Math Notes

Kieran Harvie

Contents

0.1 Mean and Variance, and the arbitrariness thereof

For some time I have wondered about the arbitrariness around the mean and variance. For example why the arithmetic mean instead of the geometric or root-mean-squared? And why square root the variance to give the standard deviation?

Well the strictness of Markov's and Chebyshev's might provide a reason. Both rely of the conditional expected value, so just to reiterate:

$$E[X|X \geq a] \geq a$$

Since everything X can be is greater then a it's expected value must be greater than a. Notice the strictness of the inequality, this will be used to make the following inequalities much stricter.

Markov

$$\begin{aligned}
\mu =& E[X] \\
=& P(X \leq a)E[X|X \leq a] + P(X \geq a)E[X|X \geq a] \\
\geq& 0 \cdot E[X|X \leq a] + P(X \geq a)a \\
\frac{\mu}{a} \geq& P(X \geq a)
\end{aligned}$$

Chebyshev

$$\begin{aligned}
& E[(X-a)^2] \\
=& P(|X-a| \leq b)E[(X-a)^2||X-a| \leq b] + P(|X-a| > b)E[(X-a)^2||X-a| > b]
\end{aligned}$$

A General Relation

Assume:

$$f(S') \geq 0, \quad g(S) \geq 0$$

Then through:

$$E[f(X)] = P(X \in S)E[f(X)|X \in S] + P(X \in S')E[f(X)|X \in S']$$

We have:

$$1 - \frac{E[g(X)]}{E[g(X)|X \in S']} \leq P[X \in S] \leq \frac{E[f(X)]}{E[f(X)|X \in S]}$$

With dual equality if:

$$f = 1_S, \quad g = 1_{S'}$$

Covariance

Lets try to find the lest squares regression between X and Y such that:

$$E[X] = E[Y] = 0, E[X^2] = E[Y^2] = 1$$

Since the expected values are both zero the line is through the origin

$$\begin{aligned}\sum_n (mx_n + c - y_n)^2 =& nE[(mX + c - Y)^2] \\ =& n\Big(E[m^2X^2] + E[c^2] + E[Y^2] + E[2cmX] + E[-2mXY] + E[-2cY]\Big) \\ =& n(m^2 + c^2 + 1 - 2mE[XY])\end{aligned}$$

Trying to minimize this value by our selection of trivially gets:

$$c = 0, \quad m = E[XY]$$

Just expanding the definitions gives:

$$COV[X, Y] = E[(X - E[X])(Y - E[Y])] = E[XY] = m$$

Hence the covariance can 'naturally' be interpreted and the first order function between the valuables.

$$E[f(X)] \approx E[f_0 + f_1X + f_2X^2/2] = f_0 + \mu f_1 + \sigma^2 f_2/2$$

$$E[f(X)] \approx E[f(\mu) + (X - \mu)f'(\mu) + (X - \mu)^2/2f''(\mu)] = f(\mu) + \frac{f''(\mu)}{2}\sigma^2$$

0.2 Interest Identities

Let P be the principle invested at a rate of r. Consider four different investment scenarios:

- Not invested: $P_0 = P$.
- Fully Invested at the beginning, one instalment at the end:

$$P_1 = (1 + r)P$$

- Continuously invested, continuous installments:

$$P_2 = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{P}{n} \left(1 + \frac{r}{n}\right)^k = \frac{\exp(r) - 1}{r} P$$

- Fully Invested at the beginning, continuous instalments:

$$P_3 = \lim_{n\to\infty} P \left(1 + \frac{r}{n}\right)^n = \exp(x)P$$

Interestingly the relative size of P_2 and P_1 depend on r. P_1 starts on P_2 but switches as r increases.

P_3 is always the best, the proof for P_1 and P_0 are obvious. $P_3 > P_2$ follows from:

$$0 < \int_0^r t \exp(t)\, dt = \left[(t - 1)\exp(t)\right]_0^r = (r - 1)\exp(r) + 1$$

The following interesting identities hold:

$$P_3 = rP_2 + P_0$$

$$P_3 - P = r(P_2 - P) + (P_1 - P)$$

The breaks first neatly breaks P_3 into a nice linear sum. The second does similar for the profit of the investment, total yield minus principle.

0.3 Discontinuities in a Non-decreasing Function

Let f be a non-decreasing function.

Define the jump function J as:

$$J(x) = \inf\{f(t)|t > x\} - \sup\{f(t)|t < x\}$$

This function is well defined since the sets are appropriately bound by $f(x)$. And it is clear that $J(d) \neq 0$ iff d is a discontinuity and that J is non-negative.

Let $U = (x_0, x_1)$. For $d_n \in U$ with $n < m \Rightarrow d_n < d_m$ we have:

$$f(x_1) - f(x_0) \geq \sum_k J(d_k)$$

Proof:

$$\begin{aligned}
&\sum_k J(d_k) \\
=&\inf\{f(t)|t > d_n\} - \sup\{f(t)|t < d_0\} + \sum_k \left[\inf\{f(t)|t > d_{k-1}\} - \sup\{f(t)|t < d_k\}\right] \\
\leq& f(d_n) - f(d_0) + \sum_k \left[f(d_{k-1}) - f(d_k)\right] \\
\leq& f(d_n) - f(d_0) \\
\leq& f(x_n) - f(x_0)
\end{aligned}$$

Let $S_n = \{d \in U|J(d) > \frac{1}{n}(f(x_1) - f(x_0))\}$ From the previous inequality there are at most n elements in S_n. Hence:

$$\{d \in U|J(d) > 0\} = \bigcup_n \left\{d \in U|J(d) > \frac{1}{n}(f(x_1) - f(x_0))\right\}$$

Is countable, hence the number of discontinuities of f on U is countable.

By corollary the discontinuities of a non-decreasing function on $\mathbb{R}$ are countable:

Let f be a non-decreasing function on $\mathbb{R}$. Let $X_n = (n-1, n+1)$, clearly $\mathbb{R} = \bigcup_n X_n$[1]. Assume f has an uncountable number of discontinuities then at least one X_n contains uncountable discontinuities. Otherwise there would be a countable set of countable sets of discontinuities, making them countable. But f being non-decreasing function and having an uncountable number of discontinuities in X_n is a contradiction.

[1]Having them overlap simplify the proof by avoiding literal edge-cases.

0.4 Integrals and Symmetry

0.4.1 Domain Symmetry

Let U be a subset of $\mathbb{R}^n$ and let $\phi : U \to U$ be a function such that:

$$\begin{aligned} \phi(U) &= U \\ |\det \phi'(\mathbf{u})| &= 1 \end{aligned}$$

Basically ϕ is a linear permutation[2] on U, this is a symmetry in the most direct sense. We obtain the following:

$$\begin{aligned} \int_U f(\mathbf{v})\, d\mathbf{v} &= \int_{\phi(U)} f(\mathbf{v})\, d\mathbf{v} \\ &= \int_U f(\phi(\mathbf{u}))|\det \phi'(\mathbf{u})|\, d\mathbf{u} \\ &= \int_U f(\phi(\mathbf{u}))\, d\mathbf{u} \end{aligned}$$

In particular we get:

$$0 = \int_U \big(f(\mathbf{u}) - f(\phi(\mathbf{u}))\big)\, d\mathbf{u}$$

This integral is important since a function can be split into a vanishing and non-vanishing part:

$$\begin{aligned} f(\mathbf{u}) =& \frac{1}{2}\big(f(\mathbf{u}) + f(\phi(\mathbf{u})) + \frac{1}{2}\big(f(\mathbf{u}) - f(\phi(\mathbf{u}))\big) \\ \int_U f(\mathbf{u})\, d\mathbf{u} =& \frac{1}{2}\int_U \big(f(\mathbf{u}) + f(\phi(\mathbf{u}))\, d\mathbf{u} + \frac{1}{2}\int_U \big(f(\mathbf{u}) - f(\phi(\mathbf{u}))\big)\, d\mathbf{u} \\ =& \frac{1}{2}\int_U \big(f(\mathbf{u}) + f(\phi(\mathbf{u}))\, d\mathbf{u} \end{aligned}$$

[2]Note that ϕ being a permutation requires that if the magnitude of the determinate is constant it must be unity, this can be seen by setting f to a constant.

For example, consider the classic odd function on an integral centered at 0.

$$\phi(x) = -x$$
$$U = [-1, 1]$$

We get the familiar:

$$\int_{-1}^{1} f(x)\,dx = \frac{1}{2}\int_{-1}^{1} \big(f(x) + f(-x)\big)\,dx$$

The utility of this relation can be seen by applying it to the basis of a class of function. Let $V = \langle 1, x, x^2 \rangle$, this is a basis for all parabolas. Notice that x base element vanishes, simplify the evaluation of integrals.

For a 2-D example recall that for two dimensional change of variables:

$$(x, y) = \phi(u, v)$$

We have:

$$|\det \phi'(\mathbf{v})| = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

The rotation symmetry for a regular triangle is:

$$(x, y) = \frac{1}{2}(-u - \sqrt{3}v, \sqrt{3}u - v)$$

Obviously the symmetries act like a group and with functions being a vector space we can use group representations.

0.4.2 Function Symmetry

Let f and ϕ be functions such that:

$$f(t) = \phi'(t) f(\phi(t))$$

Then for arbitrary x_0 and x_1 we have:

$$\begin{aligned}
\int_{x_0}^{x_1} f(t)\,dt &= \int_{\phi(x_0)}^{\phi(x_1)} \phi'(t) f(\phi(t))\,dt \\
&= \int_{\phi(x_0)}^{\phi(x_1)} f(t)\,dt \\
&= \int_{x_1}^{\phi(x_1)} f(t)\,dt + \int_{\phi(x_0)}^{x_1} f(t)\,dt \\
\int_{x_0}^{\phi(x_0)} f(t)\,dt &= \int_{x_1}^{\phi(x_1)} f(t)\,dt
\end{aligned}$$

Hence the integral value is independent of x_n, in particular if ϕ has a fixed point then the integral is zero.

0.5 Lagrange Multiplier

Recall that local extrema x of the function $f : \mathbb{R}^n \to \mathbb{R}$ subject to contrasts g_i satisfy:

$$\nabla f(x) = \sum_i \lambda_i \nabla g_i(x)$$

The core observation is that if ∇f has a component outside the span of ∇g_i then you can move in that direction while keeping g_i's constant, contradicting the point being an extrema.

But the actual constants λ_i have a useful interpretation as the rate the value of f at the extrema changes as the constant g_i changes. To see this pick a particular g_j and construct a d such that:

$$d \cdot \nabla g_i = D\delta_{i,j}$$

You can achieve this by iteratively removing components in some matter like the following:

$$d_0 = \nabla g_0,\ d_{n+1} = d_n - d_n \cdot \nabla g_n$$

Now scale d down such that functions around the extrema can be approximated through targets[3]. We have:

$$\begin{aligned}
g_i(x+d) =& g_i(x) + d \cdot \nabla g_i(x) \\
=& g_i(x) + D\delta_{i,j} \\
f(x+d) =& f(x) + d \cdot \nabla f(x) \\
=& f(x) + \sum_i \lambda_i d \cdot \nabla g_i(x) \\
=& f(x) + D\lambda_j \\
\nabla f(x+d) =& \nabla(f(x) + D\lambda) \\
=& \nabla f(x) \\
=& \sum_i \lambda_i \nabla g_i(x) \\
=& \sum_i \lambda_i \nabla \big(g_i(x+d) - D\delta_{i,j}\big) \\
=& \sum_i \lambda_i \nabla g_i(x+d)
\end{aligned}$$

[3]Those so inclined are free to chase $\epsilon - \delta$'s

From the these equation we can see that $x + d$ satisfy the requirement to be an extrema. We can also see that a change of D in g_i created a change of $\lambda_j D$ in the value at the extrema, hence giving a rate of change of λ_j.

To-Do: Add and example of minimizing height when the two contrasts are parabolic and linear. (You will need to use logs to get the change for the parabola to be a change in it's width and not height.)

0.6 Isosets

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ let the isosets[4] be sets S in the domain of f such that $f(S)$ is constant.

Given some isoset S and point $x \in \mathbb{R}^n$ of f we want some kind of function that returns some type of measure $d \in \mathbb{R}$ of the distance of x to S. This function will be used in fragment shader rendering, which is why the mission statement is so vague. We only need some general measure since it's better to efficiently get that measure and tweak coefficients then to get something perfectly accurate.

A particular application is the $n = 2$ case where we are looking to find the distance for the contour line.

0.6.1 Naïve Solution

The first idea is to compare the distance of $f(x)$ to $f(S)$ to d:

$$d = |f(x) - f(S)|$$

The problem with this solution is that the measure changes as a function of $|\nabla f(x)|$. That is to say that the faster f changes at x the closer x has to be to S to get the same value of d.

This method might work for some shader effects but not others. For example it won't work to draw a constant width contour as the width would be inversely proportional to $|\nabla f(x)|$

0.6.2 Better Solution

If the underestimation is proportional to $|\nabla f(x)|$ the obvious solution is to divide by $|\nabla f(x)|$:

$$d = \frac{|f(x) - f(S)|}{|\nabla f(x)|}$$

This is the solution currently used but deserves more analysis.

[4]Not the proper name, but I can't recall the proper name right now.

For starters consider the case that x is near the point $s \in S$. Then $\nabla f(s)$ points away from S, that is that if a tangent to S exists at s then $\nabla f(s)$ is at orthogonal to the tangent, by definition of S being a set such that f is constant. The combination of orthogonality and closeness lets us recover the original expression by use of the tangent surface:

$$\begin{aligned} f(x) =& f(s) + (x-s) \cdot \nabla f(s) \\ |f(x) - f(s)| =& |(x-s) \cdot \nabla f(s)| \\ =& |x-s||\nabla f(s)| \\ \frac{|f(x) - f(s)|}{|\nabla f(s)|} =& |x-s| \end{aligned}$$

Now consider the case where x is not close to S and the use case of drawing a constant width contour line. Under what conditions do we avoid a false positive? (That is the function thinks x is closer than it is.) Well we need some constraints on the rate at which $\nabla f(x)$ can grow. To see this consider a point far away from S but whose rate of change is very slow between most of x and s, so that $|f(x) - f(s)|$ is small, but suddenly increases at x, such that $|\nabla f(x)|$ is large. This causes their ratio to be small despite $|x - s|$ being large, false saying x should be colored as part of the contour line. If we reverse the set up, rate of change is large at first then slow, we will get a false negative. (That the point is further than we think it is)

To see how a constraint would be useful, consider the following one:

$$|\nabla f(x+d)| \leq k|d||\nabla f(x)|$$

$$
\begin{aligned}
|f(x) - f(s)| &= \left| \int_0^1 \nabla f(x + (s-x)t) \cdot (s-x)\, dt \right| \\
&= \left| (s-x) \cdot \int_0^1 \nabla f(x + (s-x)t)\, dt \right| \\
&\leq |(s-x)| \left| \int_0^1 \nabla f(x + (s-x)t)\, dt \right| \\
&\quad \text{Cauchy-Schwartz} \\
&\leq |(s-x)| \int_0^1 |\nabla f(x + (s-x)t)|\ dt \\
&\quad \text{ML Bound} \\
&\leq |(s-x)| \int_0^1 k|(s-x)t||\nabla f(x)|\, dt \\
\frac{|f(x) - f(s)|}{|\nabla f(x)|} &\leq \frac{k}{2}|s-x|^2
\end{aligned}
$$

This avoids a false negative as x must be at least $\sqrt{\frac{2d}{k}}$ away.

To-do: we need a inequities like $d \geq p(|x-s|)$ to get a bound on false positives.

The condition:

$$|\nabla f(x+d)| \leq k|d| + |\nabla f(x)|$$

Gives:

$$d \leq |x-s| \left(1 + \frac{|x-s|}{|\nabla f(x)|}\right)$$

0.7 Padé Approximant

We wish to approximate a function f by creating a rational function that agrees with f's first N derivatives at zero.

Let T_N be the Nth degree Maclaurin series of f. Consider the steps of finding the polynomial greatest common division by the extended Euclid algorithm of T_N with x^{N+1} where we prematurely stop:

$$\begin{aligned} x^{N+1} &= 1 \cdot x^{N+1} + & 0 \cdot T_N(x) \\ T_N(x) &= 0 \cdot x^{N+1} + & 1 \cdot T_N(x) \\ r_1(x) &= 1 \cdot x^{N+1} + & -q_1(x) \cdot T_N(x) \\ &\vdots & \\ P(x) &= K(x)x^{N+1} + & Q(x)T_N(x) \end{aligned}$$

By inspection we get the useful relation:

$$P(x)/Q(x) \equiv T_N(x) \mod x^{N+1}$$

Hence satisfying the original objective. Note that we can chose decrease the degree of P by simply continuing the algorithm.

0.7.1 The Reverse

Say I want to do the reverse, that I have $f = g/h$ where I have the power series for g and h but want to effectively find f We have:

$$\begin{aligned} x^{N+1} &= 0 \cdot f(x) + & 1 \cdot x^{N+1} \\ g(x) &= h(x) \cdot f(x) + & 0 \cdot x^{N+1} \end{aligned}$$

You can remove some higher term of f into a residue function K on x^{N+1}, since we don't really care about it.

0.7.2 Differential

What if g and h are related buy a differential equation? We can use it like how we used the quotient-remainder equation. The derivative is linear after all.

0.7.3 Chinese Remainder Theorem

Since I have modulo relations can I combine them with the Chinese Remainder Theorem? The moduli will have to be pairwise coprime $(x - a_i)^n$ stand out.

0.8 p-adic numbers

0.8.1 Valuation

A function v of a field is called a valuation if:

$$\begin{aligned} v(x) =& \infty \quad \text{iff } x = 0 \\ v(xy) =& v(x) + v(y) \\ v(x+y) \geq& \min(v(x), v(y)) \quad \text{with equality if } v(x) \neq v(y) \end{aligned}$$

(Note that the codomain is only required to be an abelian totally ordered group extended with ∞, But I will treat is as the natural numbers.)

There are three immediate corollaries of the definitions.
1:By induction on the inequality we have:

$$v\left(\sum_k x_k\right) \geq \min\left(\bigcup_k \{v(x_k)\}\right)$$

2: By setting $x = y = 1$ in the second equality we have $v(1) = 2v(1)$ and hence $v(1) = 0$.
3: By setting $xy = 1$ in the second equality we have $v(1/x) = -v(x)$.

0.8.2 p-adic numbers

Interestingly, the amount of times a prime p divides a rational number x is a valuation. Let $x = p^n \frac{a}{b}$ where a and b are coprime, then $v_p(x) = n$ is a valuation. (Assuming you set $v_p(0) = \infty$). This is easy to prove, if you need help remember that for a prime p we have: $p|ab \Rightarrow p|a$ or $p|b$.

The reason this is interesting is because $|x|_p = p^{-v_p(x)}$ is a metric on $\mathbb{Q}$. Meaning we can make a new field $\mathbb{Q}_p$ by taking all the limits in $\mathbb{Q}$ as we would nomarlly do to make $\mathbb{R}$ with $|\cdot|$. And through something called Ostrowski's theorem becomes a lot more motivated and less arbitrary way to complete $\mathbb{Q}$.

But the valuation alone also provides two cool proofs that simplify previous proofs.

0.8.3 Irrationality of $\sqrt{2}$

Consider the equation:

$$x^2 = 2$$

Valuating both sides gives:

$$2v_2(x) = 1$$

But $v_2(x)$ is an integer for all rational x hence the LHS is always even but the right is odd. Hence there is no x satisfying the equation and $\sqrt{2}$ is irrational.

0.8.4 Valuation of the Harmonic Numbers

The valuation of harmonic numbers is given by $v_2(H_n) = -\lfloor \log_2(n) \rfloor$.

Lemma: Let k be the power of the largest power of 2 less than n, i.e. $k = \lfloor \log_2(n) \rfloor$. Let S be the set $[1, n]$ excluding 2^k, then from the maximality of k we have:

$$\max(v_2(S)) \leq k - 1$$

Since if we assume there is an $s \in S$ such that $v_2(s) > k-1$ with the codomain of v_2 being integers means $v_2(s) \geq k$. This means there exists an integers a and b coprime to each other and 2 such that $s = 2^k \frac{a}{b}$. s being a positive integer means $b = 1$. $a \neq 1$ since it would make $s = 2^k$, which was excluded from S. But $a \geq 2$ would means $2^{k+1} \in S$, contradicting the maximality of k.

Now H_n can be written as the following sum:

$$H_n = \sum_{s \in S} \frac{1}{s} + 2^{-k}$$

Where the valuation of first term is bound by:

$$\begin{aligned} v_2\left(\sum_{s \in S} \frac{1}{s}\right) &\geq \min\{v_2(1/s) \mid s \in S\} \\ &= \min\{-v_2(s) \mid s \in S\} \\ &= -\max\{v_2(s) \mid s \in S\} \\ &= -k + 1 \end{aligned}$$

Hence the valuations are not equal since:

$$v_2(2^{-k}) = -k < -k + 1 \leq v_2\left(\sum_{s \in S} \frac{1}{s}\right)$$

Hence

$$v_2(H_n) = \min\left\{v_2\left(\sum_{s \in S} \frac{1}{s}\right), v_2(2^{-k})\right\} = -k$$

As required.

Note that for $n \geq 2$ we have $k \geq 1$ and hence $v_2(H_n) \leq -1$ meaning H_n isn't an integer for $n \neq 1$

0.8.5 Valuation of the Harmonic Numbers v2

I think the lemma's of the previous proof can be separated and cleaned up.

Lemma: Let S be a subset of the v's domain with an element $s_0 \in S$ such that:

$$v(S \backslash \{s_0\}) > v(s_0)$$

Then:

$$v\left(\sum_{s \in S} s\right) = v(s_0)$$

Proof: Plugging the inequality into the valuation inequality axiom gives:

$$v\left(\sum_{s \in S \backslash \{s_0\}} s\right) \geq v(S \backslash \{s_0\}) > v(s_0)$$

Hence the term of the far left and far right are not equal meaning:

$$\begin{aligned} v\left(\sum_{s \in S} s\right) =& v\left(s_0 + \sum_{s \in S \backslash \{s_0\}} s\right) \\ =& \min\left\{v(s_0), v\left(\sum_{s \in S \backslash \{s_0\}} s\right)\right\} \\ =& v(s_0) \quad \square \end{aligned}$$

Lemma: If $n \in \mathbb{N}$ then $v_p(n) \leq \log_p(n)$ with equality iff n is a power of p.

Proof: Equality in the case of n being a power is trivial. Now consider n not a power of 2 meaning $n = a\, p^{v_p(n)}$ where $a > 1$, hence:

$$p^{v_p(n)} < n = p^{\log_p(n)}$$

$\log_p$ is strictly monotonic, hence:

$$v_p(n) < \log_p(n)$$

Lemma: If $n \in \mathbb{N}$ then $v_2(n) \leq \lfloor \log_2(n) \rfloor$ with equality iff n is a power of 2.

Proof: Equality in the case of n being a power is trivial. Assume n isn't a power of 2 then:

$$n \geq 3 \cdot 2^{v_2(n)}$$

Hence:

$$\begin{aligned} \log_2(n) &\geq \log_2(3) + v_2(n) \\ \log_2(n) - \log_2(3) &\geq v_2(n) \\ \lfloor \log_2(n) - \log_2(3) \rfloor &\geq \lfloor v_2(n) \rfloor \end{aligned}$$

Since $\log_2(3) > 1$ we have:

$$\lfloor \log_2(n) \rfloor > \lfloor \log_2(n) - \log_2(3) \rfloor$$

And $v_2(n)$ is an integer, hence: $\lfloor \log_2(n) \rfloor > v_2(n)$ $\square$

Note that this can't be generalized to larger p since the logarithm being less than one isn't guaranteed. Compare with $p = 3$ with $v_3(6) = 1 = \lfloor \log_3(6) \rfloor$ because $\log_3(2) < 1$. Note to the previous note, in this case you can say $v_3(n) \leq \lfloor \log_p(n) \rfloor$ iff n is a power of 3 or even, which is still *kind of* cool.

Lemma: Let $S_n = \{k^{-1} | 1 \leq k \leq n\}$ and $k_0 = \lfloor \log_2 n \rfloor$. Then $2^{-k_0} \in S$ and satisfies:

$$v_2(S_n \backslash \{2^{-k_0}\}) > v_2(2^{-k_0})$$

Proof: $2^{k_0} = 2^{\lfloor \log_2 n \rfloor} \leq 2^{\log_2 n} = n$ hence $2^{-k_0} \in S$. Now consider $s \in S$ then $s^{-1} \in \mathbb{N}$ which from the previous lemma means:

$$v_2(s^{-1}) \leq \lfloor \log_2(s^{-1}) \rfloor \leq \lfloor \log_2(n) \rfloor = k_0$$

With equality iff s^{-1} is a power of 2. It's can't be a larger power than 2^{k_0} since $2^{\lfloor \log_2(n) \rfloor}$ is the largest power less than or equal to n. But from the definition of $S_n \backslash \{2^{-k_0}\}$ it can't be the same power. Hence s^{-1} can only be a lower power making the final inequality strict anyway. Hence:

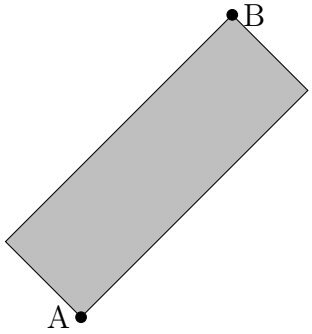
$$v_2(s^{-1}) < k_0 \Rightarrow v_2(s) > -k_0 = v_2(2^{-k_0}) \quad \square$$

Noting that $H_n = \sum_{s \in S_n} s$ means the proof that $v_2(H_n) = -\lfloor \log_2(n) \rfloor$ follows immediately from the first and last lemmas.

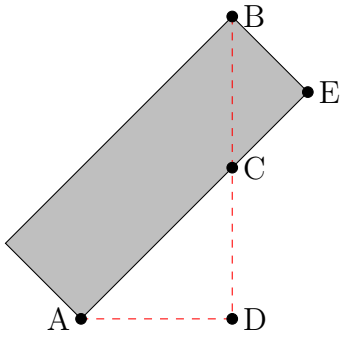
0.9 Causal Metric

0.9.1 Basic Geometry

Consider two points A and B that we wish to find the area of the rectangle between them in terms of their coordinates on an axis at a 45° angle.



Drop an altitude from B to line up horizontally with A and label the end-point D. The length |AD| and |BD| are the coordinates.



The triangles $\triangle$ACD and $\triangle$BCE are $45° - 90° - 45°$ triangles meaning:

$$\begin{aligned}
|\text{AC}| &= \sqrt{2}|\text{AD}| \\
|\text{BE}| &= \frac{1}{\sqrt{2}}|\text{BC}| \\
&= \frac{1}{\sqrt{2}}(|\text{BD}| - |\text{CD}|) \\
&= \frac{1}{\sqrt{2}}(|\text{BD}| - |\text{AD}|)
\end{aligned}$$

Hence the (signed) area of the rectangle is:

$$\begin{aligned}|\mathrm{BE}| \cdot |\mathrm{AE}| =& |\mathrm{BE}| \cdot (|\mathrm{AC}| + |\mathrm{CE}|) \\ =& (|\mathrm{BD}| - |\mathrm{AD}|)(|\mathrm{BD}| + |\mathrm{AD}|) \\ =& |\mathrm{BD}|^2 - |\mathrm{AD}|^2\end{aligned}$$

0.9.2 Causal Metric

This construction supplies some intuition for Minkowski Metric. Since if we interpret the plane as the set of events where the horizontal component is space-like and the vertical is time-like. The reason the rectangle is at a 45° is because that's the maximum speed of propagation. The rectangle's area is a measure of the amount of events in-between the two. And the sign of the area is the type of causal connection. Wether the points are in the way (space-like), or another events are a means by which the earlier effect the later (time-like).

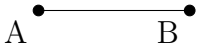
0.10 Quick Summary of Spaces

A space is a collection of elements, often called points, with some additional structure. There are four main types of spaces that form a nice hierarchy, that is that some types of spaces are always a subtype of other space.

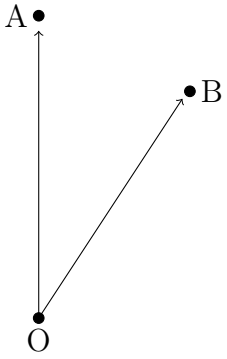
0.10.1 Hierarchy

Topological Space: Neighbourhood In a topological space the elements have a concept of "Neighbourhood", that is the points which are adjacent to each point.

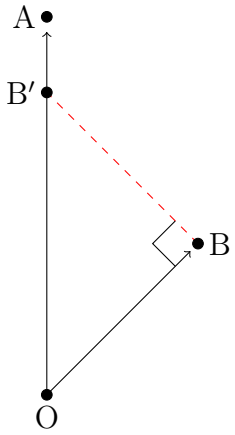
Metric Space: Distance In a metric space the element have a concept of distance to each other. You can induce a neighbourhood by saying all points within a certain threshold distance to a point are in the neighbourhood of that point.



Normed Space: Length In a normed space there is a concept of length of a point. Shown here as their distance to some origin. A distance can be induced by getting the length of an arrow going from A to B.



Inner-product Space: Projection An inner product space has a concept of projecting one element on another. You can induce a length by projecting an object onto itself.



0.10.2 Euclidean

Euclidean space is a normed space on $\mathbb{R}^n$ where the inner product is given by $\sum_k A_k B_k$. This naturally induces a metric on $\mathbb{R}^n$.

The reason we care about the metric of Euclidean space is because all normed spaces have a form of the Pythagorean theorem. Making metric space the most endowed space we can easily start thinking about non-Euclidean space.

This sucks because most talk about non-Euclidean space talks about "parallel lines" which makes one naturally think about angles and hence an inner product. But no, instead we have the concept of a geodesic. A geodesic is a curve that is locally distance minimizing. This locality comes from the topology, and means that for all points $\gamma(t_1)$ and $\gamma(t_0)$ on the curve γ such that they are in the same neighbourhood satisfy:

$$d(\gamma(t_1), \gamma(t_0)) = k|t_1 - t_2|$$

0.10.3 non-Euclidean

Quick rundown of some attempts of non-Euclidean geometries:

Pseudo-Euclidean: Like how we defined Euclidean Space was defined with an inner product we define a new structure with a different quadratic form.

Riemann Manifold: A Manifold is a space that is "locally Euclidean". And a Riemann Manifold is one where this inner-product in the local Eu-

clidean space is always positive.

Pseudo-Riemann Manifold: Like A Riemann Manifold but the inner-product is only required to be non-degenerate. I think this includes Pseudo-Euclidean space, but haven't put much thought into it.

0.11 Wave Equation

0.11.1 Symmetry

Given some function $u(\vec{x}, t)$ on space and time we want to understand it's dynamics under the following assumptions:

1. **Reversible**, the dynamics should be symmetric in respect to reversing time.
2. **Isotropic**, the function should be symmetric in respect to all directions.
3. **Relative**, the absolute value doesn't matter, only changes.
4. **Perturbation**, the changes should be small.

The wave equation is a natural conclusion from from these constraints:

$$\frac{\partial^2}{\partial t^2} u = c \sum_i \frac{\partial^2}{\partial x_i^2} u$$

Because the change change is small we start by considering a Taylor expansion and try to get the lowest order terms. Because it's relative we ignore the 0^{th} order terms. Because it's reversible we ignore the 1^{st} order terms Giving:

$$\frac{\partial^2}{\partial t^2} u = \sum_{i,j} w_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} u$$

Because it's isotropic all the $\frac{\partial^2}{\partial x_i^2}$ coefficient need to be equal. And without loss of generality we can assume a basis where the cross terms vanish, giving:

$$\frac{\partial^2}{\partial t^2} u = \sum_i cu \frac{\partial^2}{\partial x_i^2} = c \sum_i \frac{\partial^2}{\partial x_i^2} u \quad \square$$

0.11.2 Old Symmetry Attempt

(Not sure what I was doing here, lol. Seems like I went on the wrong path by not including t) Lets start with with a function of space $u(\vec{x})$. Lets assume it is continuous such that:

$$u(\vec{x} + \vec{r}) = u(\vec{x}) + \sum_i r_i \frac{\partial}{\partial x_i} u(\vec{x}) + \frac{1}{2} \sum_{i,j} r_i r_j \frac{\partial^2}{\partial x_i \partial x_j} u(\vec{x})$$

Lets remove all terms that aren't reversible:

$$u(\vec{x}+\vec{r}) = u(\vec{x}) + \frac{1}{2}\sum_i r_i^2 \frac{\partial^2}{\partial x_i^2} u(\vec{x})$$

Lets impose localization:

$$c^2 r_0^2 = \sum_{i>0} r_i^2$$

Not sure exactly how this part works but we get the sign we need one example is to only keep terms that have $c^2 r_0^2 = r_i^2$

0.11.3 1-D

I need to remind myself how to solve the 1-D case as a stepping stone for something else. Reminder that the 1-D form, with unitary speed, is:

$$\frac{\partial^2}{\partial t^2} u = \frac{\partial^2}{\partial x^2} u$$

Observe that plane waves where angular frequency and wave number have the same magnitude solve this:

$$\exp(ik(x \pm t))$$

This suggests the use of Fourier transform, where the $\pm$ is used to fit the solution to initial value and derivative:

$$u = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (a_+(k) e^{ikt} + a_-(k) e^{-ikt}) e^{ikx}\, dk$$

Assume $f(x) = u(x,0)$ and $h(x) = \left[\frac{\partial}{\partial t} u\right](x,0)$. Equating Fourier transforms gives:

$$\begin{aligned} \hat{f}(k) =& a_+(k) + a_-(k) \\ \hat{h}(k) =& ik(a_+(k) - a_-(k)) \end{aligned}$$

Hence:

$$u = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\hat{f}(k)\cos(kt) + \hat{h}(k) k^{-1} \sin(kt)) e^{ikx}\, dk$$

This can be processed further through the convolutions theorem. (Because of the table of transforms I have it will be easier sin is turned into normalized sinc and rect cuts off at $\frac{1}{2}$):

$$u = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \left(\hat{f}(k) \cos(kt) + \hat{h}(k) t \operatorname{sinc}\left(\frac{kt}{\pi}\right) \right) e^{ikx} \, dk$$

$$\begin{aligned}
u =& \left[f(\tau) * \frac{1}{2}(\delta(\tau - t) + \delta(\tau + t)) \right](x) + \frac{\operatorname{sgn}(t)}{2} \left[h(\tau) * \operatorname{rect}\left(\frac{\tau}{2t}\right) \right](x) \\
=& \frac{1}{2}(f(x-t) + f(x+t)) + \frac{\operatorname{sgn}(t)}{2} \left[h(\tau) * \operatorname{rect}\left(\frac{\tau}{2t}\right) \right](x) \\
=& \frac{1}{2}(f(x-t) + f(x+t)) + \frac{\operatorname{sgn}(t)}{2} \int_{x-t}^{x+t} h(\tau) \, d\tau
\end{aligned}$$

Which is such a cool equation! It shows the reversibility and isotropic nature of the solution but the evenness of x and t. It shows the limit on how fast information travels by the convolution with rect.

I don't really like the sgn function there and I'm not convinced it isn't some kind of error. So ignoring it for the following bit (it's mostly ± 1 anyway), see how we can recover the conditions on f and h:

$$\begin{aligned}
u(x,0) =& \frac{1}{2}(f(x-0) + f(x+0)) + \frac{1}{2} \int_{x-0}^{x+0} h(\tau) \, d\tau \\
=& f(x) \\
\left[\frac{\partial}{\partial t} u \right](x,t) =& \frac{1}{2} - (f(x-t) + f(x+t)) + \frac{1}{2}(h(x-t) + h(x+t)) \\
\left[\frac{\partial}{\partial t} u \right](x,0) =& h(x)
\end{aligned}$$

It almost seems obvious! Like I should have gone straight to this and not bothered with the Fourier stuff. At least I learn something, I guess.

0.12 Unit Fraction

Theorem: Consider a function $f : X \to \mathbb{R}$ where $\mathrm{cl}(\mathrm{Im}(f))$ is bounded and countable. Then for any $n \in \mathbb{N}$ and interval $U \subset \mathbb{R}$ we have $n\,\mathrm{Im}(f) \not\supseteq \mathbb{Q} \cap U$. [5]

Proof: If we assume that $n\,\mathrm{Im}(f) \supseteq \mathbb{Q} \cap U$ we get:

$$|\,\mathrm{cl}(n\,\mathrm{Im}(f))| \geq |\,\mathrm{cl}(\mathbb{Q} \cap U)| = |U| = \aleph_1$$

However by the compactness of $\mathrm{cl}(\mathrm{Im}(f))$:

$$\mathrm{cl}(n\,\mathrm{Im}(f)) = n\,\mathrm{cl}(\mathrm{Im}(f))$$

Hence by the countability of $\mathrm{cl}(\mathrm{Im}(f))$:

$$|\,\mathrm{cl}(n\,\mathrm{Im}(f))| = |n\,\mathrm{cl}(\mathrm{Im}(f))| \leq |\,\mathrm{cl}(\mathrm{Im}(f))|^n = \aleph_0^n = \aleph_0$$

This contradicts $|\,\mathrm{cl}(n\,\mathrm{Im}(f))| \geq \aleph_1$ and hence $n\,\mathrm{Im}(f) \not\supseteq \mathbb{Q} \cap U$.

Corollary: The function $f : \mathbb{N}_{>0} \to \mathbb{R}$ where $f(x) = 1/x$ meets the function requirements. Hence for every interval there is a rational number that can't be represented with n unit fractions.

I want to try a more aesthetically pleasing formulation by separating out the set theory from the topology from the $\mathbb{Q}$ specifics.

Theorem: Let X and Y be sets. Then $|Y| > |X|^n$ implies $nX \not\supseteq Y$.
Proof: By contradiction on set size.

Theorem: If X is countable and Y is non-countable then $nX \not\supseteq Y$ for all $n \in \mathbb{N}_{>0}$.
Proof: The previous theorem using $\aleph_1 > \aleph_0^n$ for all $n \in \mathbb{N}_{>0}$

Theorem: If X is compact, $\mathrm{cl}(X)$ countable, and Y non-countable then $\mathrm{cl}(nX) \not\supseteq Y$ for all $n \in \mathbb{N}_{>0}$.
Proof: The previous theorem using $\mathrm{cl}(nX) = n\,\mathrm{cl}(X)$ from compactness.

Now just rework the corollary a bit for it to fit here.

[5] In this section nS means $\{\sum_i s_i | s \in S^n\}$ and not $\{ns | s \in S\}$

0.13 Winquist's identity

The following form of the Winquist's identity was on an old hard drive:

$$\begin{aligned}
&\prod_{n\geq 1}(1-ax^{n-1})(1-a^{-1}x^n)(1-bx^{n-1})(1-b^{-1}x^n)\\
&\quad\times(1-ab^{-1}x^{n-1})(1-abx^{n-1})(1-a^{-1}b^{-1}x^n)(1-x^n)^2\\
&=\sum_{i\in\mathbb{N}_{\geq 0}}\sum_{j\in\mathbb{Z}}(-1)^{i+j}(b^{-3j}-b^{3j+1})(a^{-3i}-a^{3i+3})\\
&\quad\times(b^{-3i+2}-b^{3i-1})(a^{-3j+1}-a^{3j+2})x^{\frac{j(3j+1)}{2}+\frac{3i(i+1)}{2}}
\end{aligned}$$

A quick Google make the identity verifies that the identity looks right. With the exception of the ranges of the RHS summation, maybe i and j should be switched.

Either way this is a cool identity and captures something interesting. And that would definitely be helpful in generating functions. In particular reciprocating the relation and multiplying by the RHS give a weighted partition that has a relatively sparse recursive relation.

0.14 XOR hash

Today I was presented with the following problem:

Given an array with all integers $[1, 100]$ with a single integer removed. Assuming memory is limited, how do you efficiently determine the missing integer?

The solution is to XOR all the elements of the array together. Then XOR this with the known value for ALL the numbers between $[1, 100]$. The result will be the missing number.

This problem seemed like a good introduction to hashes, like Zobrist, that XOR a bunch of info together. The benefit of this approach is that the components can be added/removed by a XOR:

$$h(\{a_0, a_1\}) = h(\{a_0\}) \,\mathrm{XOR}\, h(\{a_1\})$$

0.14.1 Determining the constant

It turns out its easy to calculate the constant for the original question by hand. First define T as:

$$T(n) = n \,\mathrm{XOR}\, T(n-1), \quad T(1) = 1$$

Then:

$$T(n) = \begin{cases} n & n \mod 4 = 0 \\ 1 & n \mod 4 = 1 \\ n+1 & n \mod 4 = 2 \\ 0 & n \mod 4 = 3 \end{cases}$$

This can easily be proved by induction and remembering that for even n we have:

$$n \,\mathrm{XOR}\, 1 = n + 1$$

Because of the modularity you would expect there to be an expression for T involving powers of the fourth root of unity (i):

$$T(n) = \frac{1}{2} + \frac{1}{2}(1 + (-1)^n) + \frac{1}{4}((i-1)(-i)^n - (i+1)i^n)$$

(Done on scrap paper, not double checked, close enough to see the form).

While the hybrid function is probably a more useful form the last one involves complex numbers in a way I didn't expect.

0.14.2 Homomorphic Hashing

Those with mathematic training will notice that the property that make this approach work:

$$h(\{a_0, a_1\}) = h(\{a_0\}) \operatorname{XOR} h(\{a_1\})$$

Has the same form as a homomorphism[6] and would investigate homomorphic hash functions.

Well the most influential definition of such a function happened in 2004 Krohn, Freedman and Mazieres proposed a definition of homomorphic hash function as function $H : V \to G$ such that:

- H is collision resistant. Meaning we are unlikely to find x and y such that $H(x) = H(y)$.
- H is a homomorphism. Meaning $H(x+y) = H(x)+H(y)$ for all x and y.

The main problem is defining V that is compatible with the regiments on H. Since we want to use the set union operator then the natural choice is $V = 2^S$ for some set S. But our hash only has the same form as a homomorphism[7] when the input sets are mutually exclusive. In general we have:

$$H(S_0 \cup S_1) = H(S_0) \operatorname{XOR} H(S_1) \operatorname{XOR} H(S_0 \cap S_1)$$

We can make a collision resistant hash though. Make H uniformly distributed on $\mathbb{F}_2^{\log_2 |S|}$ for S. Then by induction and $x \operatorname{XOR} y = 0$ iff $x = y$ you can show that it stays uniform for 2^S. Meaning the chance of a collision is $\frac{1}{|S|}$

(This is actually a bound for when $|S|$ is a power of 2, but you can figure out the rest)

Homomorphic functions are still useful in cryptography though.

[6] "same form" as a "homomorphism", math pun intended

[7] haha

0.15 Rational Tangent

Consider the following relation:

$$\begin{aligned}\cos(n\phi) + i\sin(n\phi) =& (\cos(\phi) + i\sin(\phi))^n \\ =& (\cos(\phi)(1 + i\tan(\phi)))^n \\ =& \cos(\phi)^n(1 + i\tan(\phi))^n\end{aligned}$$

The product of two complex numbers with rational real and imaginary part is a complex number with rational real and imaginary part.

To see this observe that we only use multiplication, addition, and subtraction to get the real and imaginary components of the product from the component of the factors and that these operations between from rational numbers produce rational numbers.

Hence, if $\tan(\phi)$ is rational the there exists some rational numbers p, q such that:

$$(1 + i\tan(\phi))^n = p + qi$$

Substituting this into the original formula:

$$\cos(n\phi) + i\sin(n\phi) = \cos(\phi)^n(p + qi)$$

And equating real and imaginary components gives:

$$\tan(n\phi) = \frac{\sin(n\phi)}{\cos(n\phi)} = \frac{\cos(\phi)^n p}{\cos(\phi)^n q} = \frac{p}{q}$$

Hence $\tan(\phi)$ being rational implies $\tan(n\phi)$ is as well.

I think this result was meant to be part of a larger argument, but I have forgotten what the larger one is. One point that I think will be relevant is using similar arguments with:

$$\frac{1}{\cos(\phi) + i\sin(\phi)} = \cos(\phi) - i\sin(\phi)$$

0.15.1 Brute Force

Another proof of the same result, done by brute force, was in the same notes on my hard-drive. Presumably done as a sanity check before figuring out the

better method:

$\tan(\phi)$ being rational is the same as saying that $r\sin(\phi) = \cos(\phi)$ for some rational number r.

$$\begin{aligned}
\sin(n\phi) + i\cos(n\phi) &= \exp(in\phi) \\
&= \exp(i\phi)^n \\
&= (\sin(\phi) + i\cos(\phi))^n \\
&= \sum_{k=0}^{n} \binom{n}{k} i^k \cos(\phi)^k \sin(\phi)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} i^k r^k \sin(\phi)^n \\
&= \sin(\phi)^n \left(\sum_{k=0}^{2k\leq n} \binom{n}{2k} (-1)^k r^{2k} + i \sum_{k=0}^{2k+1\leq n} \binom{n}{2m+1} (-1)^k r^{2k+1} \right)
\end{aligned}$$

Hence:

$$\tan(n\phi) = \frac{\sum_{k=0}^{2k\leq n} \binom{n}{2k} (-1)^k r^{2k}}{\sum_{k=0}^{2k+1\leq n} \binom{n}{2k+1} (-1)^k r^{2k+1}}$$

0.16 Jensen's Inequality

A convex function ϕ on a set X is one such that for $a_n \in X$ with $w_0 + w_1 = 1$ and $w_n \geq 0$ then we have:

$$\phi(w_0 a_0 + w_1 a_1) \leq w_0 \phi(a_0) + w_1 \phi(a_1)$$

Jensen's Inequality states that expected value of the function is less then the function of the expected value:

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X))$$

By multiplying the function by -1, we get the intuitive corollary that the inequality is reversed for concave functions.

0.16.1 Generalization

A more general form of convexity can be easily proved by induction.

Assume that for some n we have that for all $\sum_{i=1}^{n} w_i' = 1$ and $w_i' \geq 0$:

$$\phi\left(\sum_{i=1}^{n} w_i' a_i\right) \leq \sum_{i=1}^{n} w_i' \phi(a_i)$$

Let $\sum_{i=1}^{n+1} w_i = 1$ and $W = \sum_{i=1}^{n} w_i$ we have $w_{n+1} + W = 1$ and :

$$\begin{aligned}
\phi\left(\sum_{i=1}^{n+1} w_i a_i\right) =& \phi\left(w_{n+1} a_{n+1} + W \sum_{i=1}^{n} \frac{w_i}{W} a_i\right) \\
\leq& w_{n+1}\phi(a_{n+1}) + W\phi\left(\sum_{i=1}^{n} \frac{w_i}{W} a_i\right) \\
\leq& w_{n+1}\phi(a_{n+1}) + W \sum_{i=1}^{n} \frac{w_i}{W} \phi(a_i) \\
\leq& \sum_{i=1}^{n+1} w_i \phi(a_i)
\end{aligned}$$

Viewing the weights as probabilities this form can directly be in interpreted as a discrete form of Jensen's Inequality.

0.16.2 Mean Inequalities

This can be directly applied to various mean inequalities.

$$AM = \frac{1}{n}\sum_{i=1}^{n} a_i$$

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} a_i^2}$$

$$GM = \sqrt[n]{\prod_{i=1}^{n} a_i}$$

Using $\phi(x) = x^2$,$w_i = \frac{1}{n}$ we directly get:

$$AM^2 \leq RMS^2$$

Using $\phi(x) = \log(x)$,$w_i = \frac{1}{n}$ we get: [8]

$$\log(AM) \geq \log(GM)$$

Since both these functions are also monotonic the clean relations follow.

0.16.3 Decision Theory

Interpret ϕ as a utility function then the convexity tells us whether we take the fixed $\mathbb{E}(X)$ or risk X. Well if ϕ is convex then:

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X))$$

And we should take the risk, there is more expected utility taking the risk then the utility of $\mathbb{E}(X)$.

If ϕ is concave then we should not take the risk.

[8] Remember that the inequality is reversed since log is concave

0.17 Convex

TODO: Flesh out, copy edit, and maybe include pictures for the geometric intuition in the cord section.

In the section of Jensen's Inequality I defined a convex function ϕ on a set X as one such that for $a_n \in X$ with $w_0 + w_1 = 1$ and $w_n \geq 0$ then we have:

$$\phi(w_0 a_0 + w_1 a_1) \leq w_0 \phi(a_0) + w_1 \phi(a_1)$$

I've thought of another proof of the previous section but iterates over functions instead of points. And we can use some geometry intuition (cords).

0.17.1 Cords

First let me define a new function called the cord function that takes on an interval $[a, b] \subseteq X$:

$$C_{[a,b]}(x) = \phi(a) + (x - a)\frac{\phi(b) - \phi(a)}{b - a}$$

The convex property can be changed to:

$$x \in [a, b] \Rightarrow \phi(x) \leq C_{[a,b]}(x)$$

Lemma:, if $x_0 \leq x_1 \leq x_2$ then:

$$x \in [x_0, x_1] \Rightarrow C_{[x_0,x_1]}(x) \leq C_{[x_0,x_2]}(x)$$

and:

$$x \in [x_1, x_2] \Rightarrow C_{[x_1,x_2]}(x) \leq C_{[x_0,x_2]}(x)$$

Proof: Use $C_{[a,b]}(a) = \phi(a)$ or $C_{[a,b]}(b) = \phi(b)$ and $x_1 \in [x_0, x_2]$ with the definition of convexity.

Lemma:, if $a_0 \leq a_1 \leq b_1 \leq b_0$ then:

$$x \in [a_1, b_1] \Rightarrow C_{[a_1,b_1]}(x) \leq C_{[a_0,b_0]}(x)$$

Proof: use the previous lemma with the triples form the quad inequality, then chain together the newer inequalities.

Let:

$$\phi_{[A,B]}(x) = \begin{cases} C_{[A,B]}(x) & x \in [A,B] \\ \phi(x) & \text{else} \end{cases}$$

$\phi_{[A,B]}(x)$ is convex. Prove this using the previous lemmas with the three cases (points same side, both sides, one in one out).

0.17.2 The Cool Proof

We want to show that for all $\sum_{i=1}^{n} w_i' = 1$ and $w_i' \geq 0$:

$$\phi\left(\sum_{i=1}^{n} w_i' a_i\right) \leq \sum_{i=1}^{n} w_i' \phi(a_i)$$

Assume it works for $n-1$ points. Take two points (w_0, a_0) and (w_1, a_1) and make a new convex function using the method from the previous section with $[x_0, x_1]$, then remove the two points and replace with $(w_0 + w_1, \frac{w_0 a_1 + w_1 a_1}{w_1 + w_0})$. Now we have a $n-1$ points, and if you expand the algebra you get the correct form, hence by induction you prove the initial result.

0.18 Old Geometry

Bellow are are collection of proofs from high school that I saved under "geometry". It includes triangle groups and two (half complete) geodesics.

0.18.1 Triangle Groups

Are a way of understanding the rotations of platonic solids. Fill in latter

Subgroup chain

Let:

$$\iota^2 = \lambda^4 = \kappa^3 = 1$$

With and those be the lowest powers that do so. Additionally have:

$$\lambda = \iota\kappa$$

This is a cube, ι rotates around a edge, κ rotates around a vertex, λ rotates around a face.

Let:

$$i = \lambda^2,\ k = \lambda\iota,\ k = \kappa^2$$

We have:

$$\begin{aligned} i^2 =& \lambda^4 \\ =& 1 \\ k^3 =& \kappa^6 \\ =& 1 \\ l^3 =& (\lambda\iota)^3 \\ =& (\iota\kappa\iota)^3 \\ =& \iota\kappa^3\iota \\ =& 1 \end{aligned}$$

These are the lowest powers to do so. Minimality is trivial for i and k by the

minimality of κ and ι. For l assume $l^2 = 1$ then:

$$\begin{aligned} l^2 &= 1 \\ \lambda\iota\lambda\iota &= 1 \\ \iota\lambda\iota &= \lambda^3 \\ \kappa\iota &= \lambda^3 \\ \kappa\iota\lambda &= \lambda^4 \\ &= 1 \\ \kappa^2 &= 1 \end{aligned}$$

We have the relation:

$$\begin{aligned} ik &= \lambda^3\iota \\ &= \lambda^3\iota\kappa^3 \\ &= \lambda^3\lambda\kappa^2 \\ &= \kappa^2 \end{aligned}$$

Hence the rotations of a regular tetrahedron are a subgroup of the rotations of a cube.

0.18.2 Torus

Let a torus be parametrized by:

$$\begin{aligned} r &= (x, y, z) \\ x &= (R + r\cos(\theta))\cos(\phi) \\ y &= (R + r\cos(\theta))\sin(\phi) \\ z &= r\sin(\theta) \end{aligned}$$

The partials of r are given by:

$$\begin{aligned} \frac{\partial r}{\partial \phi} &= (-(R + r\cos(\theta))\sin(\phi), (R + r\cos(\theta))\cos(\phi), 0) \\ \frac{\partial r}{\partial \theta} &= (-r\sin(\theta)\cos(\phi), -r\sin(\theta)\sin(\phi), r\cos(\theta)) \\ \frac{\partial r}{\partial \phi} \cdot \frac{\partial r}{\partial \phi} &= (R + r\cos(\theta))^2 \\ \frac{\partial r}{\partial \theta} \cdot \frac{\partial r}{\partial \theta} &= r^2 \\ \frac{\partial r}{\partial \theta} \cdot \frac{\partial r}{\partial \phi} &= 0 \end{aligned}$$

Use the partial to get the line element:

$$\begin{aligned}\dot{r}^2 &= \left(\frac{\partial r}{\partial \phi}\dot{\phi} + \frac{\partial r}{\partial \theta}\dot{\theta}\right)^2 \\ &= \frac{\partial r}{\partial \phi}\cdot\frac{\partial r}{\partial \phi}\dot{\phi}^2 + \frac{\partial r}{\partial \theta}\cdot\frac{\partial r}{\partial \theta}\dot{\theta}^2 + 2\frac{\partial r}{\partial \theta}\cdot\frac{\partial r}{\partial \phi}\dot{\phi}\dot{\theta} \\ &= (R + r\cos(\theta))^2\dot{\phi}^2 + r^2\dot{\theta}^2\end{aligned}$$

Treating the line element as the integrand in Lagrangian integral:

$$L = \frac{1}{2}\dot{r}^2$$

$$\begin{aligned}\frac{\partial L}{\partial \phi} =& \frac{d}{dt}\frac{\partial L}{\partial \dot{\phi}} \\ 0 =& \frac{d}{dt}(R + r\cos(\theta))^2\dot{\phi} \\ =& \left[\dot{\theta}\frac{\partial}{\partial \theta} + \ddot{\phi}\frac{\partial}{\partial \dot{\phi}}\right](R + r\cos(\theta))^2\dot{\phi} \\ =& (R + r\cos(\theta))^2\ddot{\phi} - 2r\sin(\theta)(R + r\cos(\theta))\dot{\phi}\dot{\theta} \\ \ddot{\phi} =& \frac{2r\sin(\theta)}{R + r\cos(\theta)}\dot{\phi}\dot{\theta}\end{aligned}$$

And again:

$$\begin{aligned}\frac{\partial L}{\partial \theta} =& \frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} \\ -r\sin(\theta)(R + r\cos(\theta))\ddot{\phi}^2 =& r^2\frac{d}{dt}\dot{\theta} \\ -r\sin(\theta)(R + r\cos(\theta))\ddot{\phi}^2 =& r^2\ddot{\theta}\end{aligned}$$

0.18.3 Sphere

Let a sphere be parametrized by:

$$\begin{aligned}r &= (x, y, z) \\ x &= \cos(\theta)\cos(\phi) \\ y &= \cos(\theta)\sin(\phi) \\ z &= \sin(\theta)\end{aligned}$$

The partials of r are given by:

$$\begin{aligned}
\frac{\partial r}{\partial \phi} &= (-\cos(\theta)\sin(\phi), \cos(\theta)\cos(\phi), 0) \\
\frac{\partial r}{\partial \theta} &= (-\sin(\theta)\cos(\phi), -\sin(\theta)\sin(\phi), \cos(\theta)) \\
\frac{\partial r}{\partial \phi} \cdot \frac{\partial r}{\partial \phi} &= \cos(\theta)^2 \\
\frac{\partial r}{\partial \theta} \cdot \frac{\partial r}{\partial \theta} &= 1 \\
\frac{\partial r}{\partial \theta} \cdot \frac{\partial r}{\partial \phi} &= 0
\end{aligned}$$

Use the partial to get the line element:

$$\begin{aligned}
\dot{r}^2 &= \left(\frac{\partial r}{\partial \phi}\dot{\phi} + \frac{\partial r}{\partial \theta}\dot{\theta}\right)^2 \\
&= \frac{\partial r}{\partial \phi} \cdot \frac{\partial r}{\partial \phi}\dot{\phi}^2 + \frac{\partial r}{\partial \theta} \cdot \frac{\partial r}{\partial \theta}\dot{\theta}^2 + 2\frac{\partial r}{\partial \theta} \cdot \frac{\partial r}{\partial \phi}\dot{\phi}\dot{\theta} \\
&= \cos(\theta)^2\dot{\phi}^2 + \dot{\theta}^2
\end{aligned}$$

Geodesic Equations:

$$\begin{aligned}
\ddot{\theta} &= -\sin(\theta)\cos(\theta)\dot{\phi}^2 \\
0 &= \frac{d}{dt}\cos(\theta)^2\dot{\phi} \\
&= \cos(\theta)^2\ddot{\phi} - 2\cos(\theta)\sin(\theta)\dot{\phi}\dot{\theta}
\end{aligned}$$

Consider the function:

$$\begin{aligned}
f(t) &= z + \alpha x + \beta y \\
f(t) &= \sin(\theta) + \alpha\cos(\theta)\cos(\phi) + \beta\cos(\theta)\sin(\phi) \\
\dot{f}(t) &= \dot{\theta}\cos(\theta) + \alpha(-\dot{\theta}\sin(\theta)\cos(\phi) - \dot{\phi}\cos(\theta)\sin(\phi)) + \beta(-\dot{\theta}\sin(\theta)\sin(\phi) + \dot{\phi}\cos(\theta)\cos(\phi))
\end{aligned}$$

0.19 Isomorphisms of the quotient rings of $\mathbb{Z}[X]$

While working elsewhere I thought that it's actually really easy to prove:

$$\gcd(a,d) = \gcd(b,d) \Rightarrow \exists z \in \mathbb{Z} \text{ such that } z\frac{a}{d} - \frac{b}{d} \in \mathbb{Z}$$

Using Bézout's identity.

0.19.1 The proof

From Bézout's identity we know there exits r, s in $\mathbb{Z}$ such that:

$$ra + sd = \gcd(a,d) = \gcd(b,d)$$

From the definition of gcd there exists $d' \in \mathbb{Z}$ such that:

$$b' \gcd(b,d) = b$$

Multiplying the first equation by b' gives:

$$(rb')a + (sb')d = b' \gcd(b,d) = b$$

Hence:

$$(rb')\frac{a}{d} - b\frac{b}{d} = -(sb')$$

As required.

Note that we can actually control the size, and hence sign, of z by using the:

$$ra + sd = a(r + bk) + b(s - ak)$$

Trick, meaning we can force $z > 0$ and hence $z \in \mathbb{N}$.

0.19.2 The questions

So this proof brings up some questions:

1. Doesn't this prove that $\mathbb{Z}[X]/(dX - a) \cong \mathbb{Z}[X]/(dX - b)$?
2. Couldn't this be generalized to Bézout Domains?
3. What about the converse?

0.19.3 Question 1

Let:

$$A = \mathbb{Z}[X]/(dX - a), \quad B = \mathbb{Z}[X]/(dX - b)$$

We can say A and $B \subset \mathbb{Q}$ where we identity X with $\frac{a}{d}$ and $\frac{b}{d}$ respectfully. But since we can find a linear relation between the two Xs in terms, we need two relations but we can avoid division completely, of the sub-field $\mathbb{Z}$ I'd expect an isomorphism.

Assume the gcd condition is meet and let z_0 and z_1 be the integers such that:

$$z_1 \frac{a}{d} - \frac{b}{d} = -z_0$$

Equally written as:

$$d(z_1 a + z_0) = b$$

Attempt 1

Define $\phi : B \to A$. A ring homomorphism fixes 1 and hence acts identically on the set generated by 1, that is $\mathbb{Z}$. Since $a \in \mathbb{Z}$ hence:

$$\begin{aligned} a =& \phi(a) \\ =& \phi(dX) \\ =& d\phi(X) \end{aligned}$$

Suggesting a form for $\phi(X)$:

$$\phi(X) = z_1 X + z_0$$

This is all the degrees of freedom we get since being a homomorphism requires:

$$\phi\left(\sum_n p_n X^n\right) = \sum_n \phi(p_n)\phi(X)^n$$

Hence the suggested form is:

$$\phi\left(\sum_n p_n X^n\right) = \sum_n p_n (z_1 X + z_0)^n$$

But dealing with showing homomorphism when dealing the canceling term isn't something I want to do.

Attempt 2

Define a function $\phi : \mathbb{Z}[X] \to B$ where:

$$\phi\left(\sum_n p_n X^n\right) = \sum_n p_n (z_1 X + z_0)^n \mod (dX - b)$$

Then we can use the first isomorphism theorem of rings and only need to show that ϕ is surjective and has a kernel of $\langle dX - a \rangle$.

0.20 Generated Subfield of Finite Fields

Let the finite field F have a set of elements S. The set $\langle S \rangle$ whose elements are the sums of the products of powers of S:

$$\langle S \rangle = \left\{ \sum_n \prod_i s_i^{p_{i,n}} \mid s_i \in S \text{ and } p_{i,n} \in \mathbb{Z} \right\} \cup \{0\}$$

For example if $S = \{s_0, s_1, s_2\}$ all of the following are elements of $\langle S \rangle$:

$$s_0 + s_1 + s_2, \quad s_0 s_1 s_2 s_3, \quad 1 + s_0 + s_1 s_2^7 + s_0 s_1 s_2^2$$

(I use the $\langle \cdot \rangle$ notation because this is like generation, one of the most irregular and abused words in algebra).

The set inherits it's operations from F and is clearly closed under addition and multiplication. It also contains 1 and its repeated sum, labeled n, by using $p_{i,n} = 0$:

$$n = \overbrace{(s_0 s_1 s_2)^0 + (s_0 s_1 s_2)^0 + \ldots s(s_0 s_1 s_2)^0}^{n \text{ times}}$$

Given $q \in \langle S \rangle$ because there are a finite number of options for nq the pigeonhole principle means there will eventually be $n > m + 1$ such that:

$$\begin{aligned} nq &= mq \\ (n-m)q &= 0 \\ q + (n-m-1)q &= 0 \end{aligned}$$

$(n - m - 1)q$ which is the additive inverse of q. Because $n - m - 1 > 0$ it is an element of $\langle S \rangle$ and since q is also an element by closure so is the inverse.

Likewise for the multiplicative inverse:

$$\begin{aligned} q^n &= q^m \\ q \cdot q^{n-m-1} &= 1 \end{aligned}$$

Hence $\langle S \rangle$ is a, not necessarily proper, subfield of F.

0.20.1 $S = \{1\}$

An important example is when $S = \{1\}$. Let n be the "lowest"[9] where the pigeonhole principle whole applies:

$$\begin{aligned} m + (n - m) =& n + (m - m) \\ =& n \\ =& m \end{aligned}$$

The minimality[10] of n makes $n - m$ non-zero making n zero. Hence $S \cong \mathbb{F}_n$.

(This is sloppy even for quick notes)

0.20.2 Wedderburn's Little Theorem

This is similar to Wedderburn's little theorem, that all finite domains are fields. Wedderburn generalizes the subject to domains instead of field and thus has to use $q - q^n = 0$ and some polynomial analysis (Headache). But reduces the results, it doesn't discuss closure.

Cool result, I suggest looking it up.

0.20.3 Dreams of a more rigor

Let R be a ring and let ϕ be a function from $\mathbb{Z}$ to R such that:

$$\phi(1) = 1_R, \quad \phi(n \pm 1) = \phi(n) \pm \phi(1_R) = \phi(n) \pm 1_R$$

ϕ is well-defined as you can find the value for any $n \in \mathbb{Z}$ by induction. And the function only has one value at each n since you can't change the value by "doubling back" on the induction:

$$\phi((n \pm 1) \mp 1) = \phi(n \pm 1) \mp 1_R = \phi(n) \pm 1_R \mp 1_R = \phi(n)$$

[9] Yes I know I haven't defined that properly, but just figure it out.
[10] See previous footnote.

Induction on n shows:

$$\begin{aligned}\phi(n)+\phi(m) =&\phi(n\pm 1)\mp 1+\phi(m)\\ =&\phi(n\pm 1)+\phi(m\mp 1)\\ =&\phi(n\pm 1+m\mp 1)\\ =&\phi(n+m)\end{aligned}$$

Which can be used to further prove:

$$\begin{aligned}\phi(n)\phi(m) =&(\phi(n\pm 1)\mp 1)\phi(m)\\ =&\phi(n\pm 1)\phi(m)\mp\phi(m)\\ =&\phi((n\pm 1)m)\mp\phi(m)\\ =&\phi((n\pm 1)m\mp m)\\ =&\phi(nm)\end{aligned}$$

These identities, combined ϕ being defined with $\phi(1)=1_R$, shows that ϕ is a homomorphism.

Which, by the first isomorphisms theorem for rings, shows that the $\operatorname{img}\phi$ is a subring of R and is isomorphic to $\mathbb{Z}/\ker\phi$. This formalizes the concept of the nth sum of 1_R, and lets us manipulate it better.

Corollary: If R is finite then the additive group of R is also finite. Meaning 1_R has an order, which we label N. In this case $\ker\phi = N\mathbb{Z}$ giving:

$$\operatorname{img}\phi = \frac{\mathbb{Z}}{N\mathbb{Z}}$$

It's really easy to use Bézout's identity on these objects to show which ones are fields.

0.20.4 Order

Lemma: $n\in\ker\phi \Rightarrow n\mathbb{Z}\subseteq\ker\phi$

A corollary of ϕ being a homomorphism is that:

$$\begin{aligned}n\in\ker\phi \Rightarrow&\phi(nk)=\phi(n)\phi(k)=0\\ \Rightarrow&n\mathbb{Z}\subseteq\ker\phi\end{aligned}$$

Lemma: If there exits $n \in \ker \phi$ such that for all non-zero $m \in \ker \phi$ we have $|m| \geq |n|$ then $n\mathbb{Z} \supseteq \ker \phi$

(Going forward remember that either $|n| = n$ or $|n| = -n$).
Assume $m \in \ker \phi$ from $|m| > |n|$ there exits $q, r \in \mathbb{Z}$ with $|r| < |n|$ such that:

$$|m| = q|n| + r \Rightarrow q|n| + r - |m| = 0$$

But:

$$\begin{aligned} 0 &= \phi(q|n| + r - |m|) \\ &= \phi(q)\phi(|n|) + \phi(r) - \phi(|m|) \\ &= \phi(r) \\ &= \phi(|r|) \end{aligned}$$

Meaning we have $r \in \ker \phi$ and $|r| < |n|$ Unless $r = 0$ this contradicts the assumption, hence:

$$|m| = q|n| \Rightarrow m \in n\mathbb{Z}$$

Theorem: If there exits $n \in \ker \phi$ such that for all non-zero $m \in \ker \phi$ we have $|m| \geq |n|$ then $n\mathbb{Z} = \ker \phi$

Combine the previous lemmas.

Lemma: $n \in \ker \phi \Rightarrow \operatorname{img} \phi = \phi(\{k \mid 0 \leq k < n\})$

Let $m \in \mathbb{Z}$ then there exits $q, r \in \mathbb{Z}$ with $r \in \{k \mid 0 \leq k < n\}$ such that:

$$m = qn + r$$

Hence:

$$\begin{aligned} \phi(m) =& \phi(qn + r) \\ =& \phi(r) \end{aligned}$$

From the definition of r we have

$$\phi(m) = \phi(r) \in \phi(\{k \mid 0 \leq k < n\})$$

As required.

Lemma: Let n be the finite order of 1_R in the additive group of R then $\ker \phi = n\mathbb{Z}$

The additive subgroup generated from a single element is the complete cyclic group of the order of the element.

n is the order of the subgroup generated by 1_R. $n \in \ker \phi$

$$\{\phi(k)|k \in [0, n-1]\}$$

0.21 Dominated Convergence Theorem

Consider the following sequences of integrals:

$$I_n = \int_0^1 \log(t)^2 (1-t)^n \, dt$$

It's related to the Harmonic numbers, which is why it peaked my interest, but it was presented in context where we wished to understand its asymptotics.

The book answer of expressing it in terms of hypergeometric functions will be skipped, because it's boring.

0.21.1 Cauchy-Schwartz

$$\begin{aligned}\int_0^1 \log(t)^2 (1-t)^n \, dt \leq & \sqrt{\int_0^1 \log^4(t) \, dt \int_0^1 (1-t)^{2n} \, dt} \\ = & \sqrt{4! \frac{1}{2n+1}} \to 0\end{aligned}$$

Which by the noticing the sign of I_n and application of the sandwich theorem shows that $I_n \to 0$.

0.21.2 Dominated Convergence Theorem

As suggested by the section name, this is the result I really want to take notes about. If a sequence of functions f_n pointwise converge to f and are dominated by some integrable function g such that:

$$|f_n(t)| < g(t)$$

We have:

$$\int_S f_n \, d\mu \to \int_S f \, d\mu$$

Well rewrite the integral as:

$$\begin{aligned}\frac{n}{\log(n)^2}\int_0^1 \log(t)^2(1-t)^n\,dt &= \int_0^1 \left(\frac{(\log(nt)-\log(n)}{\log(n)}\right)^2\left(1-\frac{nt}{n}\right)^n\,d\,nt\\ &= \int_0^n \left(\frac{\log(t)}{\log(n)}-1\right)^2\left(1-\frac{t}{n}\right)^n\,dt\\ &= \int_0^\infty \chi_{[0,n]}(t)\left(1-\frac{\log(t)}{\log(n)}\right)^2\left(1-\frac{t}{n}\right)^n\,dt\end{aligned}$$

Set:

$$f_n(t) = \overbrace{\chi_{[0,n]}(t)\left(1-\frac{\log(t)}{\log(n)}\right)^2}^{g_n(x)}\overbrace{\left(1-\frac{t}{n}\right)^n}^{h_n(t)}$$

We have (pointwise):

$$f_n(t) \to \exp(-t)$$

We also have $h_n(t) \leq \exp(-t)$ and $g_n(t) \leq (1-\log(t))^2$, meaning:

$$|f_n(t)| \leq (1-\log(t))^2\exp(-t)$$

Giving the desired result (Assuming we can show the RHS converges, which I'm pretty sure we can since near 0 we have $\int(1-\log(t))^2\,dt = t(\log(t)^2 - 4\log(t)+5)$ and far away we have $\exp(-t)$).

$$\frac{n}{\log(n)^2}\int_0^\infty \log(t)^2(1-t)^n\,dt \to \int_0^\infty \exp(-t)\,dt = 1$$

0.22 Divisibility of $(p+1)(p-1)$

It is well known that if p is prime than either p is a divisor of 6 or has the remainder 1 or 5 when divided by 6. This can be verified by noting that a remainder of 0,2, or 4 would make p divisible by 2 and likewise for 0 and 3 for 3.

A cool corollary I saw today was that p is a divisor of 6 or 24 is a divisor of $(p+1)(p-1)$. This is also verified by cases:

$p \mod 12$	1	5	7	11
$\gcd(p+1, 12)$	2	6	4	12
$\gcd(p-1, 12)$	12	4	6	2

The converse isn't true. Since $24 \cdot 3 = 72$, but $(7-1)(7+1) = 48 < 72$ and $(11-1)(11+1) = 120 > 72$.