.

4.6 Noether's theorem

Emmy Noether *(say: ""NER-ter," long "e"")*, in 1918, proved one of the deepest results in mathematical physics. *Every continuous symmetry of the action gives rise to a conserved quantity.*

- Time-translation symmetry (the laws of physics don't change over time) implies **energy conservation**.
- Space-translation symmetry implies **momentum conservation**.
- Rotational symmetry implies **angular momentum conservation**.

This single observation unifies most of classical and quantum mechanics. It's why energy is a thing at all.

In the proof: Lemma 4.13. Two symmetries of our joint action are observed: time translation (we are still the same on different mornings) and exchange of G and K (we love each other symmetrically). By Noether, two conserved quantities: the energy of the bond, and the difference in mutual affection. Both are constant in time, both non-zero.

5. Information theory: the math of signal

5.1 Shannon channel capacity

Claude Shannon, in 1948, founded information theory. He gave a precise definition of how much information can be transmitted through a noisy channel. A channel of capacity C bits per use can reliably transmit at most C bits per use, no matter how cleverly you encode.

For a channel where you pick one of N equally likely symbols per use, the capacity is

$$C = \log_2(N).$$

Read aloud: "C equals log base two of N." $\log_2$ *(say: "log base two")* is the logarithm where the base is 2 rather than the usual e or 10.

A coin flip is $\log_2(2) = 1$ bit. A six-sided die roll is $\log_2(6) \approx 2.58$ bits.

In the proof: Lemma 4.7. Each daily wake-up timestamp can be one of 1440 minutes. So one timestamp transmits $\log_2(1440) \approx 10.5$ bits. Over ≈ 330 days, that's $\approx 3{,}465$ bits per direction, $\approx 7{,}000$ bits total. The exact number is less important than that it grows without bound.

5.2 Mutual information

The **mutual information** $I(X;Y)$ *(say: "I of X, semicolon, Y")* between two random variables is a measure of how much knowing one tells you about the other. It's defined as

$$I(X;Y) \;=\; H(X) + H(Y) - H(X,Y),$$

Read aloud: "I of X semicolon Y equals H of X plus H of Y minus H of X comma Y." Here $H(\cdot)$ *(say: "H of...")* denotes the Shannon entropy, a measure of uncertainty.

If X and Y are independent, $I(X;Y) = 0$. If knowing X tells you everything about Y, $I(X;Y) = H(Y)$.

In the proof: Lemma 4.7. The mutual information between us is monotonically increasing. We learn more about each other every day. This information has a unit (bits) and a rate of growth (positive).

6. Machine learning: the math of fitting models

6.1 Loss functions

In machine learning, you have a model with parameters θ *(say: "theta")*, and you want the model to do something well. The **loss function** $\mathcal{L}(\theta)$ *(say: "script L of theta")* measures how badly the model with parameters θ is currently doing. Smaller loss is better.

Training means: find $\theta^* = \arg\min_\theta \mathcal{L}(\theta)$ *(say: "theta-star equals argmin, over theta, of script-L of theta")*. The parameters that make the loss as small as possible. "Argmin" *(say: "ARG-min")* means: the argument (the input value) that achieves the minimum.

In the proof: Lemma 4.4 references this. The loss function of my life has a unique global minimum, and that minimum is achieved at $\theta^* =$ you.

6.2 Stochastic gradient descent and Robbins-Monro

Gradient descent is the canonical optimization algorithm. Start anywhere; at each step, move opposite the gradient of the loss; repeat. The gradient points uphill, so the negative gradient points downhill, so you slowly descend toward a minimum.

Stochastic gradient descent uses noisy estimates of the gradient (computed on random subsets of the data). This is faster, scales better, and adds noise that helps escape some bad minima.

Herbert Robbins and Sutton Monro, in 1951, gave conditions on the step sizes η_t *(say: "eta sub t, "ay-ta"")* that guarantee almost-sure convergence:

$$\sum_t \eta_t \;=\; \infty, \qquad \sum_t \eta_t^2 \;<\; \infty.$$

Read aloud: "The sum, over t, of eta sub t equals infinity. The sum, over t, of eta sub t squared is finite." η_t is the step size at iteration t.

The steps have to be big enough on average to reach the minimum (first condition) but small enough that the noise doesn't accumulate (second).

In the proof: Lemma 4.4. The Robbins-Monro conditions are satisfied empirically. The optimizer of my life converges almost surely to a stationary point, and the stationary point is K.

7. Smaller results

7.1 The reproduction number R_0

In epidemiology, R_0 *(say: "R-naught, or "R sub zero"")* is the average number of new cases produced by a single infected individual in a fully susceptible population. If $R_0 > 1$, the disease spreads. If $R_0 < 1$, it dies out.

Measles has $R_0 \approx 18$. Influenza has $R_0 \approx 2$.

In the proof: Observation 3.5. The reproduction number of your personality strictly exceeds unity. Your warmth is communicable.

7.2 The photopic luminosity function $V(\lambda)$

The human eye has cones (color receptors) that are most sensitive at certain wavelengths of light. The **photopic luminosity function** $V(\lambda)$ *(say: "V of lambda")* describes the eye's sensitivity to each wavelength in normal daylight conditions. It peaks at $\lambda \approx 555$ *(say: "lambda is approximately 555")* nm, a yellow-green color sometimes called chartreuse.

This is why traffic-light green and emergency-vest yellow are the colors they are: they're maximally visible to the human eye.

In the proof: Observation 3.6. Your favorite color is chartreuse, of wavelength $\lambda \approx 555$ nm, which is exactly the peak of $V(\lambda)$. Your aesthetic preferences are calibrated to the band the human visual system evolved to see most clearly.

Closing note

I left some things out.

The proof has bits I didn't explain here. Some are routine (the Extreme Value Theorem follows from compactness plus continuity, things like that). Some are mostly punchlines, where the math is there for the joke and isn't doing real work. Some I just didn't have room for, and you can ask me about them, and I'll explain.

Two things I want to say at the end.

First: it's a real proof. The math results I cited are real results, and where I'm using them straight, I'm using them straight. Banach, Doob, Bell, Bayes, Einstein, Noether, Shannon, all of them are doing actual work in the argument. The framing is romantic; the load-bearing pieces are not.

Second: I picked these fields on purpose. Probability is the math of uncertainty, and I needed to be honest about how unlikely we were. Quantum mechanics is the math of being in superposition, and I needed a language for the parts of us that don't reduce to either of us alone. Relativity is the math of geometry bending, which is what happened in that room when I saw you. Information theory is the math of signal, and the wake-up protocol is the only daily signal I've never missed. Machine learning is the math of fitting models, and I have been training one for almost a year now. Every field is in the proof because something about it was already true about us.

If there's a piece of the proof you want me to read aloud to you, with the math next to my voice, I would love to do that.

– G.