

Jumbled up thoughts

Stefanus Koesno¹

¹ Somewhere in California

San Jose, CA 95134 USA

(Dated: October 7, 2016)

Abstract

Stray thoughts of a stray man. This is a collection of things that I thought was fun to do and also things that I always come back to after forgetting it :) There's no order or structure here just whatever my mind facies at the moment.



FIG. 1: A rotation of the coordinate axes around the \hat{z} axis.

How does a field and the associated Lagrangian change when we perform a rotation?

Intuitively say we perform an active rotation on the axes themselves and for simplicity we take a small angle rotation around \hat{z} , we'll see that \hat{x} will get a contribution in $+\hat{y}$ direction and \hat{y} will get a contribution in $-\hat{x}$ direction, see Fig. 1, which means

$$\vec{\delta}\hat{x} \Rightarrow +\hat{y} = \hat{z} \times \hat{x}$$

$$\vec{\delta}\hat{y} \Rightarrow -\hat{x} = \hat{z} \times \hat{y}$$

Note that this only works for infinitesimal rotation since for a finite rotation there will be contributions from multiple axes, *e.g.* $\vec{\delta}\hat{x} \propto \hat{y} - \hat{x}$, if you can't see this just remember the rotation matrix in 2 dimension.

How about a generic rotation axis? say we pick an arbitrary rotation axis \vec{r} and we want to rotate a vector \vec{A} , remember that the magnitude of \vec{r} is the amount of rotation, *i.e.*

$$\vec{r} = a\hat{x} + b\hat{y} + c\hat{z} \longrightarrow |\vec{r}| = \sqrt{a^2 + b^2 + c^2} = \epsilon$$

Thus the change in \vec{A} is given by

$$\vec{\delta}\vec{A} = \vec{r} \times \vec{A} = a(\hat{x} \times \vec{A}) + b(\hat{y} \times \vec{A}) + c(\hat{z} \times \vec{A})$$

Here we see why it only works for infinitesimal rotations since rotations w.r.t different axes do not commute but they only contribute at ϵ^2 .

What if instead of a vector we have a scalar A ? First we need to convert it into a vector, but how? there are many ways of doing it. If we look back to translation $A(x) \rightarrow A(x+\epsilon)$ the transformation we get is $A \rightarrow A + \epsilon \partial A$. Now, we just need to turn this into a vector, the most obvious way is $A \rightarrow A + \vec{\epsilon} \cdot (\vec{\nabla} A)$ (note: we always need a derivative to calculate δA thanks to taylor expansion).

Using the formula we have above

$$\vec{\delta}(\nabla A) = \vec{r} \times \nabla A = a(\hat{x} \times \nabla A) + b(\hat{y} \times \nabla A) + c(\hat{z} \times \nabla A)$$

however the result we have is a vector we need a scalar since A was initially a scalar. The solution is just to take a dot product with $\vec{x} = \frac{1}{\sqrt{3}}(\hat{x} + \hat{y} + \hat{z})$

$$\delta A = \vec{x} \cdot (\vec{r} \times \nabla A) = -\vec{r} \cdot (\vec{x} \times \nabla A)$$

by using a triple scalar product rule, the thing here now is that I don't know how to get rid of that minus sign LOL

So for a general scalar field you'll get $\delta\phi = \vec{\epsilon} \cdot (\vec{x} \times \nabla)\phi$, sans a minus sign, where \vec{x} is just the coordinate axes and $\vec{\epsilon}$ is the magnitude and axis of rotation.

Now, how does the lagrangian change? Well, the lagrangian is also a scalar quantity thus we can expect it to also change as $\delta\mathcal{L} = \vec{\epsilon} \cdot (\vec{x} \times \nabla)\mathcal{L}$.

The goal here is to show that $\delta\mathcal{L}$ is related to the stress energy tensor *of translation* and for simplicity we will ignore $\vec{\epsilon}$ for now

$$\begin{aligned} (\vec{x} \times \nabla)\mathcal{L} &= \epsilon^{ijk} x_j \partial_k \mathcal{L} = \partial_\mu (\epsilon^{ijk} x_j \delta_k^\mu \mathcal{L}) \\ &= \partial_\mu \mathcal{F}^{\mu i} \\ &\rightarrow \mathcal{F}^{\mu i} = \epsilon^{ijk} x_j \delta_k^\mu \mathcal{L} \end{aligned}$$

Since this is a symmetry the change must be a total derivative.

Now from the generic formula for energy stress tensor (under rotation)

$$\begin{aligned}
\tilde{T}^{\mu i} &= \frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \delta \phi - \mathcal{F}^{\mu i} \\
&= \frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} (\vec{x} \times \vec{\nabla}) \phi - \epsilon^{ijk} x_j \delta_k^\mu \mathcal{L} \\
&= \epsilon^{ijk} x_j \left(\frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \partial_k \phi - \delta_k^\mu \mathcal{L} \right) = \epsilon^{ijk} x_j (T_k^\mu) \\
&= \vec{x} \times \vec{T}^\mu
\end{aligned}$$

Notes on energy stress tensor: Under symmetry transformations, the lagrangian should change by a total derivative (since a symmetry should not change the physics)

$$\delta \mathcal{L} = \partial_\mu \mathcal{F}^\mu$$

However the transformation is also given by

$$\begin{aligned}
\delta \mathcal{L} &= \frac{\delta \mathcal{L}}{\delta \phi} \delta \phi + \frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \delta(\partial_\mu \phi) \\
&= \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \right) \delta \phi + \frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \partial_\mu (\delta \phi) \\
\partial_\mu \mathcal{F}^\mu &= \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \delta \phi \right) \\
0 &= \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \delta \phi - \mathcal{F}^\mu \right) = \partial_\mu J^\mu \\
\rightarrow J^\mu &= \frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \delta \phi - \mathcal{F}^\mu
\end{aligned}$$

where we have used the equation of motion on the first term going into the second line, this is actually tricky if we apply wrongly we will get something like (doing Integration by parts on the second term at the same time)

$$\begin{aligned}
\delta \mathcal{L} &= \frac{\delta \mathcal{L}}{\delta \phi} \delta \phi + \frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \delta(\partial_\mu \phi) = \frac{\delta \mathcal{L}}{\delta \phi} \delta \phi - \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \right) \delta \phi \\
&= \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \right) \delta \phi - \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \right) \delta \phi \\
&= 0 \quad ???
\end{aligned}$$

LOL indeed but what went wrong? One, we can't really use IBP here since we are not

really under an integral. However we can just use the usual trick, $u(dv) = d(uv) - (du)v$

$$\begin{aligned}\frac{\delta \mathcal{L}}{\delta \phi} \delta \phi + \frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \delta(\partial_\mu \phi) &= \frac{\delta \mathcal{L}}{\delta \phi} \delta \phi + \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \delta \phi \right) - \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \right) \delta \phi \\ &= \left\{ \frac{\delta \mathcal{L}}{\delta \phi} \delta \phi - \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \right) \delta \phi \right\} + \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \delta \phi \right) \\ &= 0 + \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} \delta \phi \right)\end{aligned}$$

again using the equations of motion for ϕ .

In QFT, why does $\phi(x)$ create a particle at spacetime point x , *i.e.* $\phi(x)|0\rangle = |x\rangle$? For simplicity we'll set $t = 0 \rightarrow \phi(x) = \phi(\vec{x}, 0)$. To see this take the inner product of

$$\begin{aligned}\langle \vec{x} | \vec{p} \rangle &= \langle 0 | \phi^\dagger(\vec{x}, 0) | \vec{p} \rangle \\ &= \langle 0 | \left(\int \frac{d^3 p'}{(2\pi)^3 \sqrt{\omega_{p'}}} a^\dagger e^{-i\vec{p}'\vec{x}} + a e^{i\vec{p}'\vec{x}} \right) | \vec{p} \rangle \\ &= \left(\langle 0 | \int \frac{d^3 p'}{(2\pi)^3 \sqrt{\omega_{p'}}} a e^{i\vec{p}'\vec{x}} \right) | \vec{p} \rangle, \quad (\langle 0 | a^\dagger) = 0 \\ &= \int \frac{d^3 p'}{(2\pi)^3 \sqrt{\omega_{p'}}} e^{i\vec{p}'\vec{x}} \{ \langle \vec{p}' | \vec{p} \rangle \} \\ &= \int \frac{d^3 p'}{(2\pi)^3 \sqrt{\omega_{p'}}} e^{i\vec{p}'\vec{x}} \left\{ (2\pi)^3 \sqrt{\omega_p} \delta(\vec{p}' - \vec{p}) \right\} \\ \langle \vec{x} | \vec{p} \rangle &= e^{i\vec{p}\vec{x}}\end{aligned}$$

Thus $\phi(x)$ does indeed create a particle at spacetime point x :)

What is the explicit formula for multiple integrations of a function, *i.e.* what is $\int \int \dots \int f(x) dx$? The answer is the Cauchy formula for repeated integration

$$\begin{aligned}(Jf)(x) &= \int_0^x f(t) dt \\ (J^2 f)(x) &= \int_0^x \left(\int_0^t f(s) ds \right) dt \\ (J^\alpha f)(x) &= \frac{1}{(\alpha-1)!} \int_0^x (x-t)^{\alpha-1} f(t) dt \\ &= \frac{1}{\Gamma(\alpha)} \int_0^x (x-t)^{\alpha-1} f(t) dt\end{aligned}$$

Here we have used J as an integration operator and the identity $\Gamma(\alpha) = (\alpha - 1)!$, and also that gamma function can take non integer arguments compared to factorials.

So how do we derive this? The most intuitive derivation I saw was using Laplace transform, some preliminary Laplace transforms that we need are

1. Laplace transform of an integral of a function, $\mathcal{L}\{\int_0^x f(t)dt\}$

$$\mathcal{L}\left\{\int_0^x f(t)dt\right\} = \int_0^\infty e^{-sx} \left(\int_0^x f(t)dt\right) dx$$

we now use integration by parts, $u = \left(\int_0^x f(t)dt\right)$, $dv = e^{-sx}dx$ such that

$$\begin{aligned}\mathcal{L}\left\{\int_0^x f(t)dt\right\} &= \left(\int_0^x f(t)dt\right) \frac{e^{-sx}}{(-s)} \Big|_{x=0}^{x=\infty} - \left(-\frac{1}{s}\right) \int_0^\infty e^{-sx} f(x)dx \\ &= 0 + \frac{1}{s} \int_0^\infty e^{-sx} f(x)dx \\ \mathcal{L}\{(Jf)(x)\} &= \frac{1}{s} \mathcal{L}\{f(x)\}\end{aligned}$$

the boundary terms are zero because $\int_0^0 f(t) = 0$ and $e^{-\infty} = 0$, assuming $f(t)$ is well behaved. Generalizing to $(J^\alpha f)(x)$ we have

$$\mathcal{L}\{(J^\alpha f)(x)\} = \frac{1}{s^\alpha} \mathcal{L}\{f(x)\}$$

2. Laplace transform of x^α , *i.e.* $\mathcal{L}\{x^\alpha\}$ (all function under Laplace transform are assumed to be valid only on the positive real axis)

$$\mathcal{L}\{x^\alpha\} = \int_0^\infty e^{-sx} x^\alpha dx$$

here we just need to do multiple integration by parts with $u = x^\alpha$, $dv = e^{-sx}dx$

$$\begin{aligned}\mathcal{L}\{x^\alpha\} &= x^\alpha e^{-sx} \Big|_{x=0}^{x=\infty} - \frac{\alpha}{(-s)} \int_0^\infty x^{\alpha-1} e^{-sx} dx \\ &= \frac{\alpha}{s} \int_0^\infty x^{\alpha-1} e^{-sx} dx\end{aligned}$$

here we can do 2 things, one use the definition of gamma function $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ or repeat the integration by parts again (note that the boundary terms are again zero in

this case), if we repeat the IBP $\alpha - 1$ more times we get

$$\begin{aligned}
\mathcal{L}\{x^\alpha\} &= \frac{\alpha!}{s^\alpha} \int_0^\infty e^{-sx} dx \\
&= \frac{\alpha!}{s^\alpha} \frac{1}{(-s)} e^{-sx} \Big|_0^\infty \\
&= -\frac{\alpha!}{s^{\alpha+1}} (e^{-\infty} - e^0) \\
\mathcal{L}\{x^\alpha\} &= \frac{\alpha!}{s^{\alpha+1}}
\end{aligned}$$

which can be generalized to $\mathcal{L}\{x^\alpha\} = \frac{\Gamma(\alpha+1)}{s^{\alpha+1}}$ or just by using the definition of gamma function from the beginning

$$\begin{aligned}
\frac{\alpha}{s} \int_0^\infty x^{\alpha-1} e^{-sx} dx &= \frac{\alpha}{s} \int_0^\infty \left(\frac{y}{s}\right)^{\alpha-1} e^{-y} \left(\frac{dy}{s}\right), \quad y = sx, \quad dx = \frac{dy}{s} \\
&= \frac{\alpha}{s^{\alpha+1}} \int_0^\infty y^{\alpha-1} e^{-y} dy = \frac{\alpha \Gamma(\alpha)}{s^{\alpha+1}}, \quad \alpha \Gamma(\alpha) = \Gamma(\alpha+1) \\
\mathcal{L}\{x^\alpha\} &= \frac{\Gamma(\alpha+1)}{s^{\alpha+1}}
\end{aligned}$$

which is the same result as above.

3. The infamous convolution formula

$$\begin{aligned}
\mathcal{L}\{f\}\mathcal{L}\{g\} &= \left(\int_0^\infty e^{-sx} f(x) dx \right) \left(\int_0^\infty e^{-sy} g(y) dy \right) \\
&= \int_0^\infty e^{-s(x+y)} \left(\int_0^\infty f(x) g(y) dy \right) dx, \quad z = x + y \\
&= \int_0^\infty e^{-sz} \left(\int_0^z f(z-y) g(y) dy \right) dz \\
\mathcal{L}\{f\}\mathcal{L}\{g\} &= \mathcal{L}\{f * g\}
\end{aligned}$$

where $*$ denotes convolution, some explanation on $dx \rightarrow dz$, inside the brackets of $\int dy$, x is constant while outside of this bracket y is *constant* and thus $dz = dx$.

We are now ready to derive the formula for $(J^\alpha f)(x)$

$$\begin{aligned}
(J^\alpha f)(x) &= \mathcal{L}^{-1}\{\mathcal{L}\{J^\alpha f\}\} \\
&= \mathcal{L}^{-1}\left\{\frac{1}{s^\alpha}\mathcal{L}\{f\}\right\} = \mathcal{L}^{-1}\left\{\frac{\Gamma(\alpha)}{\Gamma(\alpha)}\frac{1}{s^\alpha}\mathcal{L}\{f\}\right\} \\
&= \frac{1}{\Gamma(\alpha)}\mathcal{L}^{-1}\left\{\frac{\Gamma(\alpha)}{s^\alpha}\mathcal{L}\{f\}\right\} \\
&= \frac{1}{\Gamma(\alpha)}\left(\mathcal{L}^{-1}\left\{\frac{\Gamma(\alpha)}{s^\alpha}\right\} * f\right) \\
&= \frac{1}{\Gamma(\alpha)}(x^{\alpha-1} * f) \\
(J^\alpha f)(x) &= \frac{1}{\Gamma(\alpha)}\int_0^x (x-t)^{\alpha-1}f(t)dt
\end{aligned}$$

which is the Cauchy formula we were looking for.

Problem 1.1 Srednicki, Show that the Dirac matrices must be even dimensional. Hint: show that the eigenvalues of β are all ± 1 , and that $\text{Tr}\beta = 0$. To show that $\text{Tr}\beta = 0$, consider e.g. $\text{Tr}\alpha_1^2\beta$. Similarly, show that $\text{Tr}\alpha_i = 0$.

Some facts about Dirac matrices

$$\{\alpha^j, \alpha^k\}_{ab} = 2\delta^{jk}\delta_{ab} \quad \{\alpha^j, \beta\}_{ab} = 0 \quad (\beta)_{ab}^2 = \delta_{ab}$$

For eigenvalues of β it follows immediately from $(\beta)_{ab}^2 = \delta_{ab}$ its eigenvalues must be ± 1 .

For $\text{Tr}\beta$ we start with

$$\begin{aligned}
\text{Tr}(\alpha_1^2\beta) &= \text{Tr}(\alpha_1\beta\alpha_1), \quad \{\alpha^j, \beta\}_{ab} = 0 \\
&= \text{Tr}(-\beta\alpha_1\alpha_1) \\
\text{Tr}(\alpha_1^2\beta) &= \text{Tr}(\beta\alpha_1\alpha_1), \quad \text{Tr}(AB) = \text{Tr}(BA) \\
\rightarrow \text{Tr}(-\beta\alpha_1\alpha_1) &= \text{Tr}(\beta\alpha_1\alpha_1) \\
\text{Tr}(\beta\alpha_1\alpha_1) &= 0
\end{aligned}$$

Now from $\{\alpha^j, \alpha^k\}_{ab} = 2\delta^{jk}\delta_{ab} \rightarrow \{\alpha_i, \alpha_i\}_{ab} = 2\delta_{ab}$, it means that $(\alpha_i^2)_{ab} = \delta_{ab}$ and thus $\text{Tr}(\beta\alpha_i\alpha_i) = \text{Tr}(\beta) = 0$.

Now since $\text{Tr}\beta = 0$ and its eigenvalues are ± 1 , β must be even dimensional.

This argument applies for $\text{Tr}\alpha_i$ as well by reversing the roles of β and α , $\text{Tr}(\beta^2\alpha_i) = \text{Tr}(\alpha_i) = 0$ and since $(\alpha_i^2)_{ab} = \delta_{ab} \rightarrow$ eigenvalues must be ± 1 , α^i must be even dimensional. This means that *all* of Dirac matrices must be even dimensional.

Now we need to show that Dirac matrices must be bigger than 2×2 . I'll start with the fact that any 2×2 matrix can be written as $a_0 1 + a_i \sigma^i$, *i.e.* a combination of Pauli matrices and the identity matrix. If you remember some sort of group theory, the group $su(2)$ spans a three sphere and the basis vectors (lie algebra $SU(2)$) are spanned by the Pauli matrices.

The commutation relations of Pauli matrices are

$$\begin{aligned}\{\sigma^i, \sigma^j\}_{ab} &= 2\delta^{ij}\delta_{ab} \\ [\sigma^i, \sigma^j] &= 2i\varepsilon^{ijk}\sigma^k\end{aligned}$$

As we can see from their anticommutation relations, the pauli matrices can play the role of α^i . We now need to show that there are no fourth 2×2 matrix to play the role of β . To do this let's construct a generic 2×2 matrix $A = a_0 1 + a_i \sigma^i$ and do its commutation relation with the pauli matrices and see what coefficients (a_0, a_i) we need to fulfill the commutation relation $\{\alpha^j, \beta\}_{ab} = 0$

$$\begin{aligned}\{A, \sigma^j\} &= \{a_0 1 + a_i \sigma^i, \sigma^j\} \\ 0 &= \{a_0 1, \sigma^j\} + \{a_i \sigma^i, \sigma^j\} \\ 0 &= 2a_0 \sigma^j + a_i \delta^{ij} 1 \\ 0 &= 2a_0 \sigma^j + a_j 1\end{aligned}$$

Recall that there's only one diagonal pauli matrix, *i.e.* if you use \hat{z} as the main direction $\sigma_z = (+1, -1)$ which is just spin up and spin down and the other two pauli matrices will be off diagonal, if you choose any other axis the same will occur, only one will be diagonal and the other two will be off diagonal.

From $0 = 2a_0 \sigma^j + a_j 1$, we need σ^j to be diagonal but since $\sigma^j = (+1, -1)$ the only solution is $a_0 = 0, a_j = 0$. However the anticommutation relation $\{A, \sigma^j\} = 0$ has to be true for all values of $j = 1, 2, 3$ which means that $a_1 = a_2 = a_3 = 0 \rightarrow A = 0$, so the only 2×2 matrix that can fulfill the role of β is the zero matrix. However, $\beta^2 = 1$, hence Dirac matrices must be at least 4×4 .

Srednicki Problem 4.1 Verify eq. (4.12). Verify its limit as $m \rightarrow 0$. Eq. 4.12 is $[\varphi^+(x), \varphi^-(x')] = (m/4\pi^2 r) K_1(mr) \equiv C(r)$ where $K_1(mr)$ is a modified Bessel function of the second kind.

The subject here is that for a transition amplitude to be Lorentz invariance, the above commutator must be zero, however it's not. Why does it have to be zero? because the transition amplitude is given by time ordered product of the fields and for spacelike separations, $x - x' > 0$, the time ordering might shuffle the order of the fields as for a spacelike separation different observers might see two events happen in different orders. Note that the metric used here is $(- + + +)$. Hence the need to set $[\varphi^+(x), \varphi^-(x')] = 0$ to maintain equal transition amplitudes for all observers, $\langle \varphi(x) \varphi(y) \rangle = \langle \varphi(y) \varphi(x) \rangle$.

However, eq. 4.12 shows that $[\varphi^+(x), \varphi^-(x')] \neq 0$! Our task at hand is to derive this conclusion. First we need to choose a frame where $t - t' = 0$ as suggested by the text to simplify calculation.

$$\begin{aligned}
\int \widetilde{dk} e^{ik(x-x')} &= \int \frac{d^3k}{(2\pi)^3 2\sqrt{k^2 + m^2}} e^{(-ik_0 \Delta t + i\vec{k} \cdot \Delta \vec{x})}, \Delta t = 0 \\
&= \int \frac{d^3k}{(2\pi)^3 2\sqrt{k^2 + m^2}} e^{i\vec{k} \cdot \Delta \vec{x}} \\
&= \int \frac{d^3k}{(2\pi)^3 2\sqrt{k^2 + m^2}} e^{i|k||r| \cos \theta}, |r| = |\Delta \vec{x}| \\
&= \int \frac{k^2 dk d\phi d(\cos \theta)}{(2\pi)^3 2\sqrt{k^2 + m^2}} e^{i|k||r| \cos \theta} \\
&= \int \frac{k^2 dk 2\pi}{(2\pi)^3 2\sqrt{k^2 + m^2}} \int_{-1}^1 d(\cos \theta) e^{i|k||r| \cos \theta} \\
&= \int \frac{k^2 dk}{(2\pi)^2 2\sqrt{k^2 + m^2}} \frac{(e^{i|k||r|} - e^{-i|k||r|})}{ikr} \\
&= \int \frac{k dk}{(2\pi)^2 2\sqrt{k^2 + m^2}} \frac{2\sin(kr)}{kr} \\
&= \frac{1}{(2\pi)^2 r} \int_0^\infty dk \frac{k \sin(kr)}{\sqrt{k^2 + m^2}}
\end{aligned}$$

We now need to massage this into a Bessel function, but first let's do a change of variable

$$\rightarrow y = k/m$$

$$\begin{aligned} &= \frac{1}{(2\pi)^2 r} \int_0^\infty m \, d\left(\frac{k}{m}\right) \frac{m \left(\frac{k}{m}\right) \sin\left(\left(\frac{k}{m}\right) mr\right)}{m \sqrt{\left(\frac{k}{m}\right)^2 + 1}} \\ &= \frac{m}{(2\pi)^2 r} \int_0^\infty dy \frac{y \sin(ymr)}{\sqrt{y^2 + 1}} \end{aligned}$$

now, there are 2 facts we need about Bessel functions

$$\begin{aligned} K_0(z) &= \int_0^\infty dy \frac{\cos(zy)}{\sqrt{y^2 + 1}} \\ \frac{d}{dz} K_\nu(z) &= \frac{\nu}{z} K_\nu(z) - K_{\nu+1}(z) \\ \rightarrow \frac{d}{dz} K_0(z) &= -K_1(z) \end{aligned}$$

If we do the actual derivative on K_0 we get

$$\begin{aligned} \rightarrow \frac{d}{dz} K_0(z) &= \int_0^\infty dy \frac{d}{dz} \left(\frac{\cos(zy)}{\sqrt{y^2 + 1}} \right) \\ &= -K_1(z) = - \int_0^\infty dy \frac{y \sin(zy)}{\sqrt{y^2 + 1}} \\ \rightarrow \frac{m}{(2\pi)^2 r} K_1(z) &= \frac{m}{(2\pi)^2 r} \int_0^\infty dy \frac{y \sin(zy)}{\sqrt{y^2 + 1}}, \quad z \rightarrow mr \\ \frac{m}{4\pi^2 r} K_1(mr) &= \frac{m}{(2\pi)^2 r} \int_0^\infty dy \frac{y \sin(mry)}{\sqrt{y^2 + 1}} \\ \frac{m}{4\pi^2 r} K_1(mr) &= \int \widetilde{dk} \, e^{ik(x-x')} \end{aligned}$$

and that is the result that we want. How about its limit as $m \rightarrow 0$? One might be tempted to find or calculate the power series expansion of the Bessel function around $z \rightarrow 0$ but I found it extremely difficult (as I've forgotten most of the methods from my mathematical methods class) but (after looking at the problem long enough) I realized that we need not mess with Bessel functions at all.

You see $\int \widetilde{dk} \, e^{ik(x-x')}$ looks a lot like a fourier transform and setting $m \rightarrow 0$ makes it look like this

$$\begin{aligned} &= \int \frac{d^3 k}{(2\pi)^3 2\sqrt{k^2 + m^2}} e^{i\vec{k} \cdot \vec{r}}, \quad m \rightarrow 0 \\ &= \frac{1}{2(2\pi)^3} \int d^3 k \frac{e^{i\vec{k} \cdot \vec{r}}}{|k|} \end{aligned}$$

which looks a lot like the fourier transform of the Coulomb potential but this time it's going into the \vec{x} domain instead of the \vec{k} domain.

The evaluation of this integral is usually done using a regulator $e^{-\mu k}$ and we'll take the limit $\mu \rightarrow 0$ at the end.

$$\begin{aligned}
&= \frac{1}{2(2\pi)^3} \int d^3k \frac{e^{-\mu k} e^{i\vec{k} \cdot \vec{r}}}{|k|} \\
&= \frac{1}{2(2\pi)^3} \int k^2 dk d\phi \frac{e^{-\mu k}}{k} \int_{-1}^1 d(\cos \theta) e^{i|k||r| \cos \theta} \\
&= \frac{1}{2(2\pi)^2} \int_0^\infty k dk e^{-\mu k} \frac{(e^{ikr} - e^{-ikr})}{i k r} \\
&= \frac{1}{2ir(2\pi)^2} \int_0^\infty dk e^{-\mu k} (e^{ikr} - e^{-ikr}) \\
&= \frac{1}{2ir(2\pi)^2} \left[\frac{1}{(-\mu + ir)} e^{(-\mu + ir)k} \Big|_0^\infty - \frac{1}{(-\mu - ir)} e^{(-\mu - ir)k} \Big|_0^\infty \right] \\
&= \frac{1}{2ir(2\pi)^2} \left[\frac{1}{(-\mu + ir)} (e^{-\infty} - e^0) - \frac{1}{(-\mu - ir)} (e^{-\infty} - e^0) \right] \\
&= \frac{1}{2ir(2\pi)^2} \left[\frac{1}{(-\mu - ir)} - \frac{1}{(-\mu + ir)} \right] = \frac{1}{2ir(2\pi)^2} \left[\frac{-\mu + ir + \mu + ir}{\{(-\mu)^2 - (ir)^2\}} \right] \\
&= \frac{1}{2ir(2\pi)^2} \frac{2ir}{\{(-\mu)^2 + r^2\}}, \mu \rightarrow 0 \\
&= \frac{1}{4\pi^2 r^2}
\end{aligned}$$

which is the result we want :) without messing with Bessel functions at all

How do you do the integral of $\int_{-\infty}^{\infty} dx \frac{\sin x}{x}$? The common solution is to do a contour integration, but is there another way? This was what I attempted, the answer to that question is in short no, but I sure learned a hell lot thinking about this problem.

First, there's no function whose derivative is $\frac{\sin x}{x}$ which means that the indefinite integral $\int dx \frac{\sin x}{x}$ has no solution. Contour integral was a neat technique apropos to a numerical calculation.

My original approach was inspired by green's function method. This was because the integral $\int dx \frac{\sin x}{x}$ is akin to $\int dk \frac{e^{ikx}}{k}$ and since $\cos k$ is even, only the $\sin k$ term will survive the integration $\rightarrow \int dk \frac{i \sin(kx)}{k}$ and by taking $x \rightarrow 1$ we'll get $\int dk \frac{i \sin k}{k}$ which is

what we want. The starting point is the unit step function $u(x)$

$$\begin{aligned}\frac{d}{dx}u(x) &= \delta(x) \\ \rightarrow ik \, u(k) &= 1 \\ u(k) &= \frac{1}{ik}\end{aligned}$$

We must take a pause here. First some trivial things, I have chosen the fourier transform pair to be $f(x) = \frac{1}{2\pi} \int dk e^{ikx} F(k)$ and $F(k) = \int dx e^{-ikx} f(x)$, *i.e.* we put all factor of normalization $\frac{1}{2\pi}$ in the inverse transform. We can choose other normalization schemes, but the result will be the same.

The second reason to pause here is more serious. The unit step function is a peculiar function. Note that

$$\frac{d}{dx}(u(x) + C) = \delta(x)$$

is still true for any constant C . But there's a particular value of C that has serious ramifications, *i.e.* $C = -\frac{1}{2}$. Why? because

$$\left(u(x) - \frac{1}{2}\right) = \frac{1}{2}(u(x) - u(-x))$$

that is $(u(x) - \frac{1}{2})$ is just the antisymmetric part of $u(x)$. Now there is a fact that we need to remember, an antisymmetric (real) function has as its fourier counterpart a purely imaginary function $\int dx e^{-ikx} f(x) = -i \int dx \sin(kx) f(x)$. We see above that $u(k) = \frac{1}{ik}$ (purely imaginary) and since $\frac{d}{dx}u(x) = \frac{d}{dx}(u(x) + C)$ what we were really doing was fourier transforming only the antisymmetric part of $u(x)$, *i.e.*

$$\begin{aligned}\frac{d}{dx} \left(\frac{1}{2}(u(x) - u(-x)) \right) &= \frac{d}{dx} \left(u(x) - \frac{1}{2} \right) = \delta(x) \\ \mathcal{F} \left\{ \frac{d}{dx} \left(u(x) - \frac{1}{2} \right) \right\} &= \mathcal{F} \{ \delta(x) \} \\ ik \, \mathcal{F} \left\{ \left(u(x) - \frac{1}{2} \right) \right\} &= 1 \\ ik \left(u(k) - \frac{1}{2}(2\pi\delta(k)) \right) &= 1 \\ u(k) &= \frac{1}{ik} + \pi\delta(k)\end{aligned}$$

Note that $\mathcal{F}\{\delta(x)\} = 1$ and $\mathcal{F}\{1\} = 2\pi\delta(k)$ is affected by the normalization scheme, other schemes will produce different numerical factors. Now, what does it have to do with $\int dx \frac{\sin x}{x}$? Let's do an inverse fourier transform on the above function

$$\begin{aligned}\mathcal{F}^{-1}\{u(k)\} &= \mathcal{F}^{-1}\left\{\frac{1}{ik} + \pi\delta(k)\right\} \\ u(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \frac{e^{ikx}}{ik} + \frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{ikx} (\pi\delta(k)) \\ u(x) &= \frac{1}{2\pi i} \int_{-\infty}^{\infty} dk \frac{e^{ikx}}{k} + \frac{1}{2}\end{aligned}$$

Now we set $x \rightarrow 1, u(1) = 1$

$$\begin{aligned}u(1) &= \frac{1}{2\pi i} \int_{-\infty}^{\infty} dk \frac{e^{ik}}{k} + \frac{1}{2} \\ 1 &= \frac{1}{2\pi i} \int_{-\infty}^{\infty} dk \frac{e^{ik} - e^{-ik}}{k} + \frac{1}{2} \\ &\rightarrow \int_{-\infty}^{\infty} dk \frac{\sin k}{k} = \pi\end{aligned}$$

In the second line we have used the fact that since $\frac{\cos k}{k}$ is an odd function, its integral is zero. And this is the origin of the factor of π . For the longest time I have wondered how this integral has anything to do with π . But wait, you might ask how in turn did the fourier transform pair acquire this factor of $\frac{1}{2\pi}$ in the first place?

The answer to that question is the gaussian integral $e^{-x^2/2}$. This gaussian function is the eigenfunction of the fourier transform. Let's see how it works

$$\begin{aligned}\int_{-\infty}^{\infty} dx e^{-ikx} e^{-x^2/2} &= \sqrt{2\pi} e^{\frac{(-ik)^2}{4(\frac{1}{2})}} \\ &= \sqrt{2\pi} e^{-k^2/2}\end{aligned}$$

where we have used the gaussian integral formula $\int dx e^{-ax^2+bx+c} = \sqrt{\frac{\pi}{a}} e^{\frac{b^2}{4a}+c}$ and if we do an inverse transform we get

$$\begin{aligned}\int_{-\infty}^{\infty} dk e^{ikx} \left(\sqrt{2\pi} e^{-k^2/2}\right) &= \sqrt{2\pi} \sqrt{2\pi} e^{\frac{(ix)^2}{4(\frac{1}{2})}} \\ &= 2\pi e^{-x^2/2}\end{aligned}$$

Thus

$$\mathcal{F}^{-1}\left\{\mathcal{F}\left\{e^{-x^2/2}\right\}\right\} = 2\pi e^{-x^2/2}$$

or equivalently

$$\begin{aligned}
e^{-(0)^2/2} &\stackrel{?}{=} \int_{-\infty}^{\infty} dk e^{ik(0)} \left(\sqrt{2\pi} e^{-k^2/2} \right) \\
&= \sqrt{2\pi} \int_{-\infty}^{\infty} dk e^{-k^2/2} = \sqrt{2\pi} \sqrt{2\pi} \\
1 &\stackrel{?}{=} 2\pi
\end{aligned}$$

this means that we need to absorb the extra factor of 2π somehow, the usual way is to put a $\frac{1}{2\pi}$ in the inverse transform, another common way is to put a factor of $\frac{1}{\sqrt{2\pi}}$ on both the fourier transform and its inverse. Note that my derivation of $\int \frac{\sin k}{k}$ does not depend on the normalization scheme, you might try it if you don't believe me ;)

But then again, how did the gaussian integral get a factor of π , well it's because we use a circle to compute it, remember?

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{-x^2/2} &= \sqrt{\left(\int_{-\infty}^{\infty} e^{-x^2/2} \int_{-\infty}^{\infty} e^{-y^2/2} \right)} \\
&= \sqrt{\int_{-\infty}^{\infty} dx dy e^{-(x^2+y^2)/2}} \\
&= \sqrt{\left(\int_0^{2\pi} d\phi \right) \int_0^{\infty} r dr e^{-(r^2)/2}} \\
\int_{-\infty}^{\infty} e^{-x^2/2} &= \sqrt{2\pi}
\end{aligned}$$

And that my friend, is how you get a factor of π in the first place :) Of course in the contour integration of $\int \frac{\sin x}{x}$ the factor of π is due to the infinitesimal semi circle around $x = 0$ but again, it's all due to a circle.

It took me two days to figure this out and I learned a lot about the step function and fourier transform in the process. It all started with a simple thought of whether or not it's possible to compute $\int \frac{\sin x}{x}$ without contour integration, turns out that it's not quite possible, sometimes we do need a new technique to solve an old problem and I gained a new respect for complex numbers.

Why do we need limits and what are they? Let's start with the infamous $\lim_{x \rightarrow 0}$. Without messing with δ and ϵ as in the usual formal definition of a limit, what a limit amounts to

is (in the case $x \rightarrow 0^+$, *i.e.* we're approaching the limit from the right) $\lim_{x \rightarrow 0^+} \Rightarrow x > 0$ but x is smaller than any positive real number.

Now how can that be? If it is smaller than any positive real number (and it's not negative) then it has to be zero. One way it is possible if we are dealing with infinities, something like

$$x = 0.\underbrace{000000 \dots 00000}_{\text{infinite \# of zeros}}1$$

There are an infinite number of zeros before the digit 1 so that it doesn't matter how small the real number you present me, I can always produce a smaller one by adding more zeros in between and it has to be infinite for this to work because x is smaller than *any* positive real number. But this won't work since it is just zero! Remember that

$$1 = 0.\underbrace{999999 \dots 99999}_{\text{infinite \# of 9's}}$$

Thus limits are a (more severe?) manifestation of infinities. The real question is: why do we need it? The more obvious reason is in the calculation of a slope. To calculate a slope you need at least two points, say (x_1, y_1) and (x_2, y_2) and the slope is

$$\frac{\Delta y}{\Delta x} = \frac{(y_2 - y_1)}{(x_2 - x_1)}$$

but then whose slope is it? Is it the slope at (x_1, y_1) or at (x_2, y_2) ? because slopes at different points are usually different. What we want is to calculate the slope at *one* point but we need two points to calculate it. How do we solve this conundrum? The answer is of course limits! If we take the limit $\Delta x = (x_2 - x_1) \rightarrow 0^+$ we are practically using two distinct points located at the 'same' place because x_2 is *not* x_1 and yet it is closer to x_1 than any other points.

In closing, in the words of my high school math teacher, "limits bring into existence things that don't exist" and I believe that what real numbers do. And that might be the *real* reason behind its difficulties, real numbers were supposed to address the problem of continuum in mathematics although physical continuum may not actually exist, maybe "what God has not created we shouldn't bring about" :)



FIG. 2: A choice of contour that includes a cut.

Branch cuts, what are they for? and how do we use them? For an illustration see Fig. 2. What we want to do here is to calculate $\oint_C \sqrt{z} dz$ along a unit circle around the origin. The tricky thing here is that $1 = e^{i0} = e^{i2\pi}$ but $\sqrt{1} \rightarrow e^{i0/2} \neq e^{i\pi}$! So we need to provide a cut such that the contour doesn't go full circle. First we will calculate $\oint_C \sqrt{z} dz$ using the contour shown in Fig. 2 (left).

$$\begin{aligned}
\oint_C \sqrt{z} dz &= \oint_{C_1} \sqrt{z} dz + \int_{L_1} \sqrt{z} dz + \oint_{C_\epsilon} \sqrt{z} dz + \int_{L_2} \sqrt{z} dz \\
&\rightarrow \oint_C \sqrt{z} dz = 0, \quad \text{no pole inside the contour} \\
&\rightarrow \oint_{C_1} \sqrt{z} dz = \int_\epsilon^{2\pi-\epsilon} \sqrt{1e^{i\theta}} d\theta = \int_0^{2\pi} \sqrt{e^{i\theta}} d\theta = \oint_C \sqrt{z} dz \\
&\rightarrow \int_{L_1} \sqrt{z} dz = \int_1^\epsilon \sqrt{x e^{(2\pi-\epsilon)i}} dx = \int_1^0 \sqrt{x} e^{\pi i} dx = \int_0^1 \sqrt{x} dx \\
&\rightarrow \oint_{C_\epsilon} \sqrt{z} dz = \int_{2\pi-\epsilon}^\epsilon \sqrt{\epsilon e^{i\theta}} d\theta = 0 \text{ as } \epsilon \rightarrow 0 \\
&\rightarrow \int_{L_2} \sqrt{z} dz = \int_\epsilon^1 \sqrt{x e^{\epsilon i}} dx = \int_0^1 \sqrt{x} dx
\end{aligned}$$

and we take the limit of $\varepsilon \rightarrow 0$. Thus we have, after rearranging

$$\begin{aligned}
\oint_C \sqrt{z} dz &= \oint_{C_1} \sqrt{z} dz + \int_{L_1} \sqrt{z} dz + \oint_{C_\varepsilon} \sqrt{z} dz + \int_{L_2} \sqrt{z} dz \\
&\rightarrow 0 = \oint_C \sqrt{z} dz + \int_0^1 \sqrt{x} dx + 0 + \int_0^1 \sqrt{x} dx \\
\oint_C \sqrt{z} dz &= - \int_0^1 \sqrt{x} dx - \int_0^1 \sqrt{x} dx \\
&= -\frac{4}{3}
\end{aligned}$$

but what happens if we choose a cut that doesn't lie on the x axis? say the cut is at an angle Δ ? as shown in Fig. 2 (right), note that we want L_1 and L_2 to differ by 2π otherwise there's no point in introducing a cut, here L_2 is at angle Δ and L_1 is at angle $\Delta + 2\pi$

$$\begin{aligned}
\oint_{C^\Delta} \sqrt{z} dz &= \oint_{C_{1\Delta}} \sqrt{z} dz + \int_{L_{1\Delta}} \sqrt{z} dz + \oint_{C_{\varepsilon\Delta}} \sqrt{z} dz + \int_{L_{2\Delta}} \sqrt{z} dz \\
&\rightarrow \oint_{C^\Delta} \sqrt{z} dz = 0, \text{ no pole inside the contour} \\
&\rightarrow \oint_{C_{1\Delta}} \sqrt{z} dz = \int_{\Delta+\varepsilon}^{\Delta+2\pi-\varepsilon} \sqrt{1e^{i\theta}} d\theta = \int_{\Delta}^{\Delta+2\pi} \sqrt{e^{i\theta}} d\theta = \oint_{C^\Delta} \sqrt{z} dz \\
&\rightarrow \int_{L_{1\Delta}} \sqrt{z} dz = \int_1^\varepsilon \sqrt{re^{(\Delta+2\pi-\varepsilon)i}} dr = \int_1^0 \sqrt{r} e^{i\Delta/2} e^{\pi i} dr = e^{i\Delta/2} \int_0^1 \sqrt{r} dr \\
&\rightarrow \oint_{C_{\varepsilon\Delta}} \sqrt{z} dz = \int_{\Delta+2\pi-\varepsilon}^{\Delta+\varepsilon} \sqrt{\varepsilon e^{i\theta}} d\theta = 0 \text{ as } \varepsilon \rightarrow 0 \\
&\rightarrow \int_{L_{2\Delta}} \sqrt{z} dz = \int_\varepsilon^1 \sqrt{re^{(\Delta+\varepsilon)i}} dr = e^{i\Delta/2} \int_0^1 \sqrt{r} dr
\end{aligned}$$

Now we have

$$\begin{aligned}
\oint_{C^\Delta} \sqrt{z} dz &= \oint_{C_{1\Delta}} \sqrt{z} dz + \int_{L_{1\Delta}} \sqrt{z} dz + \oint_{C_{\varepsilon\Delta}} \sqrt{z} dz + \int_{L_{2\Delta}} \sqrt{z} dz \\
0 &= \oint_{C^\Delta} \sqrt{z} dz + e^{i\Delta/2} \int_0^1 \sqrt{r} dr + 0 + e^{i\Delta/2} \int_0^1 \sqrt{r} dr \\
\oint_{C^\Delta} \sqrt{z} dz &= -2e^{i\Delta/2} \int_0^1 \sqrt{r} dr \\
&= e^{i\Delta/2} \left(-\frac{4}{3} \right)
\end{aligned}$$

So apparently we get a different result for different contours, but if we take a closer look

$$\begin{aligned}
\oint_{C^\Delta} \sqrt{z} dz &= \int_{\Delta}^{\Delta+2\pi} \sqrt{1e^{i\theta}} d\theta \\
\text{Let } \omega &= \theta - \Delta \rightarrow d\omega = d\theta, \quad \int_{\Delta}^{\Delta+2\pi} d\theta = \int_0^{2\pi} d\omega \\
\rightarrow \oint_{C^\Delta} \sqrt{z} dz &= \int_0^{2\pi} \sqrt{1e^{i(\omega+\Delta)}} d\omega \\
&= e^{i\Delta/2} \int_0^{2\pi} \sqrt{1e^{i\omega}} d\omega \\
\oint_{C^\Delta} \sqrt{z} dz &= e^{i\Delta/2} \oint_C \sqrt{z} dz
\end{aligned}$$

so indeed

$$\begin{aligned}
\oint_{C^\Delta} \sqrt{z} dz &= e^{i\Delta/2} \oint_C \sqrt{z} dz = e^{i\Delta/2} \left(-\frac{4}{3} \right) \\
&\rightarrow \oint_C \sqrt{z} dz = -\frac{4}{3}
\end{aligned}$$

Thus rotating the branch cut (by an angle Δ) does indeed rotate the original integral $\oint_C \sqrt{z} dz$ into $e^{i\Delta/2} \oint_C \sqrt{z} dz$, so we need to be careful.

A side note, how do we specify the contour if the circle is not at the origin but say is at z_0 ? Well, we can just parametrize the circle as $z = z_0 + re^{i\theta}$ (remember that additions in the complex plane are vectorial operations), thus a contour integral of z around a point z_0 is simply given by

$$\oint z dz \rightarrow \int_0^{2\pi} (z_0 + re^{i\theta}) d\theta$$

Going back to branch cuts, what if we have more than one root, *e.g.* $\sqrt{(z - z_1)(z - z_2)}$? see Fig. 3. Then there are two choices, we can put the cut between the roots as shown in Fig. 3 (left) or we can put the cuts starting from the roots to infinity as shown in Fig. 3 (right), the point here is to prevent a full circle around each root.

Throwback Tuesday! (today is Tuesday Oct 13 2015) High school favorite, how do we compute $\sin(18^\circ)$ without approximation and without a calculator? Well, start with $\sin(18^\circ) = \sin(90^\circ/5)$, since we know that $\sin(90^\circ) = 1$. However, our real starting point is actually $\cos(90^\circ/5)$ because $\cos(90^\circ) = 0$ and the end goal here is to transform $\sin(18^\circ)$



FIG. 3: Different choices of contours that include a cut with 2 roots.

into solving a polynomial and it's easier to find the zeros of a polynomial if there's no constant.

First we need the formula for $\cos(5x)$. The easiest way to do this is to use Euler's formula

$$e^{i5x} = (e^{ix})^5$$

$$\rightarrow \cos(5x) + i \sin(5x) = (\cos(x) + i \sin(x))^5$$

In case you forget how to expand a polynomial, here's the Pascal's triangle

$$\begin{array}{ccccccc} & & & & 1 & & & \\ & & & & & & & \\ & & 1 & 2 & 1 & & & \\ & 1 & 3 & 3 & 1 & & & \\ & 1 & 4 & 6 & 4 & 1 & & \\ 1 & 5 & 10 & 10 & 5 & 1 & & \end{array}$$

the expansion is then

$$\begin{aligned}
(\cos(x) + i \sin(x))^5 &= \cos^5(x) + 5 \cos^4(x) i \sin(x) + 10 \cos^3(x) i^2 \sin^2(x) + \\
&\quad 10 \cos^2(x) i^3 \sin^3(x) + 5 \cos(x) i^4 \sin^4(x) + i^5 \sin^5(x) \\
\rightarrow \cos(5x) &= \operatorname{Re} \{ (\cos(x) + i \sin(x))^5 \} \\
&= \cos^5(x) - 10 \cos^3(x) \sin^2(x) + 5 \cos(x) \sin^4(x) \\
\rightarrow \sin(5x) &= \operatorname{Im} \{ (\cos(x) + i \sin(x))^5 \} \\
&= 5 \cos^4(x) \sin(x) - 10 \cos^2(x) \sin^3(x) + \sin^5(x)
\end{aligned}$$

simplifying

$$\begin{aligned}
\cos(5x) &= \cos^5(x) - 10 \cos^3(x) \sin^2(x) + 5 \cos(x) \sin^4(x) \\
&= \cos^5(x) - 10 \cos^3(x) (1 - \cos^2(x)) + 5 \cos(x) (1 - \cos^2(x))^2 \\
&= \cos^5(x) - 10 \cos^3(x) + 10 \cos^5(x) + 5 \cos(x) (1 - 2 \cos^2(x) + \cos^4(x)) \\
&= 11 \cos^5(x) - 10 \cos^3(x) + 5 \cos(x) - 10 \cos^3(x) + 5 \cos^5(x) \\
\cos(5x) &= 16 \cos^5(x) - 20 \cos^3(x) + 5 \cos(x)
\end{aligned}$$

while for sin

$$\begin{aligned}
\sin(5x) &= 5 \cos^4(x) \sin(x) - 10 \cos^2(x) \sin^3(x) + \sin^5(x) \\
&= 5(1 - \sin^2(x))^2 \sin(x) - 10(1 - \sin^2(x)) \sin^3(x) + \sin^5(x) \\
&= 5(1 - 2 \sin^2(x) + \sin^4(x)) \sin(x) - 10 \sin^3(x) + 10 \sin^5(x) + \sin^5(x) \\
&= 5 \sin(x) - 10 \sin^3(x) + 5 \sin^5(x) - 10 \sin^3(x) + 11 \sin^5(x) \\
\sin(5x) &= 16 \sin^5(x) - 20 \sin^3(x) + 5 \sin(x)
\end{aligned}$$

We can of course derive the above result using the usual formulae, for example

$$\begin{aligned}
\sin(5x) &= \sin(2x + 3x) \\
&= \sin(2x) \cos(3x) + \cos(2x) \sin(3x) \\
\sin(2x) &= 2 \sin(x) \cos(x) \\
\cos(3x) &= \cos(x + 2x) = \cos(x) \cos(2x) - \sin(x) \sin(2x) \\
\cos(2x) &= \cos^2(x) - \sin^2(x) = 1 - 2 \sin^2(x) \\
\sin(3x) &= \sin(x + 2x) = \sin(x) \cos(2x) + \cos(x) \sin(2x)
\end{aligned}$$

and working out the algebra, the result will be the same. Now we see as to why starting out with $\sin(18^\circ) = \sin(90^\circ/5)$ doesn't work well

$$\begin{aligned}\sin(5 \times 18^\circ) &= 1 = 16 \sin^5(18^\circ) - 20 \sin^3(18^\circ) + 5 \sin(18^\circ) \\ 1 &= 16 \sin^5(18^\circ) - 20 \sin^3(18^\circ) + 5 \sin(18^\circ)\end{aligned}$$

the LHS is not zero, while starting with $\cos(18^\circ) = \cos(90^\circ/5)$ does work

$$\begin{aligned}\cos(5 \times 18^\circ) &= 0 = 16 \cos^5(18^\circ) - 20 \cos^3(18^\circ) + 5 \cos(18^\circ) \\ 0 &= \cos(18^\circ)(16 \cos^4(18^\circ) - 20 \cos^2(18^\circ) + 5), \quad \text{Let } x = \cos^2(18^\circ) \\ 0 &= 16x^2 - 20x + 5 \\ x &= \frac{20 \pm \sqrt{400 - 320}}{32} = \frac{20 \pm 4\sqrt{5}}{32}\end{aligned}$$

we just need to take the + of the \pm since cos of small angles are closer to one. The result we want is then given by

$$\begin{aligned}\sin(18^\circ) &= \sqrt{\sin^2(18^\circ)} = \sqrt{1 - \cos^2(18^\circ)} = \sqrt{1 - x} \\ &= \sqrt{1 - \frac{20 + 4\sqrt{5}}{32}} \\ \sin(18^\circ) &= \sqrt{\frac{3 - \sqrt{5}}{8}}\end{aligned}$$

but wait, how do you calculate the square root by hand? The problem states that we should calculate it without a calculator.

To calculate square roots by hand we use the ubiquitous formula $(a+b)^2 = a^2 + 2ab + b^2$ and do it iteratively. Say we want to find the root of a number, *e.g.* \sqrt{C} , first we do a rough approximation $C_1^2 < C$ and then we find the next factor C_2 through

$$\begin{aligned}C &= (C_1 + C_2)^2 = C_1^2 + 2C_1C_2 + C_2^2 \\ (C - C_1^2) &= (2C_1 + C_2)C_2\end{aligned}$$

finding the largest C_2 such that $(2C_1 + C_2)C_2 < (C - C_1^2)$. Once we find C_2 we repeat the process

$$\begin{aligned}C &= (\{C_1 + C_2\} + C_3)^2 = \{C_1 + C_2\}^2 + 2\{C_1 + C_2\}C_3 + C_3^2 \\ (C - \{C_1 + C_2\}^2) &= (2\{C_1 + C_2\} + C_3)C_3\end{aligned}$$

by finding the largest C_3 such that $(2\{C_1 + C_2\} + C_3)C_3 < (C - \{C_1 + C_2\}^2)$. In other words

$$C_1 \rightarrow \text{largest number such that } C_1^2 < C$$

$$C_n \rightarrow \text{largest number such that } \left(2 \left(\sum_{i=1}^{n-1} C_i\right) + C_n\right) C_n < \left(C - \left(\sum_{i=1}^{n-1} C_i\right)^2\right)$$

$$\rightarrow \sqrt{C} = \sum_{i=1}^{\infty} C_i$$

You can stop anytime you want, depending on the accuracy you need. The above prescription is correct but we want to make it more systematic by dividing the number into ‘hundreds’, partitioning it into 2-digit groupings (hundreds because our digits run from 0 to 9 which translate to 0 to < 100). Let’s see how it works with a concrete example

$$\begin{array}{rcl}
& & \begin{array}{r} 2 \ 3 \ 8. \ 3 \ 0 \ 4 \ 4 \\ \sqrt{5 \ 67 \ 89. \ 01 \ 20 \ 00 \ 00} \\ \hline 4 \end{array} \\
2 \times 2 = & & \begin{array}{r} 1 \ 67 \\ \hline 1 \ 29 \end{array} - \\
((2)(10)(2) + 3) \times 3 = & & \begin{array}{r} 38 \ 89 \\ \hline 37 \ 44 \end{array} - \\
((2)(10)(23) + 8) \times 8 = & & \begin{array}{r} 1 \ 45 \ 01 \\ \hline 1 \ 42 \ 89 \end{array} - \\
((2)(10)(238) + 3) \times 3 = & & \begin{array}{r} 2 \ 12 \ 20 \\ \hline 0 \ 00 \ 00 \end{array} - \\
((2)(10)(2383) + 0) \times 0 = & & \begin{array}{r} 2 \ 12 \ 20 \ 00 \\ \hline 1 \ 90 \ 64 \ 16 \end{array} - \\
((2)(10)(23830) + 4) \times 4 = & & \begin{array}{r} 21 \ 55 \ 84 \ 00 \\ \hline 19 \ 06 \ 43 \ 36 \end{array} - \\
((2)(10)(238304) + 4) \times 4 = & &
\end{array}$$

Thus $\sqrt{56789.012} = 238.3044 \dots$

What we need to take note above is that C_1 takes care only the first digit, this makes it a lot easier to start the procedure, rather than trying to figure out 238 from the start.

In the second iteration, we multiply C_1 by 10 and at the same time ‘multiplying’ the target by 100 by including the next 2 digits as

$$100C = 100(C_1 + C_2)^2 = 100C_1^2 + 200C_1C_2 + 100C_2^2$$

$$100(C - C_1^2) = (2(10)C_1 + C_2)C_2$$

Another important point is that $C_5 = 0$! Zero is apparently permitted. I actually learned this method when I was 10 but didn’t understand the reasoning behind it.

Another way of computing square roots is using taylor expansion

$$\sqrt{1+x} = 1 + \frac{1}{2}x + \sum_{n=1}^{\infty} \frac{(-1)^n(2n-1)!!}{2^n} x^{n+1}$$

for example

$$\sqrt{101} = 10\sqrt{1+0.01} = 10 \left(1 + \frac{1}{2}(0.01) - \frac{1}{2}(0.01)^2 + \frac{(1)(3)}{2^2}(0.01)^3 + \dots \right)$$

but the convergence might be slow due to the alternating signs in the series. One important note here is that although the previous method looks a lot like the usual division algorithm, they are extremely different. In the case of roots, you need to keep using the previous results to compute the next one, not so with divisions, this is because divisions are linear while roots are not.

Why does a number that is a multiple of 9 have digits that sum up to be a multiple of 9? *e.g.*

$$3 \times 9 = 27 \rightarrow 2 + 7 = 9$$

$$13 \times 9 = 117 \rightarrow 1 + 1 + 7 = 9$$

$$5555 \times 9 = 49995 \rightarrow 4 + 9 + 9 + 9 + 5 = 36 = 4 \times 9$$

First of all, any number can be written as

$$a_N a_{N-1} a_{N-2} \dots a_2 a_1 a_0 = \sum_{i=0}^N a_i 10^i, \quad a_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

e.g. $10937 \rightarrow a_4 = 1, a_3 = 0, a_2 = 9, a_1 = 3, a_0 = 7$, now say we have a number that is a multiple of 9, $b = 9n$, with n a positive integer

$$\begin{aligned} b = 9n &= \sum_{i=0}^N a_i 10^i \\ \sum_{i=0}^N a_i 10^i &= \sum_{i=1}^N a_i (10^i - 1) + \sum_{i=0}^N a_i \end{aligned}$$

but we can rewrite

$$\begin{aligned}
(10^i - 1) &= \sum_{j=0}^{i-1} 9(10)^j = 9 \sum_{j=0}^{i-1} (10)^j \\
\rightarrow \sum_{i=0}^N a_i 10^i &= \sum_{i=1}^N a_i \left(9 \sum_{j=0}^{i-1} (10)^j \right) + \sum_{i=0}^N a_i \\
b = 9n &= 9 \left(\sum_{i=1}^N a_i \left(\sum_{j=0}^{i-1} (10)^j \right) \right) + \sum_{i=0}^N a_i \\
9n - 9 \left(\sum_{i=1}^N a_i \left(\sum_{j=0}^{i-1} (10)^j \right) \right) &= \sum_{i=0}^N a_i
\end{aligned}$$

Thus the sum of the digits $\sum_{i=0}^N a_i = 9m$ is a multiple of nine, *e.g.* $41076 = 9 \times 4654$

$$\begin{aligned}
9 \times 4654 &= 41076 = 4 \times 10000 + 1 \times 1000 + 0 \times 100 + 7 \times 10 + 6 \\
&= (4 \times (10000 - 1) + 1 \times (1000 - 1) + 0 \times (100 - 1) + 7 \times (10 - 1)) + \\
&\quad (4 + 1 + 0 + 7 + 6) \\
&= (4 \times 9999 + 1 \times 999 + 0 \times 99 + 7 \times 9) + (4 + 1 + 0 + 7 + 6) \\
9 \times 4654 &= 9 \times (4 \times 1111 + 1 \times 111 + 0 \times 11 + 7 \times 1) + (4 + 1 + 0 + 7 + 6)
\end{aligned}$$

Hence $(4 + 1 + 0 + 7 + 6)$ must be a multiple of 9.

Note that the above argument is also valid for multiples of 3, *i.e.* digits of a number that is a multiple of 3 sum up to a multiple of 3.

What is the explicit formula for $\sum_{i=1}^N i^2$? I saw this on the bus on my way to work a long time ago, took me a day to finally figure it out. You can say that we can use induction to prove the formula however, induction only allows you to prove what you already know *not* to derive the formula itself in the first place! *inductio no ex nihilo* :)

The easiest way to derive the formula for $\sum_{i=1}^N i^2$ is to actually draw it, see Fig. 4. The series we want to sum is shown in the top red box, *i.e.* $\sum_{i=1}^{N=6} i^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2$.

This sum $\sum_{i=1}^{N=6} i^2$ is also represented by the upper red triangle in Fig. 4 because that is the definition of squares, *e.g.* $2^2 = 2 + 2$, $3^2 = 3 + 3 + 3$. However, Fig. 4 also shows

$$\begin{array}{c}
N = 6 \\
\boxed{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2} \\
\updownarrow \\
\begin{array}{c}
\boxed{1 + 2 + 3 + 4 + 5 + 6} \\
\boxed{1 + 2 + 3 + 4 + 5 + 6} \\
\boxed{1 + 2 + 3 + 4 + 5 + 6} \\
\boxed{1 + 2 + 3 + 4 + 5 + 6} \\
\boxed{1 + 2 + 3 + 4 + 5 + 6} \\
\boxed{1 + 2 + 3 + 4 + 5 + 6}
\end{array} \\
c = \sum_{j=1}^{N-1} \frac{j(j+1)}{2} \\
a = N \left(\sum_{k=1}^N k \right)
\end{array}$$

FIG. 4: An illustration on how to calculate sum of squares.

how we can calculate this upper red triangle, it is nothing but the sum in the green box a minus the sum in the lower blue triangle c .

The sum in the green box is given by $a = 6 \times (1 + 2 + 3 + 4 + 5 + 6) = N \sum_{k=1}^{N=6} k$, the lower blue triangle is given by $c = \sum_{j=1}^{N-1} \sum_{m=1}^j m = \sum_{j=1}^{N-1} j(j+1)/2$, where we have used the famous gauss formula for $\sum_{m=1}^N m = N(N+1)/2$.

The sum $\sum_{i=1}^N i^2$ is then given by

$$\begin{aligned}
\sum_{i=1}^N i^2 &= a - c \\
&= \left(N \sum_{k=1}^N k \right) - \left(\sum_{j=1}^{N-1} \frac{j(j+1)}{2} \right) \\
&= \left(N \frac{N(N+1)}{2} \right) - \left(\frac{1}{2} \sum_{j=1}^{N-1} j^2 + \frac{1}{2} \sum_{j=1}^{N-1} j \right) \\
\sum_{i=1}^N i^2 &= \left(N \frac{N(N+1)}{2} \right) - \left(\left\{ \frac{1}{2} \sum_{j=1}^N j^2 - \frac{1}{2} N^2 \right\} + \frac{1}{2} \sum_{j=1}^{N-1} j \right) \\
\sum_{i=1}^N i^2 + \frac{1}{2} \sum_{j=1}^N j^2 &= \left(\frac{N^3 + N^2}{2} \right) - \left(-\frac{1}{2} N^2 + \frac{1}{2} \frac{(N-1)N}{2} \right) \\
&= \frac{N^3 + N^2}{2} + \frac{1}{2} N^2 - \frac{(N-1)N}{4} \\
\frac{3}{2} \sum_{i=1}^N i^2 &= \frac{2N^3 + 2N^2 + 2N^2 - N^2 + N}{4} \\
&= \frac{2N^3 + 3N^2 + N}{4} = \frac{N(N+1)(2N+1)}{4} \\
\rightarrow \sum_{i=1}^N i^2 &= \frac{N(N+1)(2N+1)}{6}
\end{aligned}$$

And that is the formula we are after!

Bonus !!!

We can use the above result to calculate c explicitly in terms of N

$$\begin{aligned}
c &\equiv \sum_{j=1}^N \frac{j(j+1)}{2} = \frac{1}{2} \sum_{j=1}^N j^2 + \frac{1}{2} \sum_{j=1}^N j \\
&= \frac{1}{2} \frac{2N^3 + 3N^2 + N}{6} + \frac{1}{2} \frac{N^2 + N}{2} \\
&= \frac{2N^3 + 6N^2 + 4N}{12} = \frac{N^3 + 3N^2 + 2N}{6} \\
\sum_{j=1}^N \frac{j(j+1)}{2} &= \frac{N^3 + 3N^2 + 2N}{6}
\end{aligned}$$

Now, we can use this result to get something interesting. Note that the blue lower triangle

c in Fig. 4 can be seen as a sum in the vertical direction, *i.e.*

$$\begin{aligned}
c &= 1(5) + 2(4) + 3(3) + 4(2) + 5(1) \\
&= 1(5 - 0) + 2(5 - 1) + 3(5 - 2) + 4(5 - 3) + 5(5 - 4) \\
&= 5(1 + 2 + 3 + 4 + 5) - (2(1) + 3(3) + 4(2) + 5(1)) \\
\frac{N^3 + 3N^2 + 2N}{6} &= \left(N \sum_{i=1}^N i \right) - \sum_{i=2}^N i(i-1), \quad N = 5
\end{aligned}$$

where in the last line we have used the explicit formula for c . Switching sides and generalizing to any N

$$\begin{aligned}
\sum_{i=2}^N i(i-1) &= \left(N \sum_{i=1}^N i \right) - \frac{N^3 + 3N^2 + 2N}{6} \\
&= \frac{N^3 + N^2}{2} - \frac{N^3 + 3N^2 + 2N}{6} \\
&= \frac{3N^3 + 3N^2 - N^3 - 3N^2 - 2N}{6} \\
&= \frac{2N^3 - 2N}{6} \\
\sum_{i=2}^N i(i-1) &= \frac{N^3 - N}{3}
\end{aligned}$$

We can now in turn use this result to calculate $1 + 2(1) + 3(2) + 4(3) + \dots + N(N-1)$, *i.e.*

$$1 + \sum_{i=2}^N i(i-1) = 1 + \sum_{i=2}^N \frac{i!}{(i-2)!} = \frac{N^3 - N}{3} + 1$$

Looking back at how we calculated $\sum_{i=1}^N i^2$ using Fig. 4, we can easily generalize it to calculate $\sum_{i=1}^N i^k$, $k > 2$, see Fig. 5. To simplify things let's designate $f_k(N) \equiv \sum_{i=1}^N i^k$ as the sum of k -th powers.

From Fig. 5 it's immediately apparent that we can calculate $f_k(N)$ through

$$\begin{aligned}
f_k(N) &= N f_{k-1}(N) - \sum_{m=1}^{N-1} f_{k-1}(m) \\
&= N f_{k-1}(N) - \left(\sum_{m=1}^N f_{k-1}(m) - f_{k-1}(N) \right) \\
f_k(N) &= (N+1) f_{k-1}(N) - \sum_{m=1}^N f_{k-1}(m)
\end{aligned}$$

$$\begin{array}{c}
N=6 \\
1^n + 2^n + 3^n + 4^n + 5^n + 6^n \\
\Downarrow \\
\begin{array}{c}
1^{n-1} + 2^{n-1} + 3^{n-1} + 4^{n-1} + 5^{n-1} + 6^{n-1} \\
+ 2^{n-1} + 3^{n-1} + 4^{n-1} + 5^{n-1} + 6^{n-1} \\
+ 3^{n-1} + 4^{n-1} + 5^{n-1} + 6^{n-1} \\
+ 4^{n-1} + 5^{n-1} + 6^{n-1} \\
+ 5^{n-1} + 6^{n-1} \\
+ 6^{n-1}
\end{array} \\
\Downarrow \\
\begin{array}{c}
N \\
\text{rows} \left[\begin{array}{l}
1^{n-1} + 2^{n-1} + 3^{n-1} + 4^{n-1} + 5^{n-1} + 6^{n-1} \\
1^{n-1} + 2^{n-1} + 3^{n-1} + 4^{n-1} + 5^{n-1} + 6^{n-1} - [1^{n-1}] \\
1^{n-1} + 2^{n-1} + 3^{n-1} + 4^{n-1} + 5^{n-1} + 6^{n-1} - [1^{n-1} + 2^{n-1}] \\
1^{n-1} + 2^{n-1} + 3^{n-1} + 4^{n-1} + 5^{n-1} + 6^{n-1} - [1^{n-1} + 2^{n-1} + 3^{n-1}] \\
1^{n-1} + 2^{n-1} + 3^{n-1} + 4^{n-1} + 5^{n-1} + 6^{n-1} - [1^{n-1} + 2^{n-1} + 3^{n-1} + 4^{n-1}] \\
1^{n-1} + 2^{n-1} + 3^{n-1} + 4^{n-1} + 5^{n-1} + 6^{n-1} - [1^{n-1} + 2^{n-1} + 3^{n-1} + 4^{n-1} + 5^{n-1}]
\end{array} \right.
\end{array}
\end{array}$$

FIG. 5: An illustration on how to calculate sum of n -th powers.

which is a recursion relation for the sum of any k -th powers.

Let's see a concrete example by calculating $f_3(N) = \sum_{i=1}^N i^3$

$$\begin{aligned}
f_3(N) &= (N+1)f_2(N) - \sum_{m=1}^N f_2(m) \\
&= (N+1) \frac{2N^3 + 3N^2 + N}{6} - \sum_{m=1}^N \frac{2m^3 + 3m^2 + m}{6} \\
\sum_{i=1}^N i^3 + \frac{2}{6} \sum_{m=1}^N m^3 &= \frac{2N^4 + 5N^3 + 4N^2 + N}{6} - \frac{2N^3 + 3N^2 + N}{12} - \frac{N^2 + N}{12} \\
\frac{4}{3} \sum_{i=1}^N i^3 &= \frac{4N^4 + 8N^3 + 4N^2}{12} \\
&= \frac{N^4 + 2N^3 + N^2}{3} = \frac{N^2(N+1)^2}{3} \\
\rightarrow \sum_{i=1}^N i^3 &= \frac{N^2(N+1)^2}{4}
\end{aligned}$$

so *theoretically* we can compute the explicit formula for the sum of any k -th power using

$$f_k(N) = (N+1)f_{k-1}(N) - \sum_{m=1}^N f_{k-1}(m)$$

(with $f_0(N) = \sum_{i=1}^N 1 = N$) which is the gist of the algorithm. The following list shows $f_k(N)$ up to $k = 15$ when the expression starts becoming unwieldy, but there's a pattern, the even k starts with $N(N+1)(2N+1)$ while the odd k starts with $N^2(N+1)^2$.

$$\begin{aligned}
f_0 &= N \\
f_1 &= \frac{N^2+N}{2} &= \frac{N(N+1)}{2} \\
f_2 &= \frac{2N^3+3N^2+N}{6} &= \frac{N(N+1)(2N+1)}{6} \\
f_3 &= \frac{N^4+2N^2+N}{4} &= \frac{N^2(N+1)^2}{4} \\
f_4 &= \frac{6N^5+15N^4+10N^3-N}{30} &= \frac{N(N+1)(2N+1)(3N^2+3N-1)}{30} \\
f_5 &= \frac{2N^6+6N^5+5N^4-N^2}{12} &= \frac{N^2(N+1)^2(2N^2+2N-1)}{12} \\
f_6 &= \frac{6N^7+21N^6+21N^5-7N^3+N}{42} &= \frac{N(N+1)(2N+1)(3N^4+6N^3-3N+1)}{42} \\
f_7 &= \frac{3N^8+12N^7+14N^6-7N^4+2N^2}{24} &= \frac{N^2(N+1)^2(3N^4+6N^3-N^2-4N+2)}{24} \\
f_8 &= \frac{10N^9+45N^8+60N^7-42N^5+20N^3-3N}{90} &= \frac{N(N+1)(2N+1)(5N^6+15N^5+5N^4-15N^3-N^2+9N-3)}{90} \\
f_9 &= \frac{2N^{10}+10N^9+15N^8-14N^6+10N^4-3N^2}{20} &= \frac{N^2(N+1)^2(N^2+N-1)(2N^4+4N^3-N^2-3N+3)}{20} \\
f_{10} &= \frac{6N^{11}+33N^{10}+55N^9-66N^7+66N^5-33N^3+5N}{66} &= \frac{N(N+1)(2N+1)(N^2+N-1)(3N^6+9N^5+2N^4-11N^3+3N^2+10N-5)}{66} \\
f_{11} &= \frac{2N^{12}+12N^{11}+22N^{10}-33N^8+44N^6-33N^4+10N^2}{24} &= \frac{N^2(N+1)^2(2N^8+8N^7+4N^6-16N^5-5N^4+26N^3-3N^2-20N+10)}{24} \\
f_{12} &= \frac{210N^{13}+1365N^{12}+2730N^{11}-5005N^9+8580N^7-9009N^5+4550N^3-691N}{2730} \\
&= \frac{N(N+1)(2N+1)(105N^{10}+525N^9-1050N^7-1190N^6+2310N^5+1420N^4-3285N^3-287N^2+2073N-691)}{2730} \\
f_{13} &= \frac{30N^{14}+210N^{13}+455N^{12}-1001N^{10}+2145N^8-3003N^6+2275N^4-691N^2}{420} \\
&= \frac{N^2(N+1)^2(30N^{10}+150N^9-400N^7-326N^6+1052N^5+367N^4-1786N^3-202N^2+1382N-691)}{420} \\
f_{14} &= \frac{6N^{15}+45N^{14}+105N^{13}-273N^{11}+715N^9-1287N^7+1365N^5-691N^3+105N}{90} \\
&= \frac{N(N+1)(2N+1)(3N^{12}+18N^{11}+24N^{10}-45N^9-81N^8+144N^7+182N^6-345N^5-217N^4+498N^3+44N^2-315N+105)}{90} \\
f_{15} &= \frac{3N^{16}+24N^{15}+60N^{14}-182N^{12}+572N^{10}-1287N^8+1820N^6-1382N^4+420N^2}{48} \\
&= \frac{N^2(N+1)^2(3N^{12}+18N^{11}+21N^{10}-60N^9-83N^8+226N^7+203N^6-632N^5-226N^4+1084N^3-122N^2-840N+420)}{48}
\end{aligned}$$

The above list was generated using maxima with a small C code to generate the maxima script. The algorithm is quite simple as can be seen from our calculation of $f_3(N)$ above. To compute f_k we need the coefficient of each N^p in $f_{k-1}(N)$, note that $f_m(N)$ is a

polynomial of order N^{m+1} . That is

$$f_{k-1}(N) = \sum_{p=0}^k C_p^{(k-1)} N^p$$

The superscript in $C_p^{(k-1)}$ indicates which f_{k-1} this coefficient belongs to. We then have

$$\begin{aligned} f_k &= (N+1)f_{k-1} - \sum_{m=1}^N f_{k-1}(m) \\ &= (N+1)f_{k-1} - \sum_{m=1}^N \sum_{p=0}^k C_p^{(k-1)} m^p \\ &= (N+1)f_{k-1} - \sum_{p=0}^k \left(C_p^{(k-1)} \times \sum_{m=1}^N m^p \right) \\ &= (N+1)f_{k-1} - \sum_{p=0}^k (C_p^{(k-1)} f_p) \\ f_k &= (N+1)f_{k-1} - C_k^{(k-1)} f_k - \sum_{p=0}^{k-1} C_p^{(k-1)} f_p \\ f_k + C_k^{(k-1)} f_k &= (N+1)f_{k-1} - \sum_{p=0}^{k-1} C_p^{(k-1)} f_p \\ \rightarrow f_k &= \frac{1}{(1 + C_k^{(k-1)})} \times \left((N+1)f_{k-1} - \sum_{p=0}^{k-1} C_p^{(k-1)} f_p \right) \end{aligned}$$

The moral of the story here is that although Gauss's method (adding the first and last number and second and second last and so on) works for $\sum i$ it doesn't work for any other power of i^n . A blessing and a curse, I don't know what's worse :)

Now on to something fun :)

Here's the proof that 1 is equal to any number n

$$\begin{aligned}
 a &= b \\
 ab^{n-1} &= b^n \\
 a^{n-1}b &= b^n \\
 a^{n-1}b - a^n &= b^n - a^n \\
 a^{n-1}(b - a) &= (b - a) \sum_{i=1}^n b^{n-i} a^{i-1} \\
 a^{n-1} &= \sum_{i=1}^n a^{n-i} a^{i-1} \\
 \cancel{a^{n-1}} &= n \cancel{a^{n-1}} \\
 1 &= n
 \end{aligned}$$

In case you forget how to factor $b^n - a^n = (b - a) \sum_{i=1}^n b^{n-i} a^{i-1}$, you start by pulling out the factors one at a time

$$\begin{aligned}
 b^n - a^n &= (b - a)(b^{n-1} + \dots \\
 &\rightarrow -a)(b^{n-1} = -b^{n-1}a
 \end{aligned}$$

so you need to add $(b^{n-1} + b^{n-2}a \dots \rightarrow (bb^{n-2}a)$ to cancel $-a)(b^{n-1} = -b^{n-1}a$ introduced by the first factor $(b^{n-1} + \dots$, but then again you generate $-a)(b^{n-2}a = -b^{n-2}a^2$ so you repeat the process until you get

$$b^n - a^n = (b - a)(b^{n-1} + b^{n-2}a + b^{n-3}a^2 + \dots + ba^{n-2} + a^{n-1})$$

A stroll with Monsieur Fermat

The most annoying thing about Fermat's last theorem is that it looks so simple. Proving it however, was almost impossible even for the greatest mathematical minds that ever walked this planet. So what am I doing here? Let's see.

One might be tempted to think that going from a pythagorean equation with power of two to a cubic one is trivial, after all $3 = 2 + O(1) :$) As an illustration, remember the quadratic formula? We can all derive it in five minutes or less, what it does is nothing



FIG. 6: The higher the power the sharper the bump. For all three plots, vertical axis is y and horizontal axis is x .

but completing the square. Not so with the cubic formula, we can't just complete the cube. The solution to a cubic equation is in general a cube root of a square root in the form of $(a + (b^{1/2}))^{1/3}$. Yup the roots are nested. It's not as intuitive as I thought.

Prolegomenon: some greek sounding word for “Introduction”

So of course I'm not going to even try proving Fermat, what I'm doing here is to see if there's anything I can learn from it.

We start this discussion with something more obvious and lots of doodles :) The great divide for $n \leq 2$ and $n \geq 3$ can be seen immediately from the plot $x^n + y^n = 1$. Let's start with $x^1 + y^1 = 1$ which is a straight line, Fig. 6a, bumping it into $x^3 + y^3 = 1$ adds a little “bump” to that straight line, Fig. 6b, while increasing the power to 115 makes that bump a lot sharper, Fig. 6c.

But that's not the only great division, even powers have all the same shape (which is different from the odd one), Fig. 7a, except that the corners get sharper the higher the power, Fig. 7b. The only exception to this is the Pythagorean equation $x^2 + y^2 = 1$ which is a circle and the only plot with maximum symmetry among $x^n + y^n = 1$.

One interesting thing here (thanks to Wildberger) is whether those curves actually exist, *e.g.* for $x^3 + y^3 = 1$ the only rational solution is $(1,0), (0,1)$ this means that all other points on the curve are real numbers, however, the curves in Fig. 6b, 6c, 7 are calculated by a computer program and we all know that computers can *not* process real numbers.



FIG. 7: Sharper corners for sharper powers. For both plots, vertical axis is y and horizontal axis is x .

This seems to suggest that there is something special about Pythagoras, looks like “no. 2” is actually number 1 :)

Idea #1

Thanks to Wildberger’s youtube videos the following ideas popped up in my head. The ancient greeks didn’t use numbers like we do today, what we use are actually the arabic numerals. The greeks used line segments. To add one line segment on another you just join the two line segments to get a longer one. So I was curious as to what geometric meaning Fermat’s last theorem presents us.

This prodded me to the idea that the angles between the two line segments (in the case where we are adding two numbers) actually means something. For example, in the case of $a^1 + b^1 = c^1$, we are adding two line segments a and b where the angles between them is $\pi/1$.

For the case of $a^2 + b^2 = c^2$, the angles between line segments a and b is actually $\pi/2$. Now how about the angle in $a^n + b^n = c^n$ with n any arbitrary number ≥ 3 ? is the angle between a and b given by π/n ?



(a) Angle plot for triangle $(a, 1, \sqrt{a^3 + 1^3})$ (b) Angle plot for $(a, 1, \sqrt{a^{33333} + 1^{33333}})$

FIG. 8: For both plots, vertical axis is the angle between the 2 sides of triangle a and b with a fixed b normalized to 1

Using the cosine rule, the angle between line segments a and b is

$$\cos \theta = \frac{(a^n + b^n)^{2/n} - a^2 - b^2}{-2ab}$$

and its derivatives (if they mean anything) are given by

$$\begin{aligned} \frac{d \cos \theta}{dn} &= (b^n + a^n)^{\frac{2}{n}} \cdot \left(\frac{2 \cdot (b^n \cdot \log(b) + a^n \cdot \log(a))}{(b^n + a^n) \cdot n} - \frac{2 \cdot \log(b^n + a^n)}{n^2} \right) \\ \frac{d \cos \theta}{da} &= \frac{(b^n + a^n)^{\frac{2}{n}} - b^2 - a^2}{2 \cdot a^2 \cdot b} - \frac{2 \cdot a^{n-1} \cdot (b^n + a^n)^{\frac{2}{n}-1} - 2 \cdot a}{2 \cdot a \cdot b} \\ \frac{d \cos \theta}{db} &= \frac{(b^n + a^n)^{\frac{2}{n}} - b^2 - a^2}{2 \cdot a \cdot b^2} - \frac{2 \cdot b^{n-1} \cdot (b^n + a^n)^{\frac{2}{n}-1} - 2 \cdot b}{2 \cdot a \cdot b} \end{aligned}$$

The above exercise yields a provocative result. To simplify the problem, we can ask if $r_1^n + 1 = r_2^n$ has rational solutions $r_1 \equiv a/b$ and $r_2 \equiv c/b$. This normalization seems a bit odd at first. But it is actually easier to visualize things this way if one thinks of $r_1, 1$, and r_2 as sides of a triangle with r_2 the longest side among the three. This way one of the sides of the triangle is fixed to 1 while we vary the other two sides.

The angle θ between a and b is *not* fixed for $n \geq 3$. The angle always starts off at 90° when $r_1 = 1$ and will quickly plummet as r_1 increases, Fig. 8a. It will then climb back up *asymptotically* to 90° . As we go higher in n , the initial dip and rebound will become

more drastic but the overall behavior will remain the same, Fig. 8b. So there is a clear dichotomy of the behavior of the angle for $n = 1, 2$ and $n \geq 3$.

As I was thinking about what this might mean I was reminded of Galois theory (you might want to check Wildberger's youtube lecture on this as an introduction, see his history of mathematics series). The reason why quintic equations have no solution is because the symmetry structure does not cascade as it does for quartic and below. Can it be that a symmetry structure plays a role here as well? If so, what kind of symmetry?

Since the triangle formed by a , b , and c is a right triangle only for $n = 2$ I imagine it has something to do reflection symmetry with respect to the x and y axes in the cartesian coordinate. Say the triangle is given by these three points $(0, 0)$, $(a, 0)$, and $(0, b)$, by doing horizontal and vertical reflections (plus translations) we can cover the entire cartesian plane, *e.g.* a reflection w.r.t the x axis generates the triangle $(0, 0)(a, 0)(0, -b)$ while a subsequent reflection w.r.t the y axis gives $(0, 0)(-a, 0)(0, -b)$. If we then translate this triangle using a vector $(a\hat{x}, b\hat{y})$ - and combining it with our original triangle - we will get a rectangle $(0, 0)(a, 0)(a, b)(0, b)$, repeating this process indefinitely produces a covering for the entire cartesian plane.

Not so for $n \geq 3$, since the angle can never reach 90° there is always some area that is not covered by the above symmetry operations. From this observation, the symmetry of interest is the discrete symmetry Z_2 (plus some translation symmetry).

Idea #2

At this point I'm not sure what this all means, or if it is actually profitable to pursue this line of thought since I can't find anything concrete to say about it.

My next idea is whether there is a higher dimensional object that represents $a^n + b^n = c^n$ for different n 's. Since $n = 2$ is a triangle maybe $n \geq 3$ is some 3d object (after all $a^3 + b^3$ means that we are adding two volumes to get a third one) and $n = 4$ is an even higher dimensional object and so on. But again, I don't know what to make of it.

Idea #3

I then was reminded of quantum mechanics. Before learning quantum mechanics I wouldn't even imagine of something that have discrete and continuum spectrum in one

solution but there is (there are actually plenty examples in quantum mechanics). How is it related to Fermat? Because in Fermat, $a^n + b^n = c^n$ happens to just miss all of the integers a, b, c for $n \geq 3$ (while taking in real values just fine). So these seems like the inverse hydrogen atom, instead of having discrete spectrum it has an inverse discrete spectrum where all integers are skipped (is there such a physical object in the universe?).

So maybe there's a differential equations that yield a solution where all integers are skipped? can we just do a Fourier transform of $a^n + b^n = c^n$? Just a wild guess :)

Idea #4

Another strange idea I came up with was again inspired by symmetry. Say we move to the complex plane, does $a^n + b^n = c^n$ exhibit conformal symmetry? dilatation symmetry?

Why do they even relate to Fermat? It's all about angles! Back to idea #1, the angles between the two line segments a and b is always fixed for $n = 2$, this somewhat resembles conformal symmetry where the angle between two lines is always preserved. So maybe there's a conformal/dilatation symmetry preserved by $n = 2$ which is not preserved by $n \geq 3$. Again, just a wild guess, nothing concrete :)

Fun Excursions

Since the above ideas didn't get me anywhere I was thinking about other aspects of Fermat. We know we can't find solutions with just two numbers a and b , how many numbers do we need to get a solution? *i.e.* $a^n + b^n + d^n + \dots = c^n$.

So I began with $n = 3$, and lo and behold, I found integer solutions for $a^3 + b^3 + d^3 = c^3$. Do these solutions have a pattern behind them? Euclid to the rescue ... (maybe)

Running along Euclid

Euclid (some people said it was the Babylonians) gave us an algorithm to find the solutions of $a^2 + b^2 = c^2$, which is as follows

$$(m^2 - n^2)^2 + (2mn)^2 = (m^2 + n^2)^2$$

you simply pluck in whatever integers you fancy into m and n and Voila! you get a Pythagorean solution (one is usually interested in co-prime solutions, *i.e.* a, b, c have no common divisor). So maybe a similar algorithm exists for $a^3 + b^3 + d^3 = c^3$?

Well, there are actually plenty of algorithms for $a^3 + b^3 + d^3 = c^3$. They are categorized by how many parameters are used and more importantly by what solutions they produce, it turns out that the different algorithms generate only selective solutions or families unlike the Babylonian algorithm above which generate every solution for the Pythagorean equation.

The algorithm above is a two parameter algorithm, there is actually a one parameter algorithm as well. But before we discuss all this let me show you my misadventures with these algorithms as I try to wildly guess what they might be like.

Euclidean roulette

What we really want to know is how those Babylonians or Euclid came up with the formula (or they might have just been really smart and saw the solution immediately, me no Euclid). At a glance it looked like just a good guess. So let say we start with a simplest of most guesses

$$a^2 + (a + y)^2 = (a + z)^2$$

we see that the solution for a has to be

$$\begin{aligned}(a + z)^2 - \{(a + y)^2 + a^2\} &= z^2 + 2az - y^2 - 2ay - a^2 = 0 \\ \rightarrow a &= \pm \sqrt{2} \sqrt{z^2 - yz} + z - y\end{aligned}$$

which in turn means that

$$z^2 - yz = 2w^2$$

for some integer w . One obvious solution is

$$z = 2w$$

$$y = w$$

$$a = 2w + 2w - w = 3w$$

$$\rightarrow (3w)^2 + (4w)^2 = (5w)^2$$

And we get the 3, 4, 5 solution and its integer multiples. There are actually many other solutions for $z^2 - yz = 2w^2$, this turns out to be a blessing in disguise, we shall see this in a little bit.

Now, what happens if we choose a different starting point?

$$\begin{aligned}
a^2 + (a \cdot y)^2 &= (a + z)^2 \\
(a + z)^2 - \{(a \cdot y)^2 + a^2\} &= z^2 + 2az - a^2y^2 = 0 \\
\rightarrow a &= \frac{z \pm \sqrt{y^2 + 1}}{y^2}
\end{aligned}$$

This time, we can only solve it if we change y into a rational number p/q but solving $p^2/q^2 + 1 = r^2/s^2$ is just solving the Pythagorean equation in the first place. However, the above equation also means

$$\begin{aligned}
(ay^2 - z)^2 &= y^2 + 1 \\
a^2y^4 - 2ay^2z + z^2 &= y^2 + 1 \\
a^2p^4 - 2ap^2q^2z + z^2 &= p^2q^2 + q^4
\end{aligned}$$

So why all this asinine substitutions?

Lagniappe

I wanted to say “small lagniappe” but lagniappe itself is already small :) The above two examples actually show us how to use Euclid’s algorithm to solve other equations (instead of generating a new algorithm to solve the pythagorean equation). The first example

$$z^2 - yz = 2w^2$$

can be paraphrased into the Pythagorean equation, because it was a consequence of

$$\begin{aligned}
a^2 + (a + y)^2 &= (a + z)^2 \\
(m^2 - n^2)^2 + (2mn)^2 &= (m^2 + n^2)^2
\end{aligned}$$

which means that

$$\begin{aligned}
a &= m^2 - n^2 \\
a + y &= 2mn \\
a + z &= m^2 + n^2
\end{aligned}$$

By solving the above equations, the algorithm for the solutions of $z^2 - yz = 2w^2$ is then

$$z = 2n^2, \quad y = n^2 + 2mn - m^2, \quad w = n^2(n - m)^2$$

i.e. we can choose any two integers m and n and the integer solutions for $z^2 - yz = 2w^2$ are given above (note that we can also choose $a = -(m^2 - n^2)$ generating a slightly different final result).

It's the same with $a^2p^4 - 2ap^2q^2z + z^2 = p^2q^2 + q^4$, equating terms just like before gives

$$a = m^2 - n^2, \quad z = 2n^2, \quad y = \frac{p}{q} \rightarrow p = 2mn, q = m^2 - n^2$$

I called the above lagniappe because it was just a small boon. The boon is that there's actually a way to solve these seemingly random equations, *e.g.* $z^2 - yz = 2w^2$, $a^2p^4 - 2ap^2q^2z + z^2 = p^2q^2 + q^4$. However, those solutions we found above are not unique because Euclid's algorithm itself is not unique, as we shall see.

There's also another fun thing related to this, if we start with

$$(a + z)^2 + (a - z)^2 = (a \cdot y)^2$$

$$(a + z)^2 + (a - z)^2 - (a \cdot y)^2 = -2z^2 + a^2y^2 - 2a^2 = 0$$

we can now choose which one to solve

$$\begin{aligned} \text{solving } y : y &= \pm \frac{\sqrt{2}\sqrt{z^2 + a^2}}{a} \\ \text{solving } z : z &= \pm \frac{a\sqrt{y^2 - 2}}{\sqrt{2}} \\ \text{solving } a : a &= \pm \frac{\sqrt{2}z}{\sqrt{y^2 - 2}} \end{aligned}$$

note that we are not solving y followed by z followed by a , we are only solving *one* of them but since we have three variables we can choose which one to solve. We then have the following constraints

$$\text{from solving } y \text{ we get : } z^2 + a^2 = 2w^2$$

$$\text{from solving } z \text{ or } a \text{ we get : } 2u^2 + 2 = y^2$$

matching terms $(a + z)^2 + (a - z)^2 = (a \cdot y)^2 \rightarrow (m^2 - n^2)^2 + (2mn)^2 = (m^2 + n^2)^2$

$$\begin{aligned} a &= m^2 - n^2 + 2mn \\ z &= m^2 - n^2 - 2mn \\ y &= \frac{2(m^2 + n^2)}{m^2 - n^2 + 2mn} \\ \rightarrow w &= m^2 + n^2 \\ \rightarrow u &= \frac{n^2 + 2mn - m^2}{n^2 - 2mn - m^2} \end{aligned}$$

Note that we have inserted factor of 2 where convenient :) It is now obvious that we can play this game forever, just substitute any function you fancy into the Pythagorean equation and match terms, just be careful about factors of 2.

The moral of the story here is that the Babylonians actually gave us quite a powerful tool to solve deophantine equations (if we can massage them into the Pythagorean one), these ancient people were not just good at building towers :) But in a broader sense it means that we can map one Diophantine equation into a simpler one, which is good.

Euclidean bulldozer

We now come back to the subject of utilizing unsophisticated guesses in meting out algorithms such as Euclid's. The first attempt above $a^2 + (a + y)^2 = (a + z)^2$ wasn't fruitful and in a glance this should be the same as $(a + x)^2 + (a + y)^2 = (a + z)^2$ since we can always shift $a + x \rightarrow a$. However, it is not true. Adding an extra degree of freedom like x is a big deal, how big? let's see.

We can by brute force find an algorithm for $a^2 + b^2 = c^2$ from the most generic starting point

$$\begin{aligned} (a + z)^2 &= (a + x)^2 + (a + y)^2 \\ (a + z)^2 - \{(a + x)^2 + (a + y)^2\} &= z^2 + 2az - y^2 - 2ay - x^2 - 2ax - a^2 = 0 \end{aligned}$$

Now we have four choices, we can solve for either, a , x , y , or z , which one we solve determines our fate. Solving for a gives

$$a = z - y - x \pm \sqrt{2} \sqrt{z^2 + (-y - x)z + xy}$$

which means that

$$\begin{aligned}
z^2 + (-y - x)z + xy &= 2p^2 \\
\rightarrow x &= \frac{z^2 - yz - 2p^2}{z - y} \\
\rightarrow a &= -\frac{(y + 2|p|)z - y^2 - 2|p|y - 2p^2}{z - y}
\end{aligned}$$

and there we have it, the full algorithm is

$$\begin{aligned}
a + x &= z - y - 2|p| && \rightarrow (z - y - 2|p|)(z - y) \\
a + y &= -\frac{2|p|z - 2|p|y - 2p^2}{z - y} && \rightarrow 2|p|z - 2|p|y - 2p^2 \\
a + z &= \frac{z^2 + (-2y - 2|p|)z + y^2 + 2|p|y + 2p^2}{z - y} && \rightarrow z^2 + (-2y - 2|p|)z + y^2 + 2|p|y + 2p^2
\end{aligned}$$

we can multiply each term by $(z - y)$ if we want to and flip the minus sign. So apparently we can just go by brute force to bulldoze out an algorithm, in this case we have *three* final degrees of freedom.

However, we can also solve x first

$$x = -a \pm \sqrt{z^2 + 2az - y^2 - 2ay}$$

The next step is crucial, do we solve for a or y, z ? Let's see, solving for a gives

$$a = -\frac{z^2 - y^2 - p^2}{2z - 2y}$$

which is all well and good. Had we chosen to solve for y we'd have gotten

$$\begin{aligned}
y &= -a \pm \sqrt{z^2 + 2az - p^2 + a^2}, \text{ solving for } a \text{ next} \\
\rightarrow a &= -z \pm \sqrt{q^2 + p^2} \quad \text{OR solving for } z \text{ next} \\
\rightarrow z &= -a \pm \sqrt{q^2 + p^2}
\end{aligned}$$

Solving either a or z produces no headway since now we have to solve for $q^2 + p^2 = r^2$ which is our original problem. Thus the order we choose to solve the degrees of freedom matters. Can we follow the same mechanical steps to arrive at algorithms for $a^3 + b^3 + c^3 = d^3$?

Cubic conundrum

Applying the same guessing strategy to $a^3 + b^3 + c^3 = d^3$ turns out to be tricky. My first attempt was $a^3 + (a + y)^3 + (a + z)^3 = (a + w)^3$ but the cube root of a square root (generated by solving a in terms of y, z, w) was too hard to handle.

The next natural progression is of course $(a+x)^3 + (a+y)^3 + (a+z)^3 = (a+w)^3$ to allow more flexibility but just like the preceding attempt the resulting equations quickly became intractable. My next attempt was to simplify things by adding a minus sign $(a+x)^3 + (-a+y)^3 + (a+z)^3 = (a+w)^3$ such that we no longer have a^3 , however even with this modification the equations were still too long to be discernible

$$\begin{aligned}
(a+w)^3 - \{(a+x)^3 + (-a+y)^3 + (a+z)^3\} &= 0 \\
\implies a &= (6z+6y+6x-6w)^{-1}(-3z^2+3y^2-3x^2+3w^2 \\
&\quad \pm \sqrt{3}(-z^4+(-4y-4x+4w)z^3+(-6y^2+6x^2-6w^2)z^2+ \\
&\quad (-4y^3-4x^3+4w^3)z-y^4+(4w-4x)y^3+(6w^2-6x^2)y^2+ \\
&\quad (4w^3-4x^3)y-x^4+4wx^3-6w^2x^2+4w^3x-w^4)^{1/2})
\end{aligned}$$

here we have a super long square root (no more cube root since we have no a^3). The problem with this square root is that it contains fourth powers z^4, y^4, x^4 . Solving for x, y or z inside the square root generates more problems, *i.e.* roots containing higher powers which in turn generates more complicated equations.

My final attempt was $(a+x)^3 + (-b+y)^3 + (b+z)^3 = (a+w)^3$ however, solving a or b still generates ridiculously long square roots, solving for x however seems to be a lot simpler

$$x = (-z^3 - 3az^2 - 3a^2z - y^3 - 3by^2 - 3b^2y + w^3 + 3aw^2 + 3a^2w - b^3)^{1/3} + b$$

We are now faced with the same decision tree as discussed previously, who's next? Turns out that solving for y, z or w generates $p^3 + q^3 + r^3 = s^3$ just like the Pythagorean case above. Seems like we need to solve b or a first but b gives a better outcome

$$b = (-z^3 - 3az^2 - 3a^2z + w^3 + 3aw^2 + 3a^2w - p^3)^{1/3} + y$$

It looks good as y decouples. Next we need to solve for a , solving any other variables generates $p^3 + q^3 + r^3 = s^3$

$$\begin{aligned}
a &= \left(3z^2 - 3w^2 \pm \sqrt{3} \sqrt{-z^4 + 4wz^3 - 6w^2z^2 + (4w^3 - 4q^3 - 4p^3)z - w^4 + (4q^3 + 4p^3)w} \right) \times \\
&\quad (6z - 6w)^{-1}
\end{aligned}$$

That long square root is a killer, solving it for z or w generates a root of a cube root of a root of a root, *four* nested roots! *Q.E.D* (*Q*.uickly *E*nded *D*.reams)

In short, a brute force guessing game might be theoretically possible but not recommended :) in other words, it works well for certain obvious cases and thus it only has a very limited merit.

Euclidean apery

We now go back to guessing our own version of an algorithm for the Pythagorean equation (without a bulldozer). In our case, the goal here is to solve a in terms of two completely free parameters z and y . Let's try one with a quite generic structure

$$\begin{aligned}(a - z)^2 + (a \cdot y)^2 &= (a + z)^2 \\ (a + z)^2 - \{(a \cdot y)^2 + (a - z)^2\} &= 4az - a^2y^2 = 0 \\ &\rightarrow a = \frac{4z}{y^2}\end{aligned}$$

Yay! It works! (the other solution is $a = 0$). The full solution is then

$$\begin{aligned}\left(\frac{4z}{y^2} - z\right)^2 + \left(\frac{4zy}{y^2}\right)^2 &= \left(\frac{4z}{y^2} + z\right)^2 \\ \Rightarrow (4z - zy^2)^2 + (4zy)^2 &= (4z + zy^2)^2\end{aligned}$$

Although this is different from Euclid's, given any integer z and y we can generate a solution. Looks like these algorithms are not unique (as advertised) although I'm not sure (or too lazy to prove) that this version covers all solutions of the Pythagorean equation.

A more straightforward guess will be

$$(a + z)^2 - (a - z)^2 = 4az$$

we can simply promote $a \rightarrow a^2, z \rightarrow z^2$ to get

$$(a^2 + z^2)^2 - (a^2 - z^2)^2 = 4a^2z^2 = (2az)^2$$

which is Euclid's (or the Babylonians').

So the trick in this guessing game is that you guess a term on the LHS and another on the RHS. You then want the difference between those two sides *not* to be a sum of terms, because then you're just going back to square one by needing to solve some equation in

the form of $x_1^{n_1} + x_2^{n_2} = a^m$. For example, if you're naively try to solve fermat's cubic last theorem the same way

$$\begin{aligned}(a - z)^3 + X^3 &= (a + z)^3 \\ \rightarrow X^3 &= 2z^3 + 6a^2z\end{aligned}$$

You just get another deophantine equation to solve, back to square one.

One thing to note here is that we at first assume that all terms in our guess has the same power, *e.g.* a, z both have the power of one, we can for example promote things to a^n, z^m but our goal for now is to see how far these simple guesses take us.

Jabs in the dark

We now begin our journey in guessing the form of algorithm for $a^3 + b^3 + c^3 = d^3$.

1. I first tried the following solution (which looks simplest of all)

$$a^3 + b^3 + K^3 = (a + b)^3$$

with some unknown integer K (in Euclid's case K is just $2mn$). Let's see if we get lucky. Now K^3 is obviously $3ab(b + a)$. Sadly, the only solution is $a = 1, b = 8, K = 6$ and their integer multiples.

We can try setting $a = 3^2c^3$. First, why this choice? The goal here is to make $3a$ into $3a = P^3$ for some integer P . So the first thing to do is to complete the factor 3 into 3^3 by multiplying it by 3^2 . What we need next is $b(b + a) = N^3$ which we will solve later. With this choice $3ab(b + a) = 3^3c^3N^3 = K^3 \rightarrow K = 3cN$ we would have found our algorithm. However, even with this scheme, the only solutions are integer multiples of $a = 1, b = 8, K = 6$.

2. From the list of solutions for $a^3 + b^3 + c^3 = d^3$ I saw a familiar face 3, 4, 5, 6 (by the way, the tuplet 3, 4, 5, 6 also represent the sides of a right triangle 3, 4, 5 with an area of 6). Prompting me to guess a solution of the form

$$\begin{aligned}a^3 + (a + 1)^3 + (a + 2)^3 &= (a + 3)^3 \\ (a + 3)^3 - (a^3 + (a + 1)^3 + (a + 2)^3) &= 2(3 - a)(a^2 + 3a + 3) = 0\end{aligned}$$

The last equation shows that this scheme will only work for $a = 3$, no other solution is sequential.

3. What if we mix products and additions of terms? Like this one

$$a^3 + (a \cdot x)^3 + (a \cdot y)^3 = (a + z)^3$$

$$(a + z)^3 - \{a^3 + (a \cdot x)^3 + (a \cdot y)^3\} = z^3 + 3az^2 + 3a^2z - a^3y^3 - a^3x^3$$

Now this doesn't look promising to some since we can always factor out a to get $y^3 + x^3 + 1 = w^3$, however, from our previous discussion about shifting $a + z \rightarrow a$, we know that adding a degree of freedom actually makes a difference. So let give this a shot. The solution is

$$a = \frac{z \{(y^3 + x^3 + 1)^{2/3} + (y^3 + x^3 + 1)^{1/3} + 1\}}{(y^3 + x^3)}$$

while the other two solutions give imaginary a . Which is rather disappointing since we now need to solve

$$y^3 + x^3 + 1 = w^3$$

so this time adding a degree of freedom doesn't help. If we insist in substituting the solution for a the factor z cancels from both LHS and RHS and the only free variables in $a^3 + (a \cdot x)^3 + (a \cdot y)^3 = (a + z)^3$ are just x and y which is the same free variables we have in $y^3 + x^3 + 1 = w^3$.

At this point I would like to list solutions of $y^3 + x^3 + 1 = w^3$, see Table I (sorted in ascending x). One can immediately see that $y = 138$ gives us *two* different sets of solutions (426 also gives two solutions but once as x and another as y). Also, from their prime factors we can see that the solutions are all co-prime to one another.

Another thing to note is that the difference $y - x$ and $w - y$ are all different for different solutions, showing that an algorithm, *i.e.* $a^3 + (a + m)^3 + (a + n)^3 = (a + p)^3$ where m, n, p are *fixed* will not work as each fixed difference m, n, p will only give a single solution.

In addition, this also means that it is a near miss for Fermat, $a^3 + b^3 = c^3$ has no solutions but $1 + a^3 + b^3 = c^3$ does :)

One last comment on this topic, we can arrange things differently $y^3 + x^3 + 1 = w^3 \rightarrow y^3 + x^3 = w^3 - 1$

$$\begin{aligned} y^3 + x^3 &= w^3 - 1 \\ &= (w - 1)(w^2 + w + 1) \\ &= (w - 1)((w - 1)^2 + 3w) \\ &= (w - 1)^3 + 3w(w - 1) \end{aligned}$$

We can now add degrees of freedom as usual $y = b + z$, $x = b - z$

$$\begin{aligned} (b + z)^3 + (b - z)^3 &= (w - 1)^3 + 3w(w - 1) \\ 6bz^2 + 2b^3 &= (w - 1)^3 + 3w(w - 1) \end{aligned}$$

This looks promising as we have the same number of terms in both LHS and RHS. We know that $2b^3$ can't be u^3 for some integer u due to the factor of 2, the easiest algorithm for $y^3 + x^3 + 1 = w^3$ is to set $2b^3 = 3w(w - 1)$, $6bz^2 = (w - 1)^3$. However, from the 1, 6, 8, 9 solution it is clear that we cannot make such identification.

We can try adding more variables, *e.g.* $1 = (b - z - a)$, $y = (b + z)$, but this only complicates things with no clear benefit. In other words, only the combination, *e.g.* $6bz^2 + 2b^3$, constitutes $x_1^3 + x_2^3$, we cannot cast any one term $6bz^2$ or $2b^3$ as a pure power of 3, $6bz^2 = p^3$ or $2b^3 = q^3$.

4. Another potential solution is

$$(a - b)^3 + (b - c)^3 + (c - a)^3 = 3(a - b)(b - c)(c - a)$$

The nice thing is that the RHS is not a sum of terms, so we have a chance if we set

$$(a - b)(b - c)(c - a) = 3^2 w^3$$

To simplify things let's denote $p = (a - b)$, $q = (b - c)$, $r = (c - a)$

$$p \cdot q \cdot r = 3^2 w^3$$

$$p + q + r = 0$$

We can solve the above equations by substituting $p = 3^2 w^3 / qr$ into $p + q + r = 0$

$$\begin{aligned} 0 &= 3^2 w^3 + q^2 r + qr^2 \\ \rightarrow r &= \frac{-q^2 \pm \sqrt{q^4 - 36 \cdot q \cdot w^3}}{2q} \end{aligned}$$

To ensure that r is an integer, $q^4 - 36qw^3 = v^2$ for some integer v . The interesting thing is that there are only two solutions I can find for this to happen

$$q \rightarrow q, w \rightarrow -2q : q^4 - 36q(-2q)^3 = 289q^4 = (17q^2)^2$$

$$q \rightarrow 4q, w \rightarrow -q : (4q)^4 - 36(4q)(-q)^3 = 400q^4 = (20q^2)^2$$

Regrettably this provision will only give one solution

$$(8q)^3 + (q)^3 + (-9q)^3 = (-6q)^3$$

which is just a multiple of the 1, 6, 8, 9 solution.

We can do some minor rearrangement

$$\begin{aligned} p^3 + q^3 + 3pq(p + q) &= (p + q)^3 \\ \rightarrow (3p^3)^3 + (3q^3)^3 + (3pqz)^3 &= (z^3)^3 \quad z^3 = 3p^3 + 3q^3 \quad \text{OR} \\ \rightarrow (3^2 p^3)^3 + (q^3)^3 + (3pqz)^3 &= (z^3)^3 \quad z^3 = 3^2 p^3 + q^3 \quad \text{OR} \\ \rightarrow (3p^3)^3 + (q^3)^3 + (3pqz)^3 &= (z^3)^3 \quad 3z^3 = 3p^3 + q^3 \quad \text{OR} \\ \rightarrow p^3 + q^3 + z^3 &= (p + q)^3 \quad z^3 = 3pq(p + q) \end{aligned}$$

we can either lump the factor of 3^2 into one of the variables or spread it onto any two of them or assume the whole thing $3pq(p + q)$ is a power of three z^3 . I got

some interesting result from this exercise, *e.g.* the *largest* solution I found from this simple guessing game is

$$59707533^3 + 2268747144^3 + 981751158^3 = 2328454677^3$$

this one was from $p \rightarrow 3p^3$, $(p + q) \rightarrow 3u^3$.

5. Another promising solution is

$$\begin{aligned}(m - n)^3 + n^3 + A^3 &= (m + n)^3 \\ \rightarrow A^3 &= n^3 + 6m^2n\end{aligned}$$

We can solve for m or n or A , but m seems to yield the easiest solution, which is

$$m = \pm \sqrt{\frac{\frac{A^3}{n} - n^2}{6}}$$

Sadly, this only gives 2 integer solutions, they are

$$\begin{array}{llll} m - n = 17 & n = 4 & A = 22 & m + n = 25 \\ m - n = -2 & n = 9 & A = 15 & m + n = 16 \end{array}$$

I was hoping that these conditions, *e.g.* $m = \pm \sqrt{\frac{\frac{A^3}{n} - n^2}{6}}$ or $r = \frac{-q^2 \pm \sqrt{q^4 - 36 \cdot q \cdot w^3}}{2q}$ would yield a class of integer solutions instead of just one (or two).

So it looks like there's really no solution for $a^3 + b^3 + c^3 = d^3$ if we start we just simple guesses.

Expanding the horizon

Finding a generic algorithm for $a^3 + b^3 + c^3 = d^3$ just by guessing seems to be quite problematic as we have seen above. So what if we add another number? Say

$$a^3 + b^3 + c^3 + d^3 = e^3$$

TABLE I: (x) means prime factors of x

x	y	w	(x)	(y)	(w)
6	8	9	$(2, 3)$	(2^3)	(3^2)
71	138	144	(71)	$(2, 3, 23)$	$(2^4, 3^2)$
135	138	172	$(3^3, 5)$	$(2, 3, 23)$	$(2^2, 43)$
236	1207	1210	$(2^2, 59)$	$(17, 71)$	$(2, 5, 11^2)$
242	720	729	$(2, 11^2)$	$(2^4, 3^2, 5)$	(3^6)
372	426	505	$(2^2, 3, 31)$	$(2, 3, 71)$	$(5, 101)$
426	486	577	$(2, 3, 71)$	$(2, 3^5)$	(577)
566	823	904	$(2, 283)$	(823)	$(2^3, 113)$
575	2292	2304	$(5^2, 23)$	$(2^2, 3, 191)$	$(2^8, 3^2)$
791	812	1010	$(7, 113)$	$(2^2, 7, 29)$	$(2, 5, 101)$
1124	5610	5625	$(2^2, 281)$	$(2, 3, 5, 11, 17)$	$(3^2, 5^4)$
1938	2820	3097	$(2, 3, 17, 19)$	$(2^2, 3, 5, 47)$	$(19, 163)$
1943	6702	6756	$(29, 67)$	$(2, 3, 1117)$	$(2^2, 3, 563)$
2196	5984	6081	$(2^2, 3^2, 61)$	$(2^5, 11, 17)$	$(3, 2027)$
2676	3230	3753	$(2^2, 3, 223)$	$(2, 5, 17, 19)$	$(3^3, 139)$

Going back to our symmetrical solution

$$\begin{aligned}
(a - z)^3 + \Delta &= (a + z)^3 \\
\rightarrow (a + z)^3 - (a - z)^3 &= \Delta = 2z^3 + 6a^2z \\
\Delta &= z^3 + z^3 + 6a^2z
\end{aligned}$$

so as long as $a^2 = 6^2w^6$ and $z = v^3$ the above solution works since every term is a power of 3. The full solution is then

$$(a^3 - 6^2b^6)^3 + (6^2b^6)^3 + (6^2b^6)^3 + (a^2 \cdot 6b^2)^3 = (a^3 + 6^2b^6)^3$$

But I don't like this solution since it contains a constant factor 2. There's yet another

solution, if we began with

$$\begin{aligned} (-a + b + c)^3 + (a - b + c)^3 + (a + b - c)^3 + X &= (a + b + c)^3 \\ &\rightarrow X = 24abc \end{aligned}$$

The above approach works as there is a so called “symmetry” where all complicated terms cancel for LHS and RHS. This was done by alternating the minus sign in each term to make sure there is only one a^3, b^3, c^3 on the LHS as there is on the RHS. Now, since X is just a product of terms the most obvious choice would be $a \rightarrow a^3, b \rightarrow b^3, c \rightarrow 3^2 c^3$. The full solution is then

$$(-a^3 + b^3 + 3^2 c^3)^3 + (a^3 - b^3 + 3^2 c^3)^3 + (a^3 + b^3 - 3^2 c^3)^3 + (2 \cdot 3 \cdot abc)^3 = (a^3 + b^3 + 3^2 c^3)^3$$

There are other choices as well like $a \rightarrow a^3, b \rightarrow 3b^3, c \rightarrow 3c^3$. A more generic solution would be $a \rightarrow a^{3 \cdot m_1}, b \rightarrow 3 \cdot K_1^{3 \cdot n_1} b^{3 \cdot m_2}, c \rightarrow 3 \cdot K_2^{3 \cdot n_2} c^{3 \cdot m_2}$ but this also allows for something else, namely $a^3 + b^3 + c^3 + d^N = e^3$ by setting $a \rightarrow 2^{N-3} 3^{N-1} a^N, 2^3 \cdot 3 \cdot abc \rightarrow 2^3 \cdot 3 (2^{N-3} 3^{N-1} a^N) b^N c^N = (2 \cdot 3 \cdot abc)^N, i.e.$

$$\begin{aligned} (2^{N-3} 3^{N-1} a^N + b^N + c^N)^3 &= (-2^{N-3} 3^{N-1} a^N + b^N + c^N)^3 + (2^{N-3} 3^{N-1} a^N - b^N + c^N)^3 + \\ &\quad (2^{N-3} 3^{N-1} a^N + b^N - c^N)^3 + (2 \cdot 3abc)^N \end{aligned}$$

The most general solution would be to distribute the prime factors of $P^3 (\equiv 24abc)$ into a, b , and c ,

$$P = \prod_{i=1}^L p_i^{n_i} \quad \longrightarrow \quad P^3 = \prod_{i=1}^L p_i^{3n_i}$$

where each p_i is a prime number, n_i the multiplicity of p_i and L the number of different prime factors of P and the probability is infinite :)

One way we can use our newly found expanded horizon is to set one of the terms, $(-a + b + c)$ or $(a - b + c)$ or $(a + b - c)$, to zero, since we can't set the product term $24abc$ to zero without resulting in a trivial pursuit :)

By setting one of the terms above to zero we would get our prime objective $a^3 + b^3 + c^3 = e^3$ from $a^3 + b^3 + c^3 + \cancel{d^N} = e^3$. However, this exercise proves futile as the only solution generated is the infamous 1, 6, 8, 9, *i.e.* $a = 1, b = 2, c = 1 \rightarrow -9a^3 + b^3 + c^3 = 0$.

There is a possible (and potential) use of the above algorithm. Say we start with $(a + z)^3$ and $(a + y)^3$ on the LHS and RHS respectively, what we want is

$$\begin{aligned}
(a + z)^3 &= (a + y)^3 + C_1^3 + C_2^3 \\
(a + z)^3 - (a + y)^3 &= 3a^2z + 3az^2 + z^3 - (3a^2y + 3ay^2 + y^3) \\
&= 3a^2z + 3az^2 - y^3 - (3a^2y + 3ay^2 - z^3) \\
&= (a + z)^3 - a^3 - z^3 - y^3 - ((a + y)^3 - a^3 - y^3 - z^3) \\
&= C_1^3 - C_2^3 \\
\rightarrow C_1^3 &= (a + z)^3 - a^3 - z^3 - y^3 \\
\rightarrow C_2^3 &= (a + y)^3 - a^3 - y^3 - z^3
\end{aligned}$$

in the second line we need to swap $y \leftrightarrow z$ otherwise we'll get

$$\begin{aligned}
3a^2z + 3az^2 + z^3 &= (a + z)^3 - a^3 \\
\tilde{C}^3 &= (a + z)^3 - a^3
\end{aligned}$$

which is Fermat and we don't have a solution for it. However, we now have $a^3 + b^3 + c^3 + d^3 = e^3$ for both C_1 and C_2 , the bad news is that we might not be able to use the algorithm found above to solve them. The main problem is that C_2 has the same terms as C_1 on the RHS. Another problem is, even if we only consider C_1 , matching terms to our algorithm above

$$\begin{aligned}
(a + z) &= -p^3 + q^3 + 3^2r^3 \\
a &= -p^3 + q^3 - 3^2r^3 \\
z &= -p^3 - q^3 + 3^2r^3 \\
y &= -2 \cdot 3 \cdot pqr
\end{aligned}$$

the solution will be

$$\begin{aligned}
r &= \left(\frac{q^3 + p^3}{9} \right)^{1/3} \\
a &= 2q^3 \\
y &= 2 \cdot pq \cdot 3^{1/3} (q^3 + p^3)^{1/3} \\
z &= -2(q^3 + p^3)
\end{aligned}$$

TABLE II: Some sample of twins for $x^3 + y^3 + z^3 = w^3$, including one of the biggest one I found (see bottom row)

x	y	z	w
3	36	37	46
27	30	37	46
76	165	192	229
102	157	192	229
9	58	255	229
9	183	220	229
58	201	255	292
183	201	220	292
24	159	382	391
87	150	382	391
2678	10533	43420	43629
5157	10166	43420	43629

while p and q remain free parameters. The above requires that

$$q^3 + p^3 = \frac{w^3}{3}$$

and the only solution is $q = 1$, $p = 2$, $w = 3$ (this is just the 1,6,8,9 solution with $z = 0$). We could have chosen our other algorithm, *i.e.* $a + z = -p^3 + 3q^3 + 3r^3$ but this choice leaves us no way out. But why did we choose $(a + z) = -p^3 + q^3 + 3^2r^3$ in the first place? because other choices generate complex solutions :)

Observation intermission

We now take a break from guessing and start staring at the numerical solutions of $a^3 + b^3 + c^3 = d^3$ we found and take stock of what we've found so far.

First, the solutions are not unique, we have pairs of numbers that appear in more than one solution, see Table. II. There are in total 1290 such twins from a total of 39607 solutions (around 3%, solutions are generated ascendingly by x , the smallest number in each



FIG. 9: Plot of the distribution of twins among solutions of $a^3 + b^3 + c^3 = d^3$. Each green dot indicates where each twin was born among the solutions. First data point is excluded as first twins occur as solutions #3 giving a distortedly high 33.3% value.

solution). However, this percentage depends on how we search the solution space. Twins are more likely to pop up when x, y, z are relatively close, see comments in “Intuition rumination” section.

Moreover, there are intriguing solutions like this one, where the difference between the two largest terms is just one!

$$462^3 + 1981^3 + 51227^3 = 51228^3$$

An interesting question will be: the distribution of these twins, are they spread evenly? does their percentage keep fluctuating or does it stabilize at some point or does it increase without bound? For example see Fig. 9

Second, guessing for an algorithm is not the best strategy here because there are four terms in total. It is hard to cancel terms between LHS and RHS that will result only in a product of terms as we can see from my attempts above, *e.g.*

$$(a + z)^3 - (a - z)^3 = 2z^3 + 6a^2z$$

the factor of 2 in front of z^3 spoils everything because we need $2z^3 = w^3$ for some integer w . What if we add more variables? say $(x + y + z)^3$, the problem is that there are three

cubes on the RHS x^3, y^3, z^3 but we only have 3 terms on the LHS, this means that

$$(x + \Delta_1)^3 + (y + \Delta_2)^3 + (z + \Delta_3)^3 = (x + y + z)^3$$

We can cancel x^3, y^3, z^3 on both sides but we now have $\Delta_1^3, \Delta_2^3, \Delta_3^3$ on the LHS (plus cross terms on both sides) that we cannot cancel.

Contrast this situation to $a^3 + b^3 + c^3 + d^3 = e^3$. If we start with $(x + y + z)^3$ on the RHS, we can cycle a minus sign on the LHS $(-x + y + z)^3, (x - y + z)^3, (x + y - z)^3$ and we will have exactly one x^3, y^3, z^3 on both sides with an extra term d^3 to absorb the rest of the cross terms. No such luck with just four terms $a^3 + b^3 + c^3 = d^3$, even if we start with two terms $(x + y)^3$ on the RHS the best I could get was

$$(x + a)^3 + (y - a)^3 + 3(x + a)(y - a)(x + y) = (x + y)^3$$

which was attempt #4 with some minor rearrangement.

Speaking of attempt #4 and #5, we see something peculiar. Our usual observation says that we have one less constraint than the number of degrees of freedom and so we should have a multitude of solutions instead of just one. Let's start with #4, its constraints are

$$p^3 + q^3 + r^3 = 3^3 w^3$$

$$p \cdot q \cdot r = 3^2 w^3$$

$$p + q + r = 0$$

so we have 4 degrees of freedom p, q, r, w with three constraints. For #5, its *two* constraints are

$$(m - n)^3 + n^3 + A^3 = (m + n)^3$$

$$n^3 + 6m^2 n = A^3$$

Again one extra free term, two constraints with *three* degrees of freedom m, n, A . The odd thing here is that the above two examples yield only one solution each, despite the extra *free* degree of freedom.

How can this be? This is because we are so used to thinking about real numbers. When solving a system of equations, the equations are the only constraints there are.

Not so with integers/natural numbers. By requiring that the solutions are integers we are implicitly imposing another constraint, albeit a non traditional one. Going back to Fermat, $a^3 + b^3 = c^3$, we have only one constraint and three degrees of freedom and yet it has no solutions! It's because all the solutions are real numbers.

Let's see this in more detail through a concrete example

Attempt #6 must die!

This one was exhilarating because at a glance it gives you a lot of hope until you see it more closely. The starting point is

$$a^3 + (y^{1/3})^3 + (a - z)^3 = (a + 8z)^3$$

$$(a + 8z)^3 - \{a^3 + (y^{1/3})^3 + (a - z)^3\} = 513z^3 + 189az^2 + 27a^2z - y - a^3 = 0$$

The funny term $(y^{1/3})^3$ is due to hindsight :) and even though in itself it is a constraint, it's irrelevant to the point I'm trying to make here. On the onset it seems like we have three degrees of freedom with only one constraint. Implying

$$a = \left(\frac{\sqrt{1539648z^6 - 7344yz^3 + y^2}}{2} - \frac{y - 3672z^3}{2} \right)^{1/3}$$

$$+ 144z^2 \left(\frac{\sqrt{1539648z^6 - 7344yz^3 + y^2}}{2} - \frac{y - 3672z^3}{2} \right)^{-1/3} + 9z$$

To have any chance at all the inner square root has to produce an integer

$$1539648z^6 - 7344yz^3 + y^2 = u^2$$

but this is an additional constraint. One might object that it also introduces a new degree of freedom u . However, solving the above equation

$$y = 3672z^3 \pm \sqrt{11943936z^6 + u^2}$$

This in turn means we acquire another constraint $\sqrt{11943936z^6 + u^2}$ (which comes with another degree of freedom). This is a zero sum game, we are not gaining anything. We can continue with this forever!

Say we don't want to play this game, we can set

$$1539648z^6 - 7344yz^3 + y^2 = 0$$

$$\rightarrow y = 216z^3$$

which is good news because $216 = 6^3$ (another solution is $y = 7128z^3$). However, by this time we have utilized two constraints, the original equation and $1539648z^6 - 7344yz^3 + y^2 = 0$. This is no good. To see why, let's continue.

We have one more constraint to solve, $((y - 3672z^3)/2)^{1/3}$ has to be an integer, the solution is $z = -w/12$ for some integer w . This looks promising, we have a degree of freedom, but if you substitute this solution, $z = -w/12$, $y = 216z^3 = -w^3/8$, $a = -11w/4$, *i.e.* all terms in $a^3 + (y^{1/3})^3 + (a - z)^3 = (a + 8z)^3$ are just multiples of w

$$\begin{aligned} \left(\frac{-11w}{4}\right)^3 + \left(\frac{-w}{2}\right)^3 + \left(\frac{-8w}{3}\right)^3 &= \left(\frac{-41w}{12}\right)^3 \\ &\rightarrow 33^3 + 6^3 + 32^3 = 41^3 \end{aligned}$$

Sadly we only get one final solution 6, 32, 33, 41 because we have three degrees of freedom to begin with but utilize three constraints to get a solution, $a^3 + (y^{1/3})^3 + (a - z)^3 = (a + 8z)^3$, $\sqrt{11943936z^6 + u^2}$, $((y - 3672z^3)/2)^{1/3}$. Thus the requirement that the solutions are integers is itself a powerful constraint!

Another thing to note about the example above is that since the solution for a cubic polynomial is a cube root of a square root, $(\dots + \sqrt{\dots})^{1/3}$, we immediately incur two constraints, one for each root.

One quick side note, I initially wanted to treat the numbers in $a^3 + b^3 + c^3 = d^3$ as vectors much like in the case of $a^2 + b^2 = c^2 \rightarrow \vec{a} + \vec{b} = \vec{c}$ with a right angle between \vec{a} and \vec{b} . What would become of $\vec{a} + \vec{b} + \vec{c} = \vec{d}$? maybe there's some sort of new cosine rule? The problem is we now deal with a generic "parallelogram" and for such a four-sided object the angle between any two sides is not fixed as we can always squeeze the object and the length of all sides will still stay the same. This is in stark contrast to a triangle where changing the angle between any two sides would change the length of the other side.

Observation intermission part deux

The way I initially generated the solutions was through tripply nested loops $z_1 = 1 \rightarrow 1000$, $z_2 = 1 \rightarrow 1000$, $z_3 = 1 \rightarrow 1000$ where at each iteration I checked whether $z_1^3 + z_2^3 + z_3^3$ is equal to some w^3 .

It was a symmetrical search among z_1, z_2 , and z_3 . It was very inefficient since it went through things multiple times, *e.g.* 1,6,8,9 was rediscovered multiple times through 6,1,8,9 and 8,1,6,9.

I then modified it to remove any redundancy and also made it asymmetrical $z_1 = 1 \rightarrow 5200, z_2 = z_1 \rightarrow z_1 + 10000, z_3 = z_2 \rightarrow z_2 + 50000$ because I realized that z_1, z_2, z_3 are sometimes quite far apart.

There were in total 39607 solutions generated. They exhibit some interesting facts. First, the difference between the smallest and largest number can be quite large. Second, there is some sort of clustering going on, *e.g.* z_1 and z_2 are close together and z_3 and w are also close together but the gap between z_1, z_2 and z_3, w is quite big. Third, there's no two numbers are the same, *i.e.* $2z^3 + z_3^3 = w^3$ has no solution.

Finally, there are always an even number of odd numbers among z_1, z_2, z_3, w which means that the sum total of all four is always even. This is actually constructive because this means that we can do this

$$(a - z)^3 + (b + y)^3 + (b - y)^3 = (a + z)^3$$

we can do this because $(a + z) + (a - z) = 2a$, $(b + y) + (b - y) = 2b$ and since there is always an even number of odd numbers among z_1, z_2, z_3, w , this is always true. Continuing in this train of thought, the parameters a, z, b, y are not free, they have to satisfy

$$(a + z)^3 - \{(a - z)^3 + (b + y)^3 + (b - y)^3\} = 0$$

$$z^3 + 3a^2z - 3by^2 - b^3 = 0$$

$$\rightarrow z^3 + 3a^2z = b^3 + 3y^2b$$

Some words about convention. I sorted each solution set z_1, z_2, z_3, w in ascending order, *i.e.* $z_1 < z_2 < z_3 < w$. I then paired them up according their even/odd-ness in ascending order as well. I first search if z_1 and z_2 is of the same evenness, if so $z_1 + z_2 = 2b, z_2 - z_1 = 2y$. We cannot assign z_1, z_2 to a, z because z_1, z_2 are the smallest numbers of the lot, a, z are derived from z_3, w . If z_1 and z_2 is *not* of the same evenness, I tried z_1 and z_3 , and if they're still not of the same evenness I'd try z_1, w and derive a, z, b, y accordingly.

The result was quite surprising, $b \geq z$, $b > y$, $b + y \leq a + z$, always! Also $b - z$ is always a multiple of 3, *i.e.* $b - z = 3k$ for some integer k . This is not at all surprising by

the way, let $b - z = q \rightarrow b = z + q$

$$\begin{aligned}
b^3 - z^3 &= 3a^2z - 3y^2b \\
(z + q)^3 - z^3 &= 3(a^2z - y^2b) \\
z^3 + 3z^2q + 3zq^2 + q^3 - z^3 &= 3(a^2z - y^2b) \\
q^3 &= 3(a^2z - y^2b - z^2q - zq^2) \\
\Rightarrow q &= 3h
\end{aligned}$$

This also sheds some light of the twin solutions. Let's take for example the first row in Table. II, the first twin is

$$\begin{aligned}
3^3 + 36^3 + 37^3 &= 46^3 \\
\rightarrow (b - y = 3)^3 + (a - z = 36)^3 + (b + y = 37)^3 &= (a + z = 46)^3
\end{aligned}$$

with $a = 41, z = 5, b = 20, y = 17$, while the second twin is given by

$$\begin{aligned}
27^3 + 30^3 + 37^3 &= 46^3 \\
\rightarrow (b - y = 27)^3 + (a - z = 30)^3 + (b + y = 37)^3 &= (a + z = 46)^3
\end{aligned}$$

with $a = 38, z = 8, b = 32, y = 5$.

In conclusion we can the above as a four-parameter algorithm

$$(a - z)^3 + (z + (3k + y))^3 + (z + (3k - y))^3 = (a + z)^3$$

with an extra constraint of

$$(3k)^3 = 3(a^2z - y^2(z + 3k) - z^2(3k) - z(3k)^2)$$

and since this constraint is cubic, this parameterization is not very useful.

Natural selection

Despite the limitation of simply guessing the algorithm for $a^3 + b^3 + c^3 = d^3$, other equations can benefit from it naturally, *e.g.* $a^4 + b^3 + c^3 = d^4$

$$\begin{aligned}
(y + z)^4 - (y - z)^4 &= 8y^3z + 8yz^3 \\
\rightarrow (y^3 - z^3)^4 + (2y^3z)^3 + (2yz^3)^3 &= (y^3 + z^3)^4
\end{aligned}$$

and less natural for the 5th power

$$(y+z)^5 - (y-z)^5 = 2z^5 + 20y^2z^3 + 10y^4z$$

the factor of 2 (just like that of the 3rd power) killed it. But, we can always do some substitutionary atonement $y \rightarrow 20y^3, z \rightarrow 2z^3$ and massage it into

$$(20y^3 + 2z^3)^5 - (20y^3 - 2z^3)^5 = (2z^2 \cdot 2z^3)^3 + (20y^2 \cdot 2z^3)^3 + (20y^3)^4(10 \cdot 2z^3)$$

i.e. an algorithm for $a^3 + b^3 + c^4d + e^5 = f^5$, except that this time we have a product of terms c^4d that the preceding ones didn't. Things get hairy when we go to power of 7

$$(y+z)^7 - (y-z)^7 = 2z^7 + 42y^2z^5 + 70y^4z^3 + 14y^6z$$

Going with three terms doesn't get much better but it is now salvageable

$$\begin{aligned} (x+y+z)^7 - (-x+y+z)^7 - (x-y+z)^7 - (x+y-z)^7 \\ = 56xyz(3(x^2+y^2+z^2)^2 + 4(x^2y^2+x^2z^2+y^2z^2)) \\ = 2 \cdot 7 \cdot 3xyz \cdot 2^2(x^2+y^2+z^2)^2 + 2^3x^3y^3z \cdot 7 \cdot 4 + 2^3x^3yz^3 \cdot 7 \cdot 4 + 2^3xy^3z^3 \cdot 7 \cdot 4 \\ = (42xyz)(2p^2)^2 + (2xy)^3(28z) + (2xz)^3(28y) + (2yz)^3(28x) \end{aligned}$$

where $p^2 = x^2 + y^2 + z^2$ whose algorithm will be discussed later. You really need to be creative to appease higher powers :)

Along the same line, $a^3 + b^3 + c^3 = d^3$ can be post processed

$$(a+z)^3 - (a-z)^3 = 2z^3 + 6a^2z$$

substituting $z \rightarrow 2z^3, a \rightarrow 2 \cdot 6a^3$ we get

$$\begin{aligned} (2 \cdot 6a^3 + 2z^3)^3 &= (2z^2)^3(2z^3) + (6a^2 \cdot 2z)^3 + (2 \cdot 6a^3 - 2z^3)^3 \\ \rightarrow (6a^3 + z^3)^3 &= (z^2)^3(2z^3) + (6a^2 \cdot z)^3 + (6a^3 - z^3)^3 \end{aligned}$$

which is an algorithm for $a^3e + b^3 + c^3 = d^3$, not bad. Well, although this game gets really old really quickly, let's continue. The highest power that can still give some solace is eight

$$(a+z)^8 - (a-z)^8 = 16az(z^2 + a^2)^3 + 64a^3z^3(z^2 + a^2)$$

It's tempting to conform $(z^2 + a^2)$ into $z^2 + a^2 = w^2$. To continue though, it requires $a \rightarrow 16^2 a^6, z \rightarrow z^3$ to make $16az(z^2 + a^2)^3 \rightarrow (16a^2 z \cdot w^2)^3$. But this puts too many constraints. First, $z^2 + a^2 = w^2$ means that (at least) $m^2 + n^2 = z^3$, which might not be solvable. Unless we find a solution, we will get products of terms $(16az)w^6$ for $16az(z^2 + a^2)^3$ and $(4a)3w^2$ for $64a^3 z^3(z^2 + a^2)$.

A better choice will be $a \rightarrow 4a^3, z \rightarrow z^3$, this way we'll get less products of terms

$$(4a^3 + z^3)^8 - (4a^3 - z^3)^8 = (4az \cdot (z^6 + 16a^6))^3 + (4 \cdot 4a^3 \cdot z)^3(z^6 + 16a^6)$$

which is an algorithm for $a^8 + b^3 + c^3 d = e^8$. One last example but this time with three variables

$$(x + y + z)^5 - (-x + y + z)^5 - (x - y + z)^5 - (x + y - z)^5 = 80xyz(x^2 + y^2 + z^2)$$

We can easily find an algorithm for $x^2 + y^2 + z^2 = w^2$, which we will discuss in more details later. The solution is

$$(a^2 + b^2 + c^2)^2 = (2ab)^2 + (2ac)^2 + (-a^2 + b^2 + c^2)^2$$

Even with this proviso, we still have difficulty recasting $80xyz(x^2 + y^2 + z^2) \rightarrow u^n$. For example, since we can transmute $x^2 + y^2 + z^2 = w^2$, we might want to upgrade $x \rightarrow 80x^2, y \rightarrow y^2, z \rightarrow z^2$ to transform $80xyz(x^2 + y^2 + z^2) \rightarrow 80^2 x^2 y^2 z^2 w^2$.

However, this puts a constraint on $-a^2 + b^2 + c^2 = 80x^2$ which might not be solvable; $x^2 + y^2 + z^2 = w^2$ requires that $x^2 = -a^2 + b^2 + c^2$ which we will discuss later. Thus it is more than likely that we have have a product of terms for $80xyz(x^2 + y^2 + z^2)$.

Rhythm of the rain

The case of power of two (*i.e.* the Pythagorean equation) is special because we can extend the formula for any number of terms. For example, let's go back to $x^2 + y^2 + z^2 = w^2$, if we expand

$$\begin{aligned}(a + b + c)^2 &= a^2 + b^2 + c^2 + 2ab + 2ac + 2bc \\ (-a + b + c)^2 &= a^2 + b^2 + c^2 - 2ab - 2ac + 2bc\end{aligned}$$

the difference between those two quantities will only contain terms multiplying a

$$(a + b + c)^2 - (-a + b + c)^2 = 4ab + 4ac$$

And here you see the pattern, if we extend to any number of terms $z_1^2 + z_2^2 + \dots + z_n^2 = w^2$ we can use the same result above

$$\begin{aligned} (a_1 + a_2 + \dots + a_n)^2 - (-a_1 + a_2 + \dots + a_n)^2 &= 4a_1a_2 + 4a_1a_3 + \dots + 4a_1a_n \\ \rightarrow (a_1^2 + a_2^2 + \dots + a_n^2)^2 - (-a_1^2 + a_2^2 + \dots + a_n^2)^2 &= (2a_1a_2)^2 + (2a_1a_3)^2 + \dots + (2a_1a_n)^2 \end{aligned}$$

where we have done the usual trick $a_i \rightarrow a_i^2$ going to the second line. Thus the generic algorithm for $z_1^2 + z_2^2 + \dots + z_n^2 = w^2$ is

$$\begin{aligned} z_i &= 2a_1a_{i+1}, \quad 1 \leq i < n \\ z_n &= -a_1^2 + \sum_{j=2}^n a_j^2 \\ w &= \sum_{k=1}^n a_k^2 \end{aligned}$$

well of course you can choose any other a 's beside a_1 as the one having the minus sign, also n is good from 2 onwards.

We can easily extend (or maybe distend) this even further to $z_1^{N_1} + \dots + z_{n-1}^{N_{n-1}} + z_n^2 = w^2$ with N_1, \dots, N_{n-1} any natural numbers. We simply do the following

$$\begin{aligned} z_i &= 2a_1^{M/N_i}a_{i+1}, \quad 1 \leq i < n, \quad M = \prod_{j=1}^{n-1} N_j \\ z_n &= -2^{M-2}a_1^M + \sum_{j=1}^{n-1} a_{j+1}^{N_j} \\ w &= 2^{M-2}a_1^M + \sum_{k=1}^{n-1} a_{k+1}^{N_k} \end{aligned}$$

i.e. we merely substitute $4a_1a_{j+1} \rightarrow 4 \cdot 2^{M-2}a_1^M a_{j+1}^{N_j} = (2a_1^{M/N_j}a_{j+1})^{N_j}$. We will discuss more about extending things to generic power in the next section, but this shows that Pythagoras is one flexible fellow, it can be effortlessly expanded to numerous directions from just a simple guess.

We can say there's sorta rhythm to this. No such swing with the power of 3, just the rain. Well maybe there's a bit of rhythm but it is ugly. We have to start with 3 terms

$$(a + b + c)^3 - \sum [a, b, c]^3 = 24abc$$

where $\sum[a, b, c]^3$ is the sum of

$$(-a + b + c)^3 + (a - b + c)^3 + (a + b - c)^3$$

i.e. we cycle the minus sign through each term. If we had 4 terms we would have

$$2(a + b + c + d)^3 - \sum[a, b, c, d]^3 = 24(abc + abd + acd + bcd)$$

First we need a factor of 2 in front of $(a+b+c+d)^3$ because each parameter in $\sum[a, b, c, d]^3$, *e.g.* a^3 , is multiplied by 2, $-a^3 + a^3 + a^3 + a^3 = 2a^3$, the same with b^3, c^3, d^3 . The RHS is just the combination of 4 choose 3. The general formula is thus

$$(n - 2)(a_1 + a_2 + \dots + a_n)^3 - \sum[a_1, \dots, a_n]^3 = 24C[a_1, \dots, a_n]$$

where $C[a_1, \dots, a_n]$ is the combination of n choose three $a_1a_2a_3 + a_1a_2a_4 + \dots + a_{n-2}a_{n-1}a_n$ and here is the source of the problem.

For simplicity let $n = 4$, so we have $24C[a, b, c, d] = 24(abc + abd + acd + bcd)$. We can promote $a \rightarrow 3^2a^3, b \rightarrow b^3, c \rightarrow c^3, d \rightarrow d^3$. The first three terms become $(2 \cdot 3abc)^3 + (2 \cdot 3abd)^3 + (2 \cdot 3acd)^3$, however, this ruins the last term $24bcd \rightarrow 24(bcd)^3$.

No holy tri(3)nity $\neg \setminus (^{\circ} _ o) / ^{-}$

If $n \geq 5$, we will start having trouble at $24a_2a_3a_4$ because we have promoted $a_2, \dots, a_n \rightarrow a_2^3, \dots, a_n^3$ to deal with $24a_1a_ja_k \rightarrow 8 \cdot 27a_1^3a_j^3a_k^3, j, k > 1$.

Thus for a power of 3 it seems that only $a^3 + b^3 + c^3 + d^3 = e^3$ is bueno. Another significant difference is that the number of a 's is *not* the same as the number of z 's, *e.g.* 3 a 's generate $3 + 1 = 4$ z 's while 4 a 's generate $4 + 4 = 8$ z 's. The next difference from the power of 2 is that the above generic formula is only good for $n \geq 4$ compared to $n \geq 2$ for power of 2. Also for the power of three, the number of terms $z_1^3 + \dots + z_n^3 = w^3$ increases factorially and not to mention that there's a constant factor $(n - 2)$ in front of $(a_1 + a_2 + \dots + a_n)^3$.

There is still something we can do for power of 3, we can always massage

$$24a_i^3a_j^3a_k^3 = (2a_i)^3(3a_j^3a_k^3)$$

giving the generic formula for power of 3, before we state the generic formula there are some bookkeeping to settle

- the number of a 's is not the same as the number of z 's. The number of z 's, denote it m is

$$m = n + \binom{n}{3}$$

n represents $\sum[a_1, \dots, a_n]$, *i.e.* the sum of $(-a_1 + \dots + a_n) + \dots + (a_1 + \dots - a_n)$ with the negative sign cycled through each term. $\binom{n}{3}$ is for $24(a_1a_2a_3 + \dots + a_{n-2}a_{n-1}a_n)$. Note that

$$\binom{n}{3} = \sum_j^{n-2} \binom{n-j}{2}$$

i.e. the terms that contain a_1 as the first factor can be thought of as removing a_1 from the pool as choosing 2 from the rest, and we can repeat the process for terms with a_2 as first factor and so on.

- Some of the z 's have an extra factor multiplying it, *i.e.* $(2a_i)^3(3a_j^3a_k^3) = z^3y$, thus some z 's have an extra factor multiplying it.
- The equation we want to generate an algorithm for is

$$(z_1^3 + \dots + z_n^3) + \left(z_{n+1}^3 + \dots + z_{n+\binom{n-1}{2}}^3\right) + \left(y_1z_{n+\binom{n-1}{2}+1}^3 + \dots + y_lz_m^3\right) = (n-2)w^3$$

where $l = \binom{n}{3} - \binom{n-1}{2}$

The generic formula is then

$$\begin{aligned} z_i &= [3^2a_1^3, \dots, a_n^3], \quad 1 \leq i \leq n \\ z_j &= 2 \cdot 3a_1a_pa_q, \quad n+1 \leq j \leq n + \binom{n-1}{2}, \quad 2 \leq p, q \leq n \\ z_k &= 2a_r, \quad n + \binom{n-1}{2} + 1 \leq k \leq m, \quad 2 \leq r \leq n \\ y_q &= 3a_s^3a_t^3, \quad 1 \leq q \leq \binom{n}{3} - \binom{n-1}{2}, \quad 2 \leq s, t (\neq r) \leq n \\ w &= (3^2a_1^3 + \dots + a_n^3) \end{aligned}$$

and as we can see it is a lot more complicated than that of the power of 2.

To infinity and beyond

In the above examples we were stopped in our tracks by constants

$$(a + z)^3 - (a - z)^3 = 6a^2z + 2z^3$$

Since 2 is prime we can't massage $2z^3 \rightarrow u^3$. There is a way though to cast it another way $2z^3 \rightarrow u^N$ for some integers u and N (we have actually done this for the generic power of 2 earlier). We know that 2 is prime so

$$\begin{aligned} z &= 2^m g^n \\ \rightarrow 2z^3 &= 2^{3m+1} g^{3n} \end{aligned}$$

What we want now is

$$\begin{aligned} 3m + 1 &= NK \\ 3n &= NQ \\ \rightarrow 2z^3 &= 2^{3m+1} g^{3n} = (2^K g^Q)^N \end{aligned}$$

However, we have another obstacle, $6a^2z$ has to be recast into $6a^2z \rightarrow v^q$, in this case we want v^3 since we are dealing with power of 3. We can apply the same process to a and indeed we have, $a \rightarrow 6a^3$, since $6a^2 = (6a^2)^3$ is already a power of 3

$$z = 2^m g^n = d^3 \rightarrow m = 3S, n = 3T$$

m and n have to be multiples of 3. This in turn means that $3m + 1 = 9S + 1 = NK \implies N$ is any factor of $9S + 1$ and that's it, we have arrived, we can now set $2z^3$ into u^N with N any factor of $9S + 1$. How about $3n = 9T = NQ$? That can be easily dealt with since both R and Q are free parameters, the more than obvious choice is

$$Q = 9 \quad T = N$$

Four our beloved cubic friend $(a - z)^3 + 6a^2z + 2z^3 = (a + z)^3$, we can conveniently choose $S = 1$, $9S + 1 = 10$, $T = N = 2$, $K = 5$, $Q = 9$, $m = 3S = 3$, $n = NQ/3 = 6$, $z \rightarrow 2^3 z^6$

$$\begin{aligned} (6a^3 - 2^3 z^6)^3 + 6^3 a^6 2^3 z^6 + 2 \cdot 2^9 z^{18} &= (6a^3 + 2^3 z^6)^3 \\ (6a^3 - 2^3 z^6)^3 + (6a^2 2z^2)^3 + (2^5 z^{14})^2 &= (6a^3 + 2^3 z^6)^3 \\ \rightarrow a^3 + b^3 + c^2 &= d^3 \end{aligned}$$

The only ironic thing about this is that N can never be 3 (or a multiple of 3) since if $N = 3 \rightarrow 3(K - m) = 1$ which is impossible.

The generic algorithm to massage $C^q z^p$ into u^N (with C some constant factor) is

$$\begin{aligned} z &= C^m g^n \\ C^q z^p &= C^{pm+q} g^{pn} = u^N \\ \Rightarrow pm + q &= NK \\ \Rightarrow pn &= NQ \end{aligned}$$

$N = \{pm + q\}$ with $n = N$, $Q = p$ where I have used the notation $\{A\}$ as “any factor of A ”.

With this we can extend known algorithms to “infinity and beyond” and our first guinea pig is gonna be Euclid’s algorithm

$$\begin{aligned} (a^2 + b^2)^2 - (a^2 - b^2)^2 &= 4a^2 b^2 \\ a &\rightarrow 4^m g^n \\ 4a^2 &= 4^{2m+1} g^{2n} \\ 2m + 1 &= NK \\ 2n &= NQ \\ n = N \quad Q &= 2 \end{aligned}$$

$N = \{2m + 1\}$ means N is any factor of any odd number, we can also cast this differently

$$\begin{aligned} a &\rightarrow 2^m g^n \\ 4a^2 &= 2^2 a^2 = 2^{2m+2} g^{2n} \\ 2m + 2 &= NK \\ 2n &= NQ \\ n = N \quad Q &= 2 \end{aligned}$$

which means that N is any factor of any even number, a factor that is any factor of any odd number and any even number is any number (actually we only need it to be a factor of any even number for it to be any number since even numbers are $2m$ with m

any integer), *i.e.* we can recast $4a^2b^2$ into u^N with N any number :) the only thing to do is $b \rightarrow b^N$, $(b^N)^2 = (b^2)^N$. This is actually interesting, Pythagorean equation is extra special, it not only works with $a^2 + b^2 = c^2$ but also $a^2 + b^N = c^2$ for *any* N .

Pythagoreans surely are a lucky lot, not only that they can be easily extended indefinitely in terms of the number of terms in $z_1^2 + \dots + z_n^2 = w^2$ but also in the exponent $z_1^2 + z_2^N = w^2$. Remember, we started this whole business from nothing but a simple guess!

Not so with the power of 3, a simple guess won't do. Looking at the example above, $a^3 + b^3 + c^N = d^3$, N is limited to the factors of $9S + 1$, for example for $S = 2, 4, 8, 12, 14, 18, 20$ the combination $9S + 1$ is actually a prime number, they are 19, 37, 73, 109, 127, 163, 181 respectively.

There's a much easier way to establish $z_1^2 + z_2^N = w^2 \longrightarrow z_1 = r - q, w = r + q$

$$\begin{aligned} w^2 - z_1^2 &= (r + q)^2 - (r - q)^2 \\ \rightarrow z_2^N &= 4qr \end{aligned}$$

we can then conveniently set $q = 2^{N-2}d^N$, $r = h^N$ such that $z_2 = 2dh$, *i.e.*

$$(h^N - 2^{N-2}d^N)^2 + (2dh)^N = (h^N + 2^{N-2}d^N)^2$$

Note: going to $z_1^M + z_2^N = w^2$ is much harder because even though casting $r - q = y^M$ is straightforward, satisfying it while requiring $z_2^N = 4qr$ at the same time is challenging.

Silly tally

Now comes the time to gather what I've found. In this section I'll list all the algorithms I discovered, sorted in ascending power (some might not have been previously discussed because I just stumbled upon them). Recall that I started this whole business from a silly guess, $(a + z)^n - (a - z)^n = \Delta$, thus this tally is indeed rather silly :)

- power of 2

1. $z_1^2 + z_2^2 = w^2$ (the original Pythagorean) with two and *three* degrees of freedom

Two degrees of freedom

$$z_1 = 4a - ab^2, \quad z_2 = 4ab, \quad w = 4a + ab^2$$

Three degrees of freedom

$$z_1 = (a - b - 2|c|)(a - b), \quad z_2 = 2|c|(a - b) - 2c^2$$

$$w = a^2 - 2(b + |c|)a + b^2 + 2b|c| + 2c^2$$

2. $z^2 - zy = 2w^2$ (A.I.P: An Incarnation of the Pythagorean)

$$z = 2a^2, \quad y = a^2 + 2ab - b^2, \quad w = a^2(a - b)^2$$

3. $z^2p^4 - 2zp^2q^2y + y^2 = p^2q^2 + q^4$ (Y.A.I.P: Yet Another Incarnation of the Pythagorean)

$$z = a^2 - b^2, \quad y = 2b^2, \quad p = 2ab, \quad q = a^2 - b^2$$

4. $z^2 + y^2 = 2w^2$ (A.Y.A.I.P: And Yet Another Incarnation of the Pythagorean)

$$z = a^2 - b^2 + 2ab, \quad y = a^2 - b^2 - 2ab, \quad w = a^2 + b^2$$

5. $2y^2 + 2 = z^2$ (S.Y.A.I.P: Stil Yet Another Incarnation of the Pythagorean)

$$z = \frac{2(a^2 + b^2)}{a^2 - b^2 + 2ab}, \quad y = \frac{b^2 + 2ab - a^2}{b^2 - 2ab - a^2}$$

6. $z_1^2 + \cdots + z_n^2 = w^2$ (crowd funded Pythagorean)

$$z_i = 2a_1a_{i+1}, \quad 1 \leq i < n$$

$$z_n = -a_1^2 + \sum_{j=2}^n a_j^2$$

$$w = \sum_{k=1}^n a_k^2$$

7. $z_1^2 + z_2^N = w^2$ (Buzzlightyear meet Pythagoras)

Choosing factors from odd numbers

$$z_1 = (4^m a^{NQ/2})^2 - b^{2N}, \quad z_2 = 4^K a^Q b^2, \quad w = (4^m a^{NQ/2})^2 + b^{2N}$$

$$N = \{2m + 1\}, \quad NK = 2m + 1, \quad Q = 2$$

Choosing factors from even numbers

$$z_1 = (2^m a^{NQ/2})^2 - b^{2N}, \quad z_2 = 2^K a^Q b^2, \quad w = (2^m a^{NQ/2})^2 + b^{2N}$$

$$N = \{2m + 2\}, \quad NK = 2m + 2, \quad Q = 2$$

A much easier choice

$$z_1 = a^N - 2^{N-2} b^N, \quad z_2 = 2ab, \quad w = a^N + 2^{N-2} b^N$$

8. $z_1^{N_1} + \dots + z_{n-1}^{N_{n-1}} + z_n^2 = w^2$ (Phytagoras on steroids)

$$z_i = 2a_1^{M/N_i} a_{i+1}, \quad 1 \leq i < n, \quad M = \prod_{j=1}^{n-1} N_j$$

$$z_n = -2^{M-2} a_1^M + \sum_{j=1}^{n-1} a_{j+1}^{N_j}$$

$$w = 2^{M-2} a_1^M + \sum_{k=1}^{n-1} a_{k+1}^{N_k}$$

• power of 3

1. $z_1^3 + z_2^3 + z_3^3 = w^3 \rightarrow$ no straightforward generic algo, more than just silly guesses required :)

$$z_1 = a, \quad z_2 = b, \quad z_3^3 = 3ab(b + a), \quad w = a + b \quad \longrightarrow \quad 1, 6, 8, 9$$

$$z_1 = a, \quad z_2 = a + 1, \quad z_3 = a + 2, \quad w = a + 3 \quad \longrightarrow \quad a = 3$$

$$z_1 = a, \quad z_2 = a \cdot x, \quad z_3 = a \cdot y, \quad w = a + z \quad \longrightarrow \quad 1 + z_1^3 + z_2^3 = w^3$$

$$z_1 = a - b, \quad z_2 = b - c, \quad z_3 = c - a, \quad w = 3(a - b)(b - c)(c - a) \quad \longrightarrow \quad 1, 6, 8, 9$$

$$z_1 = a - b, \quad z_2 = b, \quad z_3^3 = b^3 + 6a^2b, \quad w = a + b \quad \longrightarrow \quad 17, 4, 22, 25$$

$$z_1 = a, \quad z_2 = b^{1/3}, \quad z_3 = a - c, \quad w = a + 8c \quad \longrightarrow \quad 6, 32, 33, 41$$

2. $1 + z_1^3 + z_2^3 = w^3$, see Table. I

3. $1 + z_1^3 + z_2^3 + z_3^3 + z_4^3 = w^3$, this was an accident from trying to solve no. 2 above

$$z_1 = a, z_2 = 2a, z_3 = 3a^2, z_4 = 3a^3, w = (3a^3 + 1)$$

4. $z_1^3 + z_2^{2m+1} + 2 \cdot z_3^{2m+1} = w^3$

$$z_1 = 6^m a^{2m+1} - b^{2m+1}, z_2 = 6a^2 \cdot b, z_3 = b^3, w = 6^m a^{2m+1} + b^{2m+1}$$

5. $z_1^3 + z_2^{2m+1} + y \cdot z_3^N = w^3$, N is any factor of $2(2m+1)$

$$z_1 = 6^m a^{2m+1} - b^{2m+1}, z_2 = 6a^2 \cdot b, y = 2b^{2m+1}$$

$$z_3 = b^{2(2m+1)/N}, w = 6^m a^{2m+1} + b^{2m+1}$$

6. $z_1^3 + z_2^3 + z_3^3 + z_4^3 = w^3$

the simplest choice

$$z_1 = -a^3 + b^3 + 3^2 c^3, z_2 = a^3 - b^3 + 3^2 c^3, z_3 = a^3 + b^3 - 3^2 c^3$$

$$z_4 = 2 \cdot 3 \cdot abc, w = a^3 + b^3 + 3^2 c^3$$

another simple choice

$$z_1 = -a^3 + 3b^3 + 3c^3, z_2 = a^3 - 3b^3 + 3c^3, z_3 = a^3 + 3b^3 - 3c^3$$

$$z_4 = 2 \cdot 3 \cdot abc, w = a^3 + 3b^3 + 3c^3$$

there are many other bloated solutions as discussed previously.

7. $z_1^3 + z_2^3 + z_3^3 + z_4^N = w^3$

the simplest choice

$$z_1 = -a^N + b^N + 2^{N-3} 3^{N-1} c^N, z_2 = a^N - b^N + 2^{N-3} 3^{N-1} c^N$$

$$z_3 = a^N + b^N - 2^{N-3} 3^{N-1} c^N, z_4 = 2 \cdot 3 \cdot abc$$

$$w = a^N + b^N + 2^{N-3} 3^{N-1} c^N$$

8. *Generic power of 3*

$$(z_1^3 + \cdots + z_n^3) + \left(z_{n+1}^3 + \cdots + z_{n+\binom{n-1}{2}}^3 \right) + \left(y_1 z_{n+\binom{n-1}{2}+1}^3 + \cdots + y_l z_m^3 \right) = (n-2)w^3$$

where $n \geq 3$, $l = \binom{n}{3} - \binom{n-1}{2}$

$$z_i = [3^2 a_1^3, \dots, a_n^3], \quad 1 \leq i \leq n$$

$$z_j = 2 \cdot 3 a_1 a_p a_q, \quad n+1 \leq j \leq n + \binom{n-1}{2}, \quad 2 \leq p, q \leq n$$

$$z_k = 2a_r, \quad n + \binom{n-1}{2} + 1 \leq k \leq m, \quad 2 \leq r \leq n$$

$$y_q = 3a_s^3 a_t^3, \quad 1 \leq q \leq \binom{n}{3} - \binom{n-1}{2}, \quad 2 \leq s, t (\neq r) \leq n$$

$$w = (3^2 a_1^3 + \dots + a_n^3)$$

9. $z_1^3 + z_2^3 + z_3^N = w^3$ (to cube-finity and beyond)

$$z_1 = 6a^3 - 2^{3S} b^{NQ/3}, \quad z_2 = 6a^2 2^S b^{NQ/9}, \quad z_3 = 2^K b^Q, \quad w = 6a^3 + 2^{3S} b^{NQ/3}$$

$$N = \{9S + 1\}, \quad NK = 9S + 1, \quad Q = 9$$

- power of 4

$$1. \quad z_1^4 + z_2^3 + z_3^3 = w^4$$

$$z_1 = a^3 - b^3, \quad z_2 = 2a^3 b, \quad z_3 = 2ab^3, \quad w = a^3 + b^3$$

- power of 5

$$1. \quad z_1^3 + z_2^3 + y \cdot z_3^4 + z_4^5 = w^5$$

$$z_1 = 4b^5, \quad z_2 = 40a^2 b^3, \quad y = 20b^3, \quad z_3 = 20a^3, \quad z_4 = 20a^3 - 2b^3, \quad w = 20a^3 + 2b^3$$

$$2. \quad z_1^5 + z_2^5 + z_3^5 + y_1 \cdot z_4^3 + y_2 \cdot z_5^3 + y_3 \cdot z_6^3 = w^5$$

$$z_1 = -a + b + c, \quad z_2 = a - b + c, \quad z_3 = a + b - c$$

$$y_1 = 80bc, \quad y_2 = 80ac, \quad y_3 = 80ab$$

$$z_4 = a, \quad z_5 = b, \quad z_6 = c$$

$$w = a + b + c$$

- power of 6

$$1. \quad z_1^6 + y \cdot z_2^N = w^6$$

$$z_1 = 2^{N-2} a^N - b^N, \quad y = (3 \cdot 2^{2(N-2)} a^{2N} + b^{2N})(2^{2(N-2)} a^{2N} + 3b^{2N})$$

$$z_2 = 2ab, \quad w = 2^{N-2} a^N + b^N$$

- power of 7

$$1. z_1^7 + z_2^7 + z_3^7 + y_1 \cdot z_4^2 + y_2 \cdot z_5^3 + y_3 \cdot z_6^3 + y_4 \cdot z_7^3 = w^7$$

$$z_1 = -a + b + c, z_2 = a - b + c, z_3 = a + b - c$$

$$y_1 = 42abc, y_2 = 28c, y_3 = 28b, y_4 = 28a$$

$$z_4 = 2p^2, z_5 = 2ab, z_6 = 2ac, z_7 = 2bc$$

$$w = a + b + c, p^2 = a^2 + b^2 + c^2$$

here we need the formula for $a^2 + b^2 + c^2 = p^2$ above, *i.e.* $a = 2rs, b = 2rt, c = (-r^2 + s^2 + t^2), p = (r^2 + s^2 + t^2)$ but to avoid clutter I didn't substitute them into the above formula. This is the first time we have a cascade of algorithms

- power of 8

$$1. z_1^8 + z_2^3 + y \cdot z_3^3 = w^8$$

$$z_1 = 4a^3 - b^3, z_2 = 4ab(b^6 + 16a^6), y = b^6 + 16a^6, z_3 = 16a^3b, w = 4a^3 + b^3$$

Note that variables used above are either integers or natural numbers which can be clearly deduced from their associated context.

Most of the formulae above can be easily modified into a more general power, *e.g.*

$z_1^4 + z_2^3 + z_3^3 = w^4 \longrightarrow z_1^4 + z_2^{3N} + y \cdot z_3^N = w^4$. The starting point was

$$(a + b)^4 - (a - b)^4 = 8ab^3 + 8a^3b$$

We can of course substitute $a \rightarrow 8^{3N-1}a^{3N}, b \rightarrow b^N$ such that

$$\begin{aligned} 8ab^3 + 8a^3b &\Rightarrow 8^{3N}a^{3N}b^{3N} + 8^{9N-2}a^{9N}b^N \\ &= (8ab)^{3N} + (8^{9N-2}a^{8N})(ab)^N \\ &= z_2^{3N} + y \cdot z_3^N \end{aligned}$$

But I feel that this is too contrived, $8 = 2^3$ and I went for the most natural adaptation. This was my philosophy in building the table above although I went for the generalization if it is warranted. You can play this game endlessly so the rest is left as an exercise for the readers (beside myself, if any).

Hitting the books

So that's all I've got for fooling around with a "simple guess", the phrase I kept repeating in the previous several sections. There's no big discovery obviously, except for a few interesting things that I have mentioned above.

I then checked the literature about the "Cubic Pythagorean Equation", $a^3+b^3+c^3 = d^3$. There are many algorithms! There are even many one parameter algorithms. The really interesting point is that these algorithms usually only cover only certain solutions not all of them.

This explains why it was hard getting the overall vibe from the solutions because we need to group them, for example, in one of the one-parameter algorithm, the 1, 6, 8, 9 is on its own, it is the only solution the algorithm produces that contains 1.

Getting one parameter algorithms is actually not that hard but it surely is tedious. Let's start with Pythagoras equation. What we want to do here is cast each number into a polynomial $\sum a_n x^n$. The only question is how high a degree of a polynomial we want. We can even extend this to two or more parameters

$$\sum_{n=0, m=0}^{N_1, N_2} a_{mn} x^n y^m$$

where a_{mn} is some sort of a coefficient matrix. This formulation is more generic than $(\sum a_n x^n)(\sum b_m y^m)$ because in the case of product of sums we cannot have $a_{N_1} x^{N_1} + b_{N_2} y^{N_2}$ without having $a_{N_1} b_{N_2} x^{N_1} y^{N_2}$ as well. On the other hand, a_{mn} allows anything :) In fact, Euler's formula for the cubic Pythagorean is exactly like it, it contains y^4 and x^4 but it doesn't contain the product $x^4 \cdot y^4$. The only problem is that solving these coefficients, a_{mn} , might be quite laborious.

Let's start with Pythagoras, $z_1^2 + z_2^2 = w^2$

$$w = b_1 x + a_1$$

$$z_1 = b_2 x + a_2$$

$$z_2 = b_3 x + a_3$$

Substituting this into the Pythagorean equation $z_1^2 + z_2^2 - w^2 = 0$ and grouping coefficients

of different powers of x

$$x^2 : b_3^2 + b_2^2 - b_1^2 = 0$$

$$x^1 : 2a_3b_3 + 2a_2b_2 - 2a_1b_1 = 0$$

$$x^0 : a_3^2 + a_2^2 - a_1^2 = 0$$

Two things to note here. Coefficients of the highest and lowest powers are themselves Pythagoreas equations, this is still true with cubic and higher Diophantines and is still obviously true no matter how high the polynomial we use for x . This can be good and bad.

Bad if we don't already have solutions but good that we might be able to generate new solutions from a known one. Going the bad news route, we have to set one of the b 's and a 's to zero, say $b_3 = 0 \rightarrow b_1 = b_2$. Now for a we don't want to set a_3 to zero as well, this is generally a bad strategy as $z_3 = b_3x + a_3 = 0$, this principle still applies for the cubic Pythagorean and higher.

So we need to set one of the b 's and one of the a 's to zero but we want to stagger them so that they don't belong to the same z . However, in the case of Pythagoras with just x this strategy doesn't work, it will only generate a trivial solution.

So now let's go the good news route which is not so good in this case. For example, setting $b_1 = 13, b_2 = 12, b_3 = 5$ generates $a_2 = 12a_1/13, a_3 = 5a_1/13$ which means that we are only generating integer multiples of 5, 12, 13.

This means that we have to start with at least x^2 and indeed we do, starting with at least x^2 (and choosing the bad route) we have the following parameterizations

$$\begin{array}{lll} z_1 = 2x + 2 & z_2 = x^2 + 2x & w = x^2 + 2x + 2 \\ z_1 = -x^2 + x + 2 & z_2 = 2x - \frac{x^3}{2} & w = -\frac{x^3}{2} + x + 2 \\ z_1 = \frac{x^4}{4} - x^3 + 2x & z_2 = -x^2 + 2x + 2 & w = \frac{x^4}{4} - x^3 + 2x + 2 \end{array}$$

Things on the good side is quite promising too, say you have a solution $A^2 + B^2 = C^2$, $A <$

$B < C$ you can create new solutions $z_1^2 + z_2^2 = w^2$, $z_1 < z_2 < w$ using

$$w = A \cdot x^2 + b_1x + a_1$$

$$z_2 = B \cdot x^2 + b_2x + a_2$$

$$z_1 = C \cdot x^2 + b_3x + a_3$$

$$a_1 = \frac{C \cdot b_2^2 - 2B \cdot b_1b_2 + C \cdot b_1^2}{(2A)^2}$$

$$a_2 = -\frac{B \cdot b_2^2 - 2C \cdot b_1b_2 + B \cdot b_1^2}{(2A)^2}$$

$$a_3 = -\frac{b_2^2 - b_1^2}{4A}$$

$$b_3 = -\frac{B \cdot b_2 - C \cdot b_1}{A}$$

b_1, b_2 remain free parameters. We can also substitute the known solutions A, B, C into a_1, a_2, a_3 instead with a similar outcome. Since b_1, b_2 , and x are free parameters, we immediately have a three-parameter solution. If we substitute the Babylonian algorithm for A, B, C we will then have a *five*-parameter algorithm.

For the cubic Pythagorean, $z_1^3 + z_2^3 + z_3^3 = w^3$, $z_1 < z_2 < z_3 < w$, the minimum would be x^2 , no solution with just x . The solution I found was a generalization of a known solution, the starting point is

$$w = d_1x^3 + c_1x^2 + b_1x + a_1$$

$$z_1 = d_2x^3 + c_2x^2 + b_2x + a_2$$

$$z_2 = d_3x^3 + c_3x^2 + b_3x + a_3$$

$$z_3 = d_4x^3 + c_4x^2 + b_4x + a_4$$

To prep the solution we need to set $d_1 = d_4 = 1$, $d_2 = d_3 = 0$, $a_3 = a_1$, $a_2 = a_4 = 0$ and then $c_3 = c_4 = c_1$ followed by $b_3 = b_4 = b_1$, $b_2 = 0$, *i.e.*

$$w = x^3 + c_1x^2 + b_1x + a_1$$

$$z_1 = c_2x^2$$

$$z_2 = c_1x^2 + b_1x + a_1$$

$$z_3 = x^3 + c_1x^2 + b_1x$$

All this might seem random but if you do it you'll see the progression. First you'll need to set up a 's and d 's because they themselves have to satisfy the cubic Pythagorean. Once you do this, you'll see that other variables have to take certain values.

For our current assignment, the coefficients of the various powers of x are

$$x^6 : c_2^3 + c_1^3 - 3a_1 = 0$$

$$x^5 : 3b_1c_1^2 - 6a_1c_1 = 0$$

$$x^4 : 3b_1^2c_1 - 6a_1b_1 = 0$$

$$x^3 : b_1^3 - 3a_1^2 = 0$$

Solving them altogether generates the following one parameter algorithm for the cubic Pythagorean

$$a_1 = 3 \left(\frac{c_1}{2} \right)^3, \quad b_1 = 3 \left(\frac{c_1}{2} \right)^2, \quad c_2 = \frac{c_1}{2}$$

$$w = x^3 + c_1x^2 + 3 \left(\frac{c_1}{2} \right)^2 x + 3 \left(\frac{c_1}{2} \right)^3$$

$$z_1 = \left(\frac{c_1}{2} \right) x^2$$

$$z_2 = c_1x^2 + 3 \left(\frac{c_1}{2} \right)^2 x + 3 \left(\frac{c_1}{2} \right)^3$$

$$z_3 = x^3 + c_1x^2 + 3 \left(\frac{c_1}{2} \right)^2 x$$

while c_1 remains a free parameter. This means that we effectively have a two-parameter solution, x and c_1 .

We can now see how we have multiple algorithms. For example, choosing a parameterization with up to x^3 means that we have $4 \times 4 = 16$ coefficients, a_i, b_i, c_i, d_i , $1 \leq i \leq 4$. Since we are dealing with cubes we then have x^j , $0 \leq j \leq 12$ which means that we have 13 constraints for 16 coefficients, some freedom is warranted. We must be careful however, as the requirement that all coefficients must be integers is a very strong condition.

One interesting thing to note here is that solving the coefficients of the polynomials x^n in itself is actually solving a set of Diophantine equations although in this case they are **coupled** Diophantine equations. The good thing is that we *just* need to find *one* solution not all of them, it may sometimes even be a trivial solution, *e.g.* in the cubic Pythagorean

example above we set $d_1 = d_4 = 1$, $d_2 = d_3 = 0$ which means that $-d_1^3 + d_2^3 + d_3^3 + d_4^3 = 0 \rightarrow 1^3 + 0 + 0 + 1^3 = 0$, a trivial solution. *One solution to rule them all :*)

Another interesting thing is that we can generate two-parameter algorithms from one-parameter algorithms. In Chamberland's paper there is an interesting formula

$$(ax^2 + cx - c)^3 + (bx^2 - bx + d)^3 + (cx^2 - cx - a)^3 = (dx^2 - bx + b)^3$$

where a, b, c, d are the solution of cubic Pythagorean $a^3 + b^3 + c^3 = d^3$, however, there's an extra constraint on a, b, c, d

$$c(c^3 - a^3) = b(d^2 - b^2)$$

Now, if we use the parameterization of a, b, c, d that obey the above constraint, we can generate a two-parameter algorithm by substituting it into the above algorithm. This process can *potentially* be repeated indefinitely if the new parameterization utilizes solutions of the cubic Pythagorean. Chamberland actually provides the *only* parameterization of a, b, c, d obeying the above constraint.

Intuition rumination

This short section is dedicated to the reflection of my numerical digression above. My biggest mistake was sticking with that stupid guess mechanism for too long. I was hoping that by starting from $(a + z)^n - (a - z)^n$, a and z will sort themselves out, *i.e.* they will transform into some polynomial automatically.

Sadly, that was not the case. This is actually my personal weakness, I always want to know the deeper reason behind anything. And so I was sucked into trying to understand why those silly guesses wouldn't work. I was also preoccupied with finding some sort of first principles that would allow me to universally generate parametrized algorithms for Diophantine's solutions.

In this kind of problems, there's no use in trying to get a generic principle, speed is all that matters. We just need to try as many potential candidates as fast as possible. Something that I do not like since in solving textbook physics problems you always start with some underlying principle and then work with it.

Now, let's see why this simple guess trick won't do but before that let's why it works for the power of 2. The answer is **l.u.c.k**, $(a + z)^2 - (a - z)^2 = 4az$, 4 is a square and

we have only product of terms, gold mine! This is why we can easily extend this to any number of terms. Other powers are more complicated.

Recall that the formula for a cubic equation is a cube root of a square root $(a+(b^{1/2}))^{1/3}$ and not just a cube root from “completing the cube”?

Looking back, this is actually expected because $(a+b)^n$ contains all polynomials of degree $\leq n$ or in our current lingo, there’s some “Exponent-Inception” going on, *e.g.*

$$(a+b)^5 = ((a+b)^2)^3 = (a^2 + 2ab + b^2)^3$$

Again, we see why $n = 2$ is the lucky one, $(a+b)^2$ only contains squares and power of 1, on the other hand $(a+b)^5$ contains all powers of 5 and below, and from the equation above we can say that in a sense the powers are nested, and as such it is natural that in solving a polynomial of degree 3 we get a cube root of a square root.

This is somewhat reminiscent of Galois theory where the symmetry groups are nested although as to why and how they are precisely related, I do not know.

One more pertinent point, the constant coefficients in expanding a polynomial like this one $(a+b)^n$ are computed from Pascal triangle. And we all know that Pascal triangle is a nested (recursive) structure; you need its previous result before computing the next one. Hold on you say, we can use Newton’s binomial expansion to compute it, we are modern people after all.

True, but binomial expansion requires factorials and a factorial itself is a recursive structure, there’s no explicit formula for a factorial, you have to compute it one step at a time, which is the very definition of recursion. In other words, when we were trying to find algorithms for $z_1^n + \dots + z_m^n = w^m$ what we were trying to do was recast a factorial into a polynomial, *e.g.*

$$\begin{aligned} (a+z)^3 &= \binom{3}{0}a^3 + \binom{3}{1}a^2z + \binom{3}{2}az^2 + \binom{3}{3}z^3 \\ \rightarrow (a+z)^3 - (a-z)^3 &= 6a^2z + 2z^3 \end{aligned}$$

The constant factors, 6 and 2, are (2 times) the binomial expansions (factorials) from $(a+z)^3$ and we are trying to somehow cast it into some polynomial $6 \rightarrow u^3$. It is of no surprise then that this will not always succeed, except for a few serendipitous cases.

If we trace it back even further, this recursivity is derived directly from the definition of exponentiation, *e.g.* 5^n is $5 \times 5 \times 5 \times 5 \times 5$. There is no explicit formula for this, we recurse some multiplications and happily denote it 5^n , in the words of a supermodel “image is powerful but image is also superficial”. (*Note: addition is the basic operation in arithmetic, multiplication is a derived operation, i.e. repeated additions. Taking a number to some power is repeated multiplications, a derived operation of a derived operation, continuing this trend we can define something like $a[n] \equiv a^{a^{\dots}}$, i.e. recursively exponentiating n times and see what kind of number theoretical implications we’ll get from it but this might be taking it too far*)

This recursive business might also be the culprit behind the indomitability of Fermat’s last theorem. Increasing the power by one is not just going one more above the previous one, the recursive nature requires that it is at least “one order of magnitude” more complicated so to say (although the dichotomy between $n = 2$ and $n \geq 3$ is still a mystery to me).

On a related but distant note, the formula for $\sum_{i=1}^N i$ is misleadingly simple. Misleading because it gives us the impression that we can avoid doing N additions by merely calculating $(N^2 + N)/2$. However, this is a pipe dream, sorry Gauss.

We are so used to multiplying things that we don’t even realize it. To multiply things we need to add repeatedly, there’s no other way. Even when we use our familiar algorithm of multiplying each digit what we are really doing is using the multiplication table. And that’s all we do, we memorize the multiplication table for small numbers and we use it to calculate larger multiplications. The notation for multiplication is just that, a notation, a shorthand for $a \times b \equiv \sum_{i=1}^b a$.

The formula for $\sum_i i^n, n \geq 2$ is even more involved as it now concerns recursions. In fact, as far as I know, the formulae for $n \geq 3$ can only be derived recursively from $\sum_i i^2$, see my other article on this.

In short, unrolling a recursion into something straightforward is difficile (if not impossible) and that’s what I was inflexibly trying to do using the simple guess tactic.

Now, why does polynomial work then? Simple, Taylor expansion. Remember that any (well behaved) function can be Taylor expanded including factorials. But more than that,

we now have a lot more flexibility. However, the point here is not about the underlying principle behind all this numerical discourse, the lesson is that different problems require different approaches no one single approach is supreme, sorry string theory.

Farewell to Fermat

The main reason Fermat (or number theory for that matter) is alluring is because the subject matter is widely accessible. But accessibility does not equal simplicity. It is in fact extremely convoluted.

The biggest disappointment is that I was too stubborn to change my strategy but it was fun. Although I must say that this wasn't my best work, I wasn't feeling too well emotionally either, I was kinda depressed. This was mainly my escape; those bursts of excitements when you solved/realized something.

Well, that's all for now with me on this subject, Au Revoir Monsieur Fermat!

=====

What other operations can we add to our beloved arithmetic that might yield useful results?

We already have

1. additions
2. multiplications \Leftarrow repeated additions
3. exponentiations \Leftarrow repeated multiplications
4. factorial (and double factorial) \Leftarrow modified repeated multiplications
5. higher op 1 \Leftarrow repeated exponentiations ???
6. random op 1 \Leftarrow anything we fancy ???

Substraction and division are inverse operators of addition and multiplication respectively while logarithm is the inverse of exponentiation. The interesting question would be what the inverse of factorial is. Denote $!^{-1}$ the inverse of factorial then

$$!^{-1}120 = 5$$

$$!^{-1}15 = \text{undefined?}$$

Although $!^{-1}15$ is undefined, can we still assign some real number value to it? in analogy to the square root of two, $\sqrt{2} = 1.414\dots$

Some potential higher/random ops

1. $[a * b]_{\pm} \equiv a^b \pm b^a$
2. $[a, b]_{\pm} \equiv a \cdot b \pm b \cdot a$, proven useful in quantum mechanics
3. $a !+! b \equiv a + (b - 1) + (a - 2) + (b - 3) + (a - 4) + \dots + 1$
4. $a!b \equiv a \cdot (a + 1) \cdot (a + 2) \cdot \dots \cdot (a + b)$

=====

For $a, b \in \mathbb{Z}$ there's a unique $q, r \in \mathbb{Z}$ such that $a = bq + r$.

Proof. For simplicity let's assume that $a, b \geq 0$, the argument for negative integers is similar. We assume $b \neq a$ otherwise it is trivial, *i.e.* $q = 1, r = 0$. If $b > a$ then $q = 0, r = a$. If $b < a$ then $r < b$. If $r > b$ we can keep subtracting r until it is less than b say

$$\begin{aligned}
 a &= bq + r \\
 &= bq + \underbrace{b + b + b + b + \dots}_{n \text{ number of } b\text{'s}} + (r - \underbrace{b - b - b - b - \dots}_{n \text{ number of } b\text{'s}}) \\
 a &= b(q + n) + r' \\
 a &= bq' + r'
 \end{aligned}$$

where $q' = q + bn$ and $r' = r - bn$. Suppose there are two sets of q, r that satisfy $a = bq + r$, let's denote them $a = bq_1 + r_1$ and $a = bq_2 + r_2$ then

$$\begin{aligned}
 bq_1 + r_1 &= bq_2 + r_2 \\
 b(q_1 - q_2) &= (r_2 - r_1)
 \end{aligned}$$

this means that b divides $(r_2 - r_1)$, *i.e.* $r_2 - r_1 = bm$ for some $m \in \mathbb{Z}$ but since $r_1, r_2 < b$ this is only possible if $m = r_2 - r_1 = 0$ thus

$$\begin{aligned}
 b(q_1 - q_2) &= 0 \\
 \rightarrow q_1 &= q_2
 \end{aligned}$$

since $b \neq 0$.

This is my (baby step) attempt to prove the fundamental theorem of arithmetic that states that every integer is a unique product of primes, we start with a product of two primes.

The product of primes $p_1 p_2$ will be equal to the product of other primes p_3, p_4 , *i.e.* $p_1 p_2 = p_3 p_4$, if and only if p_1, p_2 and p_3, p_4 are the same, *i.e.* $p_1 = p_3$, $p_2 = p_4$ or $p_1 = p_4$, $p_2 = p_3$.

I tried different things and I only got circular arguments :(this is the first circular argument.

Proof? I'm not sure about this proof but here it goes. Assume the opposite, $p_1 p_2 = p_3 p_4$ but $\{p_1, p_2\} \neq \{p_3, p_4\}$.

First observation, say $p_1 > p_3$, this means that $p_4 > p_2$ such that the equality has a chance to hold, this also means that

$$p_1 = qp_3 + r$$

$$p_4 = \tilde{q}p_2 + \tilde{r}$$

for some $q, r, \tilde{q}, \tilde{r} \in \mathbb{Z}$

$$p_1 p_2 = p_2 (qp_3 + r)$$

$$p_3 p_4 = p_3 (\tilde{q}p_2 + \tilde{r})$$

$$qp_2 p_3 + p_2 r = \tilde{q}p_2 p_3 + p_3 \tilde{r}$$

$$p_2 p_3 (q - \tilde{q}) = p_3 \tilde{r} - p_2 r$$

$$\cancel{p_2 p_3} (q - \tilde{q}) = \cancel{p_2 p_3} \left(\frac{\tilde{r}}{p_2} - \frac{r}{p_3} \right)$$

since $r < p_3$, $\tilde{r} < p_2 \rightarrow \frac{r}{p_3} < 1$, $\frac{\tilde{r}}{p_2} < 1$. Now the minimum of $q - \tilde{q}$ is 1 while the maximum of $\frac{\tilde{r}}{p_2} - \frac{r}{p_3}$ is < 1 . The only way this can be consistent is if $\frac{\tilde{r}}{p_2} - \frac{r}{p_3} = q - \tilde{q} = 0$.

But this means

$$\begin{aligned} \frac{\tilde{r}}{p_2} &= \frac{r}{p_3} \\ p_3 \tilde{r} &= p_2 r \end{aligned}$$

This means that p_2 divides $p_3\tilde{r}$, since p_3 is prime p_2 can only divide \tilde{r} but since $\tilde{r} < p_2$, p_2 doesn't divide \tilde{r} either so there's a contradiction.

The above statement is a circular argument, why? There's no theorem about $a|bc \rightarrow a|c$ if and only if a and b are primes, unless you already know the fundamental theorem of arithmetic. I tried proving it in various ways but it always comes down to a circular argument.

My first attempt was that $a|bc$ means that a is the product of the prime factors of b and c , now since a is prime, it is not a product of anything. We also have $a \neq b$, the only conclusion is that a is one of the prime factor of c . But this is only true if bc can be decomposed into a *unique* product of primes which is a circular argument (if the product is not unique bc might be cast into $bc = az$ and $a|bc$ is just a tautology, *i.e.* something like $0 = 0$, something that is obviously true).

Another argument I tried was using rational numbers

$$\begin{aligned} p_3\tilde{r} &= p_2r \\ \frac{p_3}{p_2} &= \frac{r}{\tilde{r}} \end{aligned}$$

but this is a contradiction as p_2, p_3 are primes and we cannot simplify the ratio p_3/p_2 , but since r/\tilde{r} has $r < p_3, \tilde{r} < p_2$ we have a "smaller" ratio which in turn means that p_3/p_2 can be simplified, a contradiction indeed. However, the canonical form of rationals is a consequence of the fundamental theorem. We always know that we can simplify ratios because we know that every number is a unique product of primes.

I then saw that $p_1p_2 = p_3p_4$ is a common multiple of p_2 and p_3 . I then proceeded to prove that the least common multiple of p_2 and p_3 is p_2p_3 , the proof can be found below. However, a common multiple is not necessarily the least common multiple, so even though the least common multiple is the product p_2p_3 we might have other common multiples that are larger, *e.g.* $p_1p_2 = p_3p_4, p_1 > p_3, p_4 > p_2$.

+++++

Say we have two relatively prime numbers p_2 and p_3 , what is their least common multiple? Say their least common multiple is $mp_2 = p_3n$ since this is the **least** common

multiple m and n are the smallest (relatively prime) numbers that satisfy this equation, relatively prime such that we cannot cancel common factors and create an even lesser common multiple.

My goal here is to show that $m = p_3$, $n = p_2$.

Say that $m > p_3$, this means that $n > p_2$ such that the equality has a chance to hold, this also means that

$$m = qp_3 + r$$

$$n = \tilde{q}p_2 + \tilde{r}$$

for some $q, r, \tilde{q}, \tilde{r} \in \mathbb{Z}$

$$mp_2 = p_2(qp_3 + r)$$

$$p_3n = p_3(\tilde{q}p_2 + \tilde{r})$$

$$qp_2p_3 + p_2r = \tilde{q}p_2p_3 + p_3\tilde{r}$$

$$p_2p_3(q - \tilde{q}) = p_3\tilde{r} - p_2r$$

$$p_2p_3(q - \tilde{q}) = p_2p_3\left(\frac{\tilde{r}}{p_2} - \frac{r}{p_3}\right)$$

since $r < p_3$, $\tilde{r} < p_2 \rightarrow \frac{r}{p_3} < 1$, $\frac{\tilde{r}}{p_2} < 1$. Now the minimum nonzero value of $q - \tilde{q}$ is 1 while the maximum of $\frac{\tilde{r}}{p_2} - \frac{r}{p_3}$ is < 1 . The only way this can be consistent is if $\frac{\tilde{r}}{p_2} - \frac{r}{p_3} = q - \tilde{q} = 0$ such that

$$\begin{aligned}\frac{r}{p_3} &= \frac{\tilde{r}}{p_2} \\ rp_2 &= p_3\tilde{r}\end{aligned}$$

but this means that $N = rp_2 = p_3\tilde{r}$ is divisible by p_2 and p_3 , *i.e.* N is a common multiple of p_2, p_3 also since $r < m$, $\tilde{r} < n \rightarrow N < mp_2$, $N < np_3$ this is a contradiction since there shouldn't be a lesser common multiple than the least common multiple!

Next option is $m < p_3$, this means that $n < p_2$ such that the equality has a chance to hold, this also means that

$$p_3 = qm + r$$

$$p_2 = \tilde{q}n + \tilde{r}$$

for some $q, r, \tilde{q}, \tilde{r} \in \mathbb{Z}$

$$\begin{aligned} mp_2 &= m(\tilde{q}n + \tilde{r}) \\ p_3n &= n(qm + r) \\ \cancel{mn} \left(\tilde{q} + \frac{\tilde{r}}{n} \right) &= \cancel{nm} \left(q + \frac{r}{m} \right) \\ \tilde{q} - q &= \frac{r}{m} - \frac{\tilde{r}}{n} \end{aligned}$$

since $r < m$, $\tilde{r} < n \rightarrow \frac{r}{m} < 1$, $\frac{\tilde{r}}{n} < 1$. Now the minimum nonzero value of $q - \tilde{q}$ is 1 while the maximum of $\frac{r}{m} - \frac{\tilde{r}}{n}$ is < 1 . The only way this can be consistent is if $\frac{r}{m} - \frac{\tilde{r}}{n} = \tilde{q} - q = 0$. But this means

$$\begin{aligned} \frac{\tilde{r}}{n} &= \frac{r}{m} \\ m\tilde{r} &= rn \\ \rightarrow \frac{\cancel{m}\tilde{r}}{\cancel{m}p_2} &= \frac{r\cancel{m}}{p_3\cancel{m}} \\ rp_2 &= p_3\tilde{r} \end{aligned}$$

since $r < m$, $\tilde{r} < n$, this is again a contradiction as we claim that m, n are the smallest numbers that satisfy $mp_2 = p_3n$. This means that $m < p_3$ is wrong but we have shown that $m > p_3$ is also wrong, the only conclusion is that $m = p_3$ and therefore $n = p_2$.

+++++

Algorithm to find least common multiple, an accidental byproduct of the above mishap.

You want to find the least common multiple of m and n , say $m > n$ (and of course they are co prime).

$$\begin{aligned} m &= nq + r \\ am &= nb \\ n(aq) + ar &= nb \\ ar &= n(b - aq) \end{aligned}$$

Now we have the common multiple of r (the remainder of m) and n . So the question of finding the least common multiple of m and n becomes the question of finding the

common multiple of r and n . This is reminiscent of the gcd algorithm of Euclid. The next step is repeating the process, since $r < n$, $n = rq' + r'$ and we now search for the common multiple r and r' and so on.

+++++

My version of the proof of the fundamental theorem of arithmetic.

Instead of using gcd (greatest common divisor) I'll use lcm (least common multiple). First, I need to show that every common multiple of two numbers is itself a multiple of the least common multiple.

Proof. Assume otherwise, denote the least common multiple of two numbers a and b is denoted L and another common multiple M (larger than L obviously since L is the *least* common multiple) we assume M is not a multiple of L and $M > L$. Now we do some arithmetic :)

$$M - L = R_0$$

we know that because $a, b | M$ and $a, b | L \longrightarrow a, b | R_0$, which means that R_0 is also a common multiple of a, b . But since L is the *least* common multiple of a and b , $R_0 > L$. We can further deduce that $R_0 > 0$ since $M > L$. We also know that $L \nmid R_0$ because if R_0 is a multiple of L then M must be a multiple of L too.

But this also means that we can repeat the process

$$R_0 - L = R_1$$

with the same conclusion for R_1 , it is a common multiple of a, b which is not a multiple of L and $R_1 > L$. We can then repeat it, and repeat it infinitely many times, which becomes "*reductio ad absurdum*" since we start with two *finite* numbers M and L . Thus any common multiple of two numbers must be a multiple of their least common multiples.

The above proof can also be done using a more orthodox number theory stuff, *i.e.* the proposition that for $a, b \in \mathbb{Z}$, $a > b$ there exist unique q, r such that $a = bq + r$ where $0 \leq r < b$ and $q > 0$. Therefore

$$M = Lq + r$$

but because $a, b | M$ and $a, b | L \longrightarrow a, b | r$ and since $0 \leq r < L$, r is a common multiple of a and b that is smaller than the least common multiple, which is a contradiction.

Now, if a and b were primes it can be shown that $L = ab$. A word of caution, you can't say that since a, b are primes the only factors of ab are $1, a, b$ therefore their lcm must be ab . This is circular reasoning. We want to show that any number can be factorized into a unique set of primes, however, we haven't shown that that is the case yet.

To show that L is ab we need to use the above result. We know that since ab is a common multiple of a and b it has to be a multiple of L

$$ab = Lq$$

but $L = an$ since $a, b | L$ and so

$$ab = anq$$

$$b = nq$$

since b is prime n, q must be $1, b$, the question is which one is which? n can't be one since then $L = a$ which is a contradiction, therefore $q = 1, n = b$ and thus $L = ab$.

We have therefore shown that the least common multiple of two primes is the product of those two. This combined with the fact that any common multiple is a multiple of the least common multiple are the only ingredients we need to prove the fundamental theorem of arithmetic.

Now, to the fundamental theorem. Suppose we can factor a number N into two different factorization

$$p_1 p_2 \dots p_n = q_1 q_2 \dots q_m$$

where p_i, q_j are all primes. We will show that $m = n$ (the same number of factors for both sides) and that the p 's are equal to the q 's.

First, $N = p_1 p_2 \dots p_n = q_1 q_2 \dots q_m$ means that N is a common multiple of p_1 and q_1 and since they are both primes N must be a multiple of $p_1 q_1$ therefore

$$N = q_1 R_{p_1} = p_1 p_2 \dots p_n$$

$$q_1 R_{p_1} = p_2 \dots p_n$$

where R_{p_1} is just some integer. But the above result means that $N_{p_1} = q_1 R_{p_1} = p_2 \dots p_n$ is a common multiple of q_1 and p_2 and since they are both primes N_{p_1} must be a multiple of $p_2 q_1$ therefore

$$\begin{aligned} N_{p_1} &= p_2 q_1 R_{p_2} = p_2 p_3 \dots p_n \\ q_1 R_{p_2} &= p_3 \dots p_n \end{aligned}$$

We can then repeat the process until $q_1 R_{p_{n-1}} = p_n$, since both q_1 and p_n are prime the only solution is to have $R_{p_{n-1}} = 1$ and $q_1 = p_n$, we can therefore factor them out from the equation

$$\begin{aligned} p_1 p_2 \dots p_{n-1} &= q_1 q_2 \dots q_m \\ \rightarrow p_1 p_2 \dots p_{n-1} &= q_2 q_3 \dots q_m \end{aligned}$$

We can now repeat the procedure above to get $q_2 = p_{n-1}$ and so on.

Now if $m \neq n$ we will be in trouble sooner or later because one side will be left with just 1

$$\begin{aligned} p_1 p_2 \dots p_k &= 1 \quad \text{if } n > m \\ 1 &= q_j q_{j+1} \dots q_n \quad \text{if } n < m \end{aligned}$$

which is again a contradiction since p 's and q 's are all prime numbers. The only way out is to have $m = n$ (the number of prime factors are the same) but that means p 's and q 's constitute the same set of prime numbers. We therefore have the result that the prime factors of any number is unique which is the fundamental theorem of arithmetic. As a bonus, our procedure above also demonstrates that the prime factors is unique up to an ordering since we can choose any p and q to begin with. $\mathcal{Q.E.D}$

+++++

Alternative (or more aptly generalized) proof that there is an infinite number of primes. Euclid's proof of $p_1 \dots p_n + 1$ can be generalized further, resulting in ways to produce primes from known ones.

$S = p_1 + p_2 \dots p_n$

where $p_1 \dots p_n$ are all the known primes we have. This means that S must contain a new prime as its factor. Otherwise, if p_1 is a factor of S

$$\begin{aligned} S - p_1 &= p_2 \dots p_n, & S &= p_1 S' \\ p_1(S' - 1) &= p_2 \dots p_n \\ \rightarrow p_1 &| p_2 \dots p_n \end{aligned}$$

$$\begin{array}{rcl} S - p_2 \dots p_n & = & p_1, \quad S = p_{13}S' \\ p_{13}(S' - p_2 \dots p_{12}p_{14} \dots p_n) & = & p_1 \\ \rightarrow p_{13} & | & p_1 \end{array}$$
$$S = \pm \prod_{i=1}^l p_i^{\alpha_i} \mp \prod_{j=l+1}^n p_j^{\alpha_j}$$

OR

$$S = \pm 1 \mp \prod_{i=1}^n p_i^{\alpha_i}$$
[illegible]

Say $8a^2+1 = d^2$ has non-trivial integer solutions and indeed it does, is there a formula to generate these solutions? Apparently there is, but it is a recurrence relation, proposition: say we have two initial solutions

$$8a_0^2 + 1 = d_0^2$$

$$8a_1^2 + 1 = d_1^2$$

where

$$a_0 = 0 \quad d_0 = 1$$

$$a_1 = 1 \quad d_1 = 3$$

It might seem a bit curious that the very first solution is a trivial one but it has to be what it has to be :) The subsequent solutions are given by

$$a_i = 6a_{i-1} - a_{i-2}$$

$$d_i = 6d_{i-1} - d_{i-2}$$

If we substitute this back into $8a^2 + 1 = d^2$ we get

$$8a_i^2 + 1 = d_i^2$$

$$8(6a_{i-1} - a_{i-2})^2 + 1 = (6d_{i-1} - d_{i-2})^2$$

$$36(8a_{i-1}^2) - 12 \cdot 8a_{i-1}a_{i-2} + (8a_{i-2}^2 + 1) = 36d_{i-1}^2 - 12d_{i-1}d_{i-2} + d_{i-2}^2$$

$$36(8a_{i-1}^2) - 12 \cdot 8a_{i-1}a_{i-2} + \cancel{d_{i-2}^2} = 36(8a_{i-1}^2 + 1) - 12d_{i-1}d_{i-2} + \cancel{d_{i-2}^2}$$

$$\cancel{36(8a_{i-1}^2)} - 12 \cdot 8a_{i-1}a_{i-2} = \cancel{36(8a_{i-1}^2)} + 36 - 12d_{i-1}d_{i-2}$$

$$8a_{i-1}a_{i-2} = d_{i-1}d_{i-2} - 3$$

The above relation will prove to be a crucial one :)

To prove this we can use induction, say the above relations are true up till $8a_i^2 + 1 = d_i^2$, *i.e.* $a_i = 6a_{i-1} - a_{i-2}$, $d_i = 6d_{i-1} - d_{i-2}$, let's see what happens if we apply the same rule

to $a_{i+1} = 6a_i - a_{i-1}$, in which case

$$\begin{aligned}
8a_i a_{i-1} &= 8(6a_{i-1} - a_{i-2})a_{i-1} \\
&= 6(8a_{i-1}^2) - 8a_{i-1}a_{i-2} \\
&= 6(d_{i-1}^2 - 1) - d_{i-1}d_{i-2} + 3 \\
&= (6d_{i-1} - d_{i-2})d_{i-1} - 3 \\
8a_i a_{i-1} &= d_i d_{i-1} - 3
\end{aligned}$$

Thus we if back substitute this to our expansion above we have

$$\begin{aligned}
8a_{i+1}^2 + 1 &= 8(6a_i - a_{i-1})^2 + 1 \\
&= 36(8a_i^2) - 12 \cdot 8a_i a_{i-1} + (8a_{i-1}^2 + 1) \\
&= 36(d_i^2 - 1) - 12(d_i d_{i-1} - 3) + d_{i-1}^2 \\
&= (6d_i)^2 - 12d_i d_{i-1} + d_{i-1}^2 \\
8a_{i+1}^2 + 1 &= (6d_i - d_{i-1})^2
\end{aligned}$$

and it works :)

Note that for $8a + 1 = d^2$ the answer is easy because the square of any odd number is

$$\begin{aligned}
(2\gamma + 1)^2 &= 4(\gamma(\gamma + 1)) + 1 \\
&= 8 \left(\frac{\gamma(\gamma + 1)}{2} \right) + 1
\end{aligned}$$

and since $\gamma(\gamma + 1)$ is always even we can always extract a factor of two out of it, thus for $8a + 1 = d^2$, d is any odd number.

The reason I wanted to solve for $8a^2 + 1 = d^2$ is because I wanted to show whether the following Pythagorean equation has a solution

$$x^2 + (x + 1)^2 = (x + k)^2$$

where k is a free parameter, *i.e.* the two terms on the LHS only differ by one, using the quadratic formula we get

$$x = k - 1 \pm \sqrt{2(k^2 - k)}$$

this means that $k^2 - k = 2a^2$, solving k using the quadratic formula we get

$$k = \frac{1 \pm \sqrt{8a^2 + 1}}{2}$$

which in turn means that $8a^2 + 1 = d^2$ for some odd integer d .

Note that

$$x^2 + (x + 2)^2 = y^2$$

reduces to $x^2 + (x + 1)^2 = (x + k)^2$. We can see this by doing mod 4. In mod 4, any square $a^2 \pmod{4}$ is either 0 or 1. If $x = 1 \pmod{4}$ we get

$$1^2 + (1 + 2)^2 = 2 \pmod{4}$$

and it is certainly not a quadratic residue modulo 4, If $x = 3 \pmod{4}$ we get

$$3^2 + (3 + 2)^2 = 2 \pmod{4}$$

again, not a quadratic residue modulo 4, while $x = 0 \pmod{4}$

$$(4w)^2 + (4w + 2)^2 = 4(2w)^2 + 4(2w + 1)^2$$

and $x = 2 \pmod{4}$

$$(4w + 2)^2 + (4w + 4)^2 = 4(2w + 1)^2 + 4(2w + 2)^2$$

we can of course cancel the common factor of 4 from LHS and RHS.

We can take this further by asking if the following has integer solutions

$$8a^3 + 1 = d^2$$

or more generally since $8 = 2^3$

$$x^3 + 1 = d^2$$

Trial by fire, round one

This is somewhat interesting as anything with cubes is usually hard :) My line of attack (which came to me while walking home from buying yogurt) is to rearrange

$$\begin{aligned} x^3 &= d^2 - 1 = (d - 1)(d + 1) \\ &= m(m + 2) \end{aligned}$$

where $m = d - 1$, I then break x down into its primal constituents

$$x = \prod_i p_i^{n_i}$$

The simplest case is when x only has one kind of primal factor $x = p_0^{n_0}$, one important thing to note is that the common factor between m and $m + 2$ is either 1 or 2, there can't be anything else otherwise say $g|m$, $g|(m + 2)$, $g > 2$

$$g|(m + 2) - m$$

$$g|2$$

but $g > 2$ so it just isn't so. Therefore, when we distribute the prime factors of x into m and $m + 2$ we can only do it if $p_0 = 2$, well, we can set $m = 1$, $m + 2 = p_0^{n_0}$ but then $(m + 2) - m \neq 2$, even when $p_0 = 2$ we can only set $m = 2$ because if we set

$$m = 2^i \quad m + 2 = 2^{n_0-i}$$

then $\gcd(m, m + 2) > 2$ which is a violation. But if $m = 2$ then $m + 2 = 4$ and $x^2 = m(m + 2) = 8$ and $x = 2$, so if x only contains a single prime type as a factor it has to be $x = 2$.

Trial by fire, round two

Next, what happens if x has two prime types, *i.e.* $x = p_1^{n_1} p_2^{n_2}$? Well, the requirement that $\gcd(m, m + 2) \leq 2$ forces either m to take all of $p_1^{n_1}$ and $m + 2$ all of $p_2^{n_2}$ or vice versa. But this way there's no way m and $m + 2$ to differ by just 2 because

$$p_1^{3n_1} p_2^{3n_2} = m(m + 2)$$

$$\rightarrow m = p_1^{3n_1}, \quad m(m + 2) = p_2^{3n_2}$$

and the difference between any two cubes is

$$(x + k)^3 - x^3 = 3x^2k + 3xk^2 + k^3$$

where the minimum is 7 (when $x = k = 1$), so no bueno. This can be avoided only if one of those two primes is 2

$$x = 2^{n_1} p_2^{n_2}$$

$$\rightarrow x^3 = 2^{3n_1} p_2^{3n_2} = m(m + 2)$$

but this case is immediately a no go as $m = 2$, $m + 2 = 2^{3n_1-1}p_2^{3n_2}$ as dictated by the fact that $\gcd(m, m + 2) = 2$ and $m < m + 2$.

Trial by fire, round three

The next easiest thing is three prime factors, $x = p_1p_2p_3$ such that

$$x^3 = p_1^3p_2^3p_3^3 = m(m + 2)$$

again in this case one of those primes must be 2 to avoid $\gcd(m, m + 2) > 2$

$$2^3p_2^3p_3^3 = m(m + 2)$$

and to avoid having perfect cubes for m and $m + 2$, depending on the actual values of p_2 and p_3 we have four possibilities

$$\begin{aligned} m &= 2p_2^3 & m + 2 &= 4p_3^3 \\ m &= 4p_2^3 & m + 2 &= 2p_3^3 \\ m &= 2p_3^3 & m + 2 &= 4p_2^3 \\ m &= 4p_3^3 & m + 2 &= 2p_2^3 \end{aligned}$$

to simplify things let's denote the prime number that multiplies 2 as B and the prime factor that multiplies 4 as A , *e.g.* $m = 2p_2^3 = 2B^3$ but for $m = 4p_2^3 = 4A^3$, etc, therefore

$$\begin{aligned} (m + 2) - m &= 2 = \mp 2(2A^3 - B^3) \\ \pm 1 &= 2A^3 - B^3 \end{aligned}$$

$$B^3 \pm 1 = (B \pm 1)(B^2 \mp B + 1) = 2A^3$$

the \mp sign on the first line is to cater for the two choices of the constant coefficient for m and $m + 2$. Now for $B \geq 3$, $B^2 \mp B + 1 > B \pm 1$. Obviously $B + 1 = 2$, $B^2 - B + 1 = A^3$ can't be so since then $B = p_{2,3} = 1$, therefore

$$B \pm 1 = 2A, \quad B^2 \mp B + 1 = A^2$$

but take note that

$$\begin{aligned} (B \pm 1)^2 &= B^2 \pm 2B + 1 = 4A^2 \\ \rightarrow (B \pm 1)^2 - (B^2 \mp B + 1) &= \pm(4A^2 - A^2) \\ \pm 3B &= \pm 3A^2 \\ B &= A^2 \end{aligned}$$

but remembering that $B \pm 1 = 2A$ we have

$$\begin{aligned} B \pm 1 &= 2A \\ A^2 \pm 1 &= 2A \\ \pm 1 &= A(2 - A) \\ \rightarrow A &= p_{2,3} \mid \pm 1 \end{aligned}$$

which in turn is a contradiction, thus $x = 2p_2p_3$ does not work. The next difficulty up is $x = 2^{n_1}p_2^{n_2}p_3^{n_3}$, same trick

$$2^{3n_1}p_2^{3n_2}p_3^{3n_3} = m(m+2)$$

again to simplify things, denote $A = 2^{3n_1-3}p_{2,3}^{3n_{2,3}}$, $B = p_{2,3}^{3n_{2,3}}$, therefore

$$\begin{aligned} \frac{(m+2) - m}{2} &= \pm 1 = 2A^3 - B^3 \\ B^3 \pm 1 &= 2A^3 \\ (B \pm 1)(B^2 \mp B + 1) &= 2A^3 \end{aligned}$$

Before moving on, let's find out what the $\gcd(B \pm 1, B^2 \mp B + 1)$ is. Say the gcd is d this means that

$$\begin{aligned} d &\mid (B \pm 1)^2 - (B^2 \mp B + 1) \\ d &\mid \pm 3B \end{aligned}$$

which in turn means that either d is 3 or it is 1 because if it's not 3 then $d \mid B \rightarrow B = dB'$, let's denote $d\Lambda = (B \pm 1)$

$$\begin{aligned} (B \pm 1)^2 &= B^2 \pm 2B + 1 \\ d\Lambda &= d(dB'^2 \pm 2B') + 1 \\ d(\Lambda - (B'^2 \pm 2B')) &= 1 \\ d &\mid 1 \end{aligned}$$

Well, that's one interesting fact, but even with that fact it's still a mess to continue, especially if we expand into a generic territory $x = 2^{n_1}p_2^{n_2}p_3^{n_3} \dots p_i^{n_i}$, we need a new way in.

Trial by fire no more

One possible way in is to note that since $\gcd(m, m+2) = 2$ at the very most, for a generic

$$x = 2^{n_1} \prod_i p_i^{n_i}, \quad p_i > 2$$

we need to distribute the prime factors of x^3 into m and $(m+2)$ in such a way that one of them have exactly one factor of two and the other absorbs the rest, if $\gcd(m, m+2) = 1$ then $m = C^3$ and $m+2 = D^3$ but the difference between two cubes is at least 7 as shown above, thus only $\gcd(m, m+2) = 2$ will work, so it is in a way very similar to the approach above.

Now, say $(m+2) = 2^{3n_1-1}D^3$ and $c = 2C^3$ and

$$\begin{aligned}(m+2) - m &= 2 = 2^{3n_1-1}D^3 - 2C^3 \\ 1 &= 2(2^{3n_1-3}D^3) - C^3 \\ 1 &= 2E^3 - C^3\end{aligned}$$

where $(2^{3n_1-1}D^3)(2C^3) = x^3$, we might be tempted to say that the difference between two cubes is at least 7, but in this case it might be the case that $E^3 < C^3$ and that argument falls through. If the assignment is the other way round $(m+2) = 2C^3$ and $2^{3n_1-1}D^3$ we get

$$\begin{aligned}(m+2) - m &= 2 = 2C^3 - 2^{3n_1-1}D^3 \\ 1 &= C^3 - 2E^3\end{aligned}$$

or in other words $\frac{(m+2)-m}{2}$ is

$$\pm 1 = 2E^3 - C^3$$

Our problem of finding integral solutions for $1 + x^3 = y^2$ has now been reduced to finding integral solutions of $\pm 1 = 2E^3 - C^3$

My first attempt was of course to parametrize E and C and then deploying our awesome cubic formula but it turned out to be quite disastrous though there were some silver linings which we will discuss later.

My next idea was to utilize the fact that

$$0 = \pm 1 - 2E^3 + C^3 = a^3 + b^3 + c^3 + d^3$$

with $b = c = -E$, the equation $0 = a^3 + b^3 + c^3 + d^3$ is called the cubic quadruple. There's a famous parametrization of this equation due to Euler

$$a = (1 - (p - 3q)(p^2 + 3q^2))r$$

$$b = ((p + 3q)(p^2 + 3q^2) - 1)r$$

$$c = ((p^2 + 3q^2)^2 - (p + 3q))r$$

$$d = (-(p^2 + 3q^2)^2 + (p - 3q))r$$

however, this only works for *rational* parameters p, q, r , there's no known integral parametrization for the cubic quadruples that cover all integral solutions. Note that even though p, q, r are rational numbers they still cover all integral solutions a, b, c, d (and rational solutions as well) of the cubic quadruples (this is something that confused me before because it's usually mentioned that Euler came up with the complete rational parametrization of cubic quadruples but there is no known integral parametrization, aren't integers a subset of rationals? The point is that there are two things here, the solutions which can be integers or rationals and the parameters which again can be integers or rationals. Rational parameters can cover all of integral and rational solutions but integral parameters can only cover some integral solutions, no integral parameters can cover all integral solutions).

From my experience with the cubic quadruples I never see a solution where two of the numbers are the same, *i.e.* all solutions are of the form $a \neq b \neq c \neq d$, but again an observation is not a proof.

Armed with Euler's parametrization above I set out to proof that no two numbers in the cubic quadruple can be the same but it presents more difficulties. First difficulty is that p, q, r are rationals and we are dealing with integers therefore we need to expand p, q into ratios (r can be canceled out).

The next difficulty is that working with p, q, r parametrization means that we are now dealing with quartic Diophantine equations, something that is more difficult than our original problem.

Fermat to the rescue

I then thought that maybe the proof for FLT with $n = 3$, *i.e.* $a^3 + b^3 + c^3 = 0$, can be used to proof that $a^3 + b^3 + 2c^3 = 0$ has no nontrivial integer solutions.

We will now follow the proof for $a^3 + b^3 + c^3 = 0$ closely (which is FLT with $n = 3$)

First, if (a, b, c) are not pairwise coprime we can always divide the three of them by their common gcd. If both a and b are even then

$$(2a')^3 + (2b')^3 + 2c^3 = 0$$

$$8(a'^3 + b'^3) + 2c^3 = 0$$

$$4(a'^3 + b'^3) + c^3 = 0$$

therefore $4|c^3$ which means that c has to be even as well. In this case we can keep dividing (a, b, c) by two until both a and b are odd. We can't have one of them to be even and the other odd because $a^3 + b^3$ will then be odd which is a contradiction since $a^3 + b^3$ must be divisible by 2 by the virtue of the factor of 2 multiplying c^3 .

Thus we can just consider coprime (a, b, c) with a and b both odd. In this case

$$a + b = 2u$$

$$a - b = 2v$$

$$\rightarrow a = u + v$$

$$b = u - v$$

because the sum/difference of two odd numbers must be even. Also, u and v must have opposite parities (one is odd while the other is even) because otherwise their sum/difference is even and both a and b will be even and thus not coprime, also $\gcd(u, v) = 1$ because otherwise if $d = \gcd(u, v) > 1$

$$a = d(u' + v')$$

$$b = d(u' - v')$$

which means that $d|a, d|b$ and $\gcd(a, b) > 1$.

Continuing

$$-2c^3 = (u + v)^3 + (u - v)^3$$

$$-2c^3 = 2u(u^2 + 3v^2)$$

Since u and v have opposite parities $(u^2 + 3v^2)$ must be odd. Thus the parity of c is determined by the parity of u .

Using the fact that $\gcd(u, v) = 1$ we would like to know $\gcd(u, u^2 + 3v^2)$, say $d|u$ and $d > 1$, well d must have prime factors, let's denote a prime divisor of d as g , and let us assume $g > 3$ for now, then $g|u \rightarrow g|u^2$ and $g|(u^2 + 3v^2)$, therefore

$$\begin{aligned} g|(u^2 + 3v^2) - u^2 \\ g|3v^2 \end{aligned}$$

Now since g is prime and it's bigger than 3 therefore $g|3v^2 \rightarrow g|v^2 \rightarrow g|v$ but this is a contradiction since $g > 3$, $g|u, g|v$ and $\gcd(u, v) = 1$. So $1 \leq g \leq 3$, but following the same steps as above $g|u$ and $g|3v^2 \rightarrow g|v^2 \rightarrow g|v$. However, either u or v must be odd so their common divisor cannot be even, thus $d = 3^{n_2}$, $n_2 > 1$, which in turn means

$$\begin{aligned} 3^{n_2}|u^2 &\rightarrow 3^{n_2-1}|u^2 \\ 3^{n_2}|3v^2 \\ &\rightarrow 3^{n_2-1}|v^2 \end{aligned}$$

In this case $\gcd(u, v) \geq 3^{\lceil (n_2-1)/2 \rceil}$ which is again a contradiction since u and v are coprime. Therefore if $d > 1$ then $d = 3$. This means that $\gcd(u, u^2 + 3v^2)$ is either 1 or 3.

First Case

If $\gcd(u, u^2 + 3v^2) = 1$ then due to the fact that $-c^3 = u(u^2 + 3v^2)$

$$\begin{aligned} u &= r^3 \\ u^2 + 3v^2 &= s^3 \end{aligned}$$

Since $(u^2 + 3v^2)$ is odd, so is s . A crucial lemma shows that if s is odd and if it satisfies $s^3 = u^2 + 3v^2$ then there are two coprime integers e and f

$$s = e^2 + 3f^2$$

so that

$$\begin{aligned} u &= e(e^2 - 9f^2) \\ v &= 3f(e^2 - f^2) \end{aligned}$$

e and f must also have opposite parities because u and v have opposite parities and

$$u = e(e^2 - 9f^2) = e(e - 3f)(e + 3f)$$

Next just like the argument above say $\gcd(e, e \pm 3f) = d > 1$, take a prime divisor g of d , and

$$\begin{aligned} g|e \\ g|e \pm 3f \\ \rightarrow g|\pm 3f \end{aligned}$$

Since g is prime, if $g > 3$ then $g|f$ and since $g|e$, $g = \gcd(e, f) > 1$ which is a contradiction since e and f are coprime. So $1 \leq g \leq 3$, now $g \neq 3$ because then $g = 3|e \rightarrow 3|u$ but $3|v$ as $v = 3f(e^2 - f^2)$ but u and v are coprime, so $g = 1, 2$ but 2 is out of the question since then $g = 2|e \rightarrow 2|u$ and $2|3f \rightarrow 2|f \rightarrow 2|v$ but $\gcd(u, v) = 1$, so $d = g = 1 = \gcd(e, e \pm 3f)$.

Now we need to take care of $\gcd(e - 3f, e + 3f) = d$, in this case since $d|e \pm 3f$

$$\begin{aligned} d|((e - 3f) + (e + 3f)) &\rightarrow d|2e \\ d|((e + 3f) - (e - 3f)) &\rightarrow d|6f \end{aligned}$$

if we like previously choose a prime factor g of d then $g = 1, 2, 3$ but 3 is ruled out because then $g = 3|2e \rightarrow 3|e \rightarrow 3|u$ as $u = e(e^2 - 9f^2)$ and $3|v$ since $v = 3f(e^2 - f^2) \rightarrow \gcd(u, v) \geq 3$, contradicted by the fact that $\gcd(u, v) = 1$, d can't be 2 because e or f have opposite parities and therefore $e \pm 3f$ is always odd and since $d|(e \pm 3f)$, an odd number, $d \neq 2$, so the only choice is $d = 1$.

Thus $e, (e - 3f), (e + 3f)$ are pairwise coprime and that means each of them has to be a cube as they are divisors of r^3

$$\begin{aligned} r^3 = u &= e(e^2 - 9f^2) = e(e - 3f)(e + 3f) \\ &\rightarrow -e = k^3 \\ e - 3f &= l^3 \\ e + 3f &= m^3 \end{aligned}$$

Now note that

$$\begin{aligned} -2e + (e - 3f) + (e + 3f) &= 0 \\ \rightarrow 2k^3 + l^3 + m^3 &= 0 \end{aligned}$$

which is a smaller solution than $a^3 + b^3 + 2c^3 = 0$ since $k|r$ and $r|c$.

Second Case

If $\gcd(u, u^2 + 3v^2) = 3$ then we can express $u = 3w$ and since $\gcd(u, v) = 1$ then $3 \nmid v$

$$\begin{aligned} -c^3 &= 3w(9w^2 + 3v^2) \\ &= 9w(3w^2 + v^2) \end{aligned}$$

Because $\gcd(v, w) = 1$ and $3 \nmid v$ then $\gcd(9w, 3w^2 + v^2) = 1$ we can see this following a similar argument above, say $\gcd(9w, 3w^2 + v^2) = d$ and we pick a prime divisor g of d

$$\begin{aligned} g|9w &\rightarrow g|81w^2 \\ g|(27(3w^2 + v^2) - 81w^2) \\ g|3^3v^2 \end{aligned}$$

again, if $g > 3$ then since g is prime then $g|v^2 \rightarrow g|v$ and also $g|3^2w \rightarrow g|w$ but $\gcd(w, v) = 1$ so $1 \leq g \leq 3$ but if $g = 3$ then

$$\begin{aligned} g = 3 &\mid 3w^2 + v^2 \\ &\rightarrow 3b = 3w^2 + v^2 \\ 3(b - w^2) &= v^2 \\ 3 &\mid v^2 \rightarrow 3|v \end{aligned}$$

but $3 \nmid v$ thus $g = 1, 2$ but 2 is ruled out since u and v are of opposite parities, this means that if u is even then w is even then $9w$ is even and $3w^2 + v^2$ is odd, if w is odd and v is even then $9w$ is odd and $3w^2 + v^2$ is also odd, so no matter what $\gcd(9w, 3w^2 + v^2) \neq 2$ thus the only conclusion is that $9w$ and $3w^2 + v^2$ are coprime. In this case since $-c^3 = 9w(3w^2 + v^2)$

$$\begin{aligned} 9w &= r^3 \\ 3w^2 + v^2 &= s^3 \end{aligned}$$

By the same lemma, since s is odd and equal to $3w^2 + v^2$, it too can be expressed in terms of smaller coprime numbers, e and f , and therefore $s = e^2 + 3f^2$

$$\begin{aligned} v &= e(e^2 - 9f^2) \\ w &= 3f(e^2 - f^2) \end{aligned}$$

we can immediately see that e and f have opposite parities because otherwise both u and w will be both even, contradicting $\gcd(u, w) = 1$, substituting them back into r^3 we get

$$\begin{aligned} r^3 &= 9w = 27f(e^2 - f^2) = 3^3 f(e + f)(e - f) \\ \left(\frac{r}{3}\right)^3 &= f(e + f)(e - f) \end{aligned}$$

$(r/3)$ has to be an integer since r, e and f are all integers. And just like before $f, (e+f), (e-f)$ are all pairwise coprime. We can see this easily by first considering $\gcd(f, e \pm f) = d > 1$

$$\begin{aligned} e \pm f &= dm, \quad d|f \\ e &= d(m \mp f') \\ &\rightarrow d \mid e \end{aligned}$$

which is a contradiction since e and f are coprime. Now we want to know what $d = \gcd(e - f, e + f)$ is, again take a prime divisor g of d

$$\begin{aligned} g &| 2e \\ g &| 2f \end{aligned}$$

if $g > 3$ then since g is prime $g|e, g|f$ but it contradicts the fact that e and f are coprime. If $g = 3$ it will cause $3|2e \rightarrow 3|e \rightarrow 3|v$ by virtue of $v = e(e^2 - 9f^2)$ but since $w = 3f(e^2 - f^2)$, $3|w$ thus $\gcd(w, u) \geq 3$ which contradicts the fact that u and w are coprime, $d \neq 2$ because e and f are of opposite parities and $e \pm f$ is always odd, therefore their divisors cannot be 2, this means that $f, (e - f), (e + f)$ are pairwise coprime which means that

$$\begin{aligned} \left(\frac{r}{3}\right)^3 &= f(e + f)(e - f) \\ &\rightarrow -f = k^3 \\ e + f &= l^3 \\ -(e - f) &= m^3 \end{aligned}$$

but

$$\begin{aligned} 2f &= (e + f) - (e - f) \\ \rightarrow -2k^3 &= l^3 + m^3 \\ 0 &= l^3 + m^3 + 2k^3 \end{aligned}$$

which will lead to another infinite descent since $k|s$ and $s|c$.

In conclusion, any non-trivial integral solutions of $a^3 + b^3 + 2c^3 = 0$ will lead to an infinite descent, but wait a minute, does the above argument even indicate that the presupposed solutions used in the proof were nontrivial?

Well, the only thing we assumed was that a and b were both odd, so this already excluded the trivial solution $a = b = c = 0$. Two other trivial solutions are $(1, -1, 0)$ and $(\pm 1, \pm 1, \mp 1)$. For $(1, -1, 0)$ this means that $u = 0, v = 1$, recall that $u = (a + b)/2$ and $v = (a - b)/2$ while for $(\pm 1, \pm 1, \mp 1)$ it means $u = \pm 1, v = 0$.

Now, any number divides 0 and therefore $\gcd(n, 0) = n$ thus for the trivial cases above $\gcd(u, v) = \gcd(\pm 1, 0) = 1$, thus this falls under the first case where $\gcd(u, u^2 + 3v^2) = 1$.

Since $v = 3f(e^2 - f^2)$, to get $v = 0$ we need $f = 0$, why can't $e = f$? it's because e and f are coprime, and therefore we can't have $v = 1$ no matter what and so $(1, -1, 0)$ is excluded from this proof.

So $v = 0$ and $f = 0$, to get $\gcd(e, f) = \gcd(e, 0) = 1 \rightarrow e = \pm 1$ and $u = e(e - 3f)(e + 3f) = \pm 1$ which translates to $(\pm 1, \pm 1, \mp 1)$, thus if we start with $(\pm 1, \pm 1, \mp 1)$ we'll get back $(\pm 1, \pm 1, \mp 1)$.

Therefore, trivial solutions are either excluded or they don't make any difference in the proof.

The insight here is that if we want to utilize infinite descent we need to break down the equation into smaller compartments and then show that those also obey the same relationship as the original equation, we need to keep this in mind, just moving around variables and expanding/substituting them won't yield good fruits unless we are going the right direction.

Crucial Lemma is Crucial

Let's talk about the crucial lemma used in the proof above which is if we have an odd

$s \rightarrow s^3 = 3w^2 + v^2$ with $\gcd(w, v) = 1$ then $s = 3f^2 + e^2$ and $\gcd(e, f) = 1$.

First, why must s be odd? it's because any odd factor of $3w^2 + v^2$ must have the form $3a^2 + b^2$, intuitively we can see that

$$\begin{aligned}(3a^2 + b^2)(3c^2 + d^2) &= 3(3a^2c^2 + a^2d^2 + b^2c^2) + b^2d^2 \\ &= 3g^2 + h^2\end{aligned}$$

and if the constant coefficient is not 3 we will have some problem. The proof is quite involved so it won't be discussed here, the thing is that we can't just do substitution because we first have to prove the existence of e and f , *i.e.* that there exist two whole numbers e and f such that $(3f^2 + e^2)^3 = s^3$ and this is not easy.

As to why e and f have to be coprime it's because

$$\begin{aligned}u &= e(e^2 - 9f^2) \\ v &= 3f(e^2 - f^2)\end{aligned}$$

and u and v are coprime thus if e and f are not coprime neither are u and v .

Summary

Before we discuss my disastrous attempt utilizing the cunic formula, let's tie some loose ends and do a quick recap on how we proved that

$$x^3 + 1 = y^2$$

only has the following solutions $(x, y) = \{(-1, 0), (0, \pm 1), (2, \pm 3)\}$.

First, the above equation was actually an elliptic curve so a proof might be derived using an elliptical machinery.

Second, without the elliptical machinery, my attempt was the following

- Rewrite the equation as

$$\begin{aligned}x^3 &= y^2 - 1 \\ &= (y - 1)(y + 1) \\ &= m(m + 2)\end{aligned}$$

where $m = y - 1$

- Next, we know that $\gcd(m, m+2)$ is either 1 or 2
- We then write x in terms of its primal constituents
- From the fact that $\gcd(m, m+2)$ is restricted to 1 or 2 we deduce that either $x = 2^nb^m$ or the trivial solutions, *i.e.* $x = 2$ with $m = 2$ or $x = 0$ with $m = 0$ or $x = -1$ with $m = -1$
- Substituting $x = 2^nb^m$ we see that it leads to the equation $\pm 1 = 2E^3 - C^3$ where $m+2 = 4E^3$ and $m = 2C^3$
- Lucky for us $(\pm 1)^3 = \pm 1$ so we can rewrite the above as a cubic quadruple $0 = a^3 + b^3 + c^3 + d^3$ where one of them is ± 1 while the other two are the same, *i.e.* we want to prove that $0 = a^3 + b^3 + 2c^3$ only has trivial solutions
- Fine, so it has only trivial solutions, how do they affect our original problem of $y^2 = x^3 + 1$? The trivial solutions of $0 = a^3 + b^3 + 2c^3$ are $(a, b, c) = (0, 0, 0)$ which won't fit $\pm 1 = 2E^3 - C^3$ since one of (a, b, c) has to be ± 1 , next will be $(a, b, c) = (1, -1, 0)$ which in turn means that $E = 0$ and therefore $m+2 = 0 \rightarrow m = -2 \rightarrow y = -1$ which is the trivial solution we already know, and lastly $(a, b, c) = (\pm 1, \pm 1, \mp 1)$ in this case $C = \pm 1 \rightarrow m = \pm 2 \rightarrow y = 3$ or $y = -1$, where both are trivial solutions we already know

Mistakes and Pitfalls

For a moment, let's travel to the *realistic* universe, assume E, C are real numbers, we can then factor them out as follows

$$2E^3 - C^3 = (2^{1/3}E - C)((2^{1/3}E)^2 + (2^{1/3}E)C + C^2) = \pm 1$$

This means that $0 < |2^{1/3}E - C| < 1$, if $2^{1/3}E > C$ then $C = \lfloor 2^{1/3}E \rfloor$ and if $2^{1/3}E < C$ then $C = \lceil 2^{1/3}E \rceil$

Now if $E > C$ then we can use the argument that the difference of two cubes is at least

7 and we are through, this must mean $E < C$, so let's denote $C = E + F$ giving

$$\begin{aligned}\pm 1 &= 2E^3 - C^3 \\ &= 2E^3 - (E + F)^3 \\ \pm 1 &= E^3 - 3FE^2 - 3F^2E - F^3\end{aligned}$$

If we straight away use the cubic formula we get three solutions for E

$$\begin{aligned}E &= \left(\frac{\sqrt{4F^6 \pm 12F^3 + 1}}{2} + \frac{6F^3 \pm 1}{2} \right)^{1/3} + \frac{2F^2}{\left(\frac{\sqrt{4F^6 \pm 12F^3 + 1}}{2} + \frac{6F^3 \pm 1}{2} \right)^{1/3}} + F \\ E &= \left(-\frac{\sqrt{3}i}{2} - \frac{1}{2} \right) \left(\frac{\sqrt{4F^6 \pm 12F^3 + 1}}{2} + \frac{6F^3 \pm 1}{2} \right)^{1/3} + \frac{2 \left(\frac{\sqrt{3}i}{2} - \frac{1}{2} \right) F^2}{\left(\frac{\sqrt{4F^6 \pm 12F^3 + 1}}{2} + \frac{6F^3 \pm 1}{2} \right)^{1/3}} + F \\ E &= \left(\frac{\sqrt{3}i}{2} - \frac{1}{2} \right) \left(\frac{\sqrt{4F^6 \pm 12F^3 + 1}}{2} + \frac{6F^3 \pm 1}{2} \right)^{1/3} + \frac{2 \left(-\frac{\sqrt{3}i}{2} - \frac{1}{2} \right) F^2}{\left(\frac{\sqrt{4F^6 \pm 12F^3 + 1}}{2} + \frac{6F^3 \pm 1}{2} \right)^{1/3}} + F\end{aligned}$$

For E to be real, the imaginary parts have to cancel, *i.e.*

$$\begin{aligned}\left(\frac{\sqrt{3}i}{2} \right) \left(\frac{\sqrt{4F^6 \pm 12F^3 + 1}}{2} + \frac{6F^3 \pm 1}{2} \right)^{1/3} &= \frac{2 \left(\frac{\sqrt{3}i}{2} \right) F^2}{\left(\frac{\sqrt{4F^6 \pm 12F^3 + 1}}{2} + \frac{6F^3 \pm 1}{2} \right)^{1/3}} \\ \rightarrow \left(\frac{\sqrt{4F^6 \pm 12F^3 + 1}}{2} + \frac{6F^3 \pm 1}{2} \right)^{2/3} &= 2F^2 \\ \left(\sqrt{4F^6 \pm 12F^3 + 1} + 6F^3 \pm 1 \right)^2 &= 32F^6\end{aligned}$$

the above equality is valid for the two solutions for E that contain i . At this point we have to be careful, if we take the square root of both LHS and RHS we need to add the \pm sign on the RHS $\pm\sqrt{32}F^3$ but this is violation of the convention because we already have \pm sign on the LHS, the correct way is to write *two* equations

$$\begin{aligned}\left(\sqrt{4F^6 + 12F^3 + 1} + 6F^3 + 1 \right) &= \pm\sqrt{32}F^3 \\ \left(\sqrt{4F^6 - 12F^3 + 1} + 6F^3 - 1 \right) &= \pm\sqrt{32}F^3\end{aligned}$$

Solving for F for the above *four* equations we get integer solutions for F only if $F = 0$ which won't work for our original problem $\pm 1 = 2C^3 - (C + F)^3$. So there we have it,

the only solution for E that works is the one without any imaginary part, *i.e.*

$$E = \frac{M^2 + 2F^2 + FM}{M}$$

$$M = \left(\frac{\sqrt{4F^6 \pm 12F^3 + 1}}{2} + \frac{6F^3 \pm 1}{2} \right)^{1/3}$$

Now say that M is irrational, it doesn't mean that E must be irrational, E can still be rational because the ratio of two irrational numbers might still be rational, heck, even the sum of two irrational numbers can still be rational, *e.g.* $\pi - (\pi - 3) = 3$ and $\pi - 3$ is irrational.

At this point we need to recall how we got the cubic formula in the first place. A summary of the cubic formula is as follows, first we need to depress the cubic equation $ax^3 + bx^2 + cx + d = 0$ by removing the x^2 term by setting

$$x = y - \frac{b}{3a}$$

to get the depressed cubic equation

$$y^3 + \left(\frac{3ca - b^2}{3a^2} \right) y + \left(\frac{27da^2 + 2b^3 - 9abc}{27a^3} \right) = 0$$

What we have is now an equation of the form $y^3 + Ay = B$ which can be solved by the following substitutions

$$A = 3st$$

$$B = s^3 - t^3$$

yielding $y = s - t$ as the solution. But first we need to solve for s and t by substituting $s = (A/3t)$ into $B = s^3 - t^3$, *i.e.*

$$B = \left(\frac{A}{3t} \right)^3 - t^3$$

$$0 = t^6 + Bt^3 - \frac{A^3}{27}, \quad \text{setting } u = t^3$$

$$\rightarrow 0 = u^2 + Bu - \frac{A^3}{27}$$

and so we get a quadratic equation, this is how we get the square roots in the first place.

Once we solve for u we just take the *three* cube roots of u to get t and that's how we get

the complex coefficient since in the complex plane the n^{th} root of unity always have n solutions.

If the above is how we get the square root in the first place then to have rational solutions that square root must yield rational numbers, we can't have irrational or complex solution even at this intermediate point since u is a rational number, therefore in our case it is sufficient to show that whatever inside the square root does not yield perfect rational squares (here we deal with rational because we divide all coefficients by a in the beginning).

We now check if there's a rational solution for $M = \left(\frac{\sqrt{4F^6 \pm 12F^3 + 1}}{2} + \frac{6F^3 \pm 1}{2} \right)^{1/3}$, we know that $\frac{6F^3 \pm 1}{2}$ is rational and since F is an integer, the square root must produce an integer as well, why not rationals? It's because F is an integer, using the quadratic formula for whatever is inside the square root we solve for F

$$G^2 = 4F^6 \pm 12F^3 + 1$$

$$\rightarrow F = \left(\frac{\sqrt{G^2 + 8} \pm 3}{2} \right)^{1/3}$$

having F as an integer means that G is an integer and therefore $G^2 + 8$ is an integer as well. But there's no integer H such that $H^2 = G^2 + 8$ because

$$H^2 - G^2 = (H - G)(H + G) = 2 \cdot 2 \cdot 2$$

if $H, G > 0$ then $H = 3, G = 1$ is the only solution, if $H < 0, G > 0$

$$(-|H| - G)(-|H| + G) = (|H| + G)(|H| - G)$$

if $H, G < 0$

$$(-|H| - (-|G|))(-|H| + (-|G|)) = (|H| - |G|)(|H| + |G|)$$

if $H > 0, G < 0$

$$(H - (-|G|))(H + (-|G|)) = (H + |G|)(H - |G|)$$

Thus the only solutions are $(3, 1), (-3, 1), (3, -1), (-3, -1)$ in any case $\sqrt{G^2 + 8} = 3$ giving

$$F = \left(\frac{3 \pm 3}{2} \right)^{1/3}$$

$$F = \begin{cases} 3^{1/3} \\ 0 \end{cases}$$

and since F is an integer $F = 0$ but this violates $\pm 1 = 2C^3 - (C + F)^3$.

So it seems like we have solved it but in fact we haven't, the problem lies with the statement that " u is a rational number", it might not be, recall that $u = t^3$ and $B = s^3 - t^3$, $A = 3st$, $y = s - t$, it's true that B, A, y are all rational but that doesn't mean that s, t have to be rational as well. A product of two irrational numbers can still be rational, e.g. $\sqrt{18} \cdot \sqrt{2} = \sqrt{36} = 4$, and as shown above the difference of two irrational numbers can still be rational, thus the argument above fails.

Lessons learned

Using the cubic or quadratic formula to solve Diophantine equations is not a good idea, here's why, say we have a quadratic Diophantine

$$u^2 + wu + z^2 = 0$$

and we deploy our lovely friend the quadratic formula

$$u = \frac{-w \pm \sqrt{w^2 - 4z^4}}{2}$$

we now need to show that there is no integer solutions for $\sqrt{w^2 - 4z^4}$ but this is now a quartic Diophantine, even if we only have z instead of z^2 it is still a quadratic Diophantine and we are back to square one and the situation is even worse for a cubic Diophantine.

 =====

What we want to show here is that $x^4 + y^4 = z^2$ has no nontrivial coprime solution

Solution. We start by noting that (pairwise coprime) Pythagorean triples are always of the form $a = m^2 - n^2$, $b = 2mn$, $c = m^2 + n^2$ where $a^2 + b^2 = c^2$ and $\gcd(m, n) = 1$ and m and n have opposite parities, one of the is odd while the other is even, this is to ensure that a, b, c are pairwise coprime and that a and b have opposite parities as well, this is due to the fact that going modulo 4, if a and b are both odd then

$$\begin{aligned} a^2 + b^2 &= 1 + 1 \pmod{4} \\ &= 2 \pmod{4} \end{aligned}$$

but $2 \pmod{4}$ is not a quadratic residue modulo 4, so a and b must have opposite parities.

With the above in mind, we assign the following for our current problem $x^4 + y^4 = z^2$

$$x^2 = m^2 - n^2$$

$$y^2 = 2mn$$

$$z = m^2 + n^2$$

In this case x is odd and y is even but they're interchangeable due to the symmetry of the problem. Rearranging the equation for x we get $x^2 + n^2 = m^2$, we know that $\gcd(m, n) = 1$ therefore $\gcd(x, m) = \gcd(x, n) = 1$ as well, this means that (x, n, m) also form a Pythagorean triplet and since x is odd, n has to be even and m odd.

And from $y^2 = 2mn$, since m and n are coprime and n even as shown above we have $m = e^2$ and $n = 2f^2$ with $\gcd(e, f) = 1$.

Since (x, n, m) form a Pythagorean triplet they must obey

$$x = g^2 - h^2$$

$$n = 2gh$$

$$m = g^2 + h^2$$

but $n = 2f^2 = 2gh \rightarrow f^2 = gh$ and since g and h are coprime we have $g = r^2$, $h = s^2$, substituting all we have

$$m = g^2 + h^2$$

$$e^2 = r^4 + s^4$$

which leads to an infinite descent as $e < m < z$. And in this case we can't have nontrivial pairwise coprime solutions for $x^4 + y^4 = z^4 = (z^2)^2$ either, FTL $n = 4$ done :)

 =====

What we want to show here is that $x^4 - y^4 = z^2$ has no nontrivial coprime solution

Solution. My way of solving this problem is a bit odd :) First I factorize

$$z^2 = x^4 - y^4 = (x^2 - y^2)(x^2 + y^2)$$

We know that x^2, y^2, z form a Pythagorean triplet. Since z^2 is positive we know that $x > y$ and therefore x must be odd, y , however, can be either even or odd (x must be odd

because according to the Pythagorean triplet formula $x^2 = m^2 + n^2$ with m and n having opposite parities).

Because they are Pythagorean triplet, we only need to consider pairwise coprime solutions, *i.e.* $\gcd(x^2, y^2) = \gcd(x^2, z) = \gcd(y^2, z) = 1$. But this means that x, y, z must be coprime as well.

First if there's a common divisor d between $(x^2 - y^2)$ and $(x^2 + y^2)$ then

$$d|2x^2$$

$$d|2y^2$$

Thus $d = 1, 2$ since x and y are coprime, therefore the only common divisors between $(x^2 - y^2), (x^2 + y^2)$ are either 1 or 2 which is expected since if x and y have opposite parities then the common divisors between them must be 1 and if x and y are both odd then the common divisor between $(x^2 - y^2), (x^2 + y^2)$ is 2.

If the common divisor $(x^2 - y^2), (x^2 + y^2)$ is 1 then by virtue of $z^2 = (x^2 - y^2)(x^2 + y^2)$ we have

$$x^2 - y^2 = a^2$$

$$x^2 + y^2 = b^2$$

with $\gcd(a, b) = 1$ which means that a and b are odd. If the common divisor $(x^2 - y^2), (x^2 + y^2)$ is 2 then we have

$$x^2 - y^2 = 2c^2$$

$$x^2 + y^2 = 2d^2$$

with $\gcd(c, d) = 1$. We can rearrange the above “judiciously” to get

$$y^2 = d^2 - c^2$$

$$x^2 = d^2 + c^2$$

Note that c, d, x, y are all pairwise coprime, also since x and y are both odd, according to the Pythagorean triplet formula c has to be even and d has to be odd.

In either case we actually generate two other Pythagorean triplets $x^2 = y^2 + a^2$ and $x^2 + y^2 = b^2$ with four coprime variables (in the second case d plays the role of x of the

first case, and $c \leftrightarrow y, y \leftrightarrow a, x \leftrightarrow b$). So now we can just concentrate on the first case where x is odd, y is even, a is odd, and b is odd.

From $x^2 = y^2 + a^2$ we have

$$x = m^2 + n^2$$

$$y = 2mn$$

$$a = m^2 - n^2$$

while from $x^2 + y^2 = b^2$ we get

$$x = e^2 - f^2$$

$$y = 2ef$$

$$b = e^2 + f^2$$

Therefore

$$y = 2mn = 2ef$$

My next idea took me through a three day wild goose chase LOL From the fact that

$$\begin{aligned} x &= m^2 + n^2 = e^2 - f^2 \\ \rightarrow e^2 &= m^2 + n^2 + f^2 \end{aligned}$$

we get a Pythagorean quadruple. This also shows that $e > m > n > f$, $m > n$ because $x = m^2 - n^2$, $n > f$ because $mn = ef$ and $e > m$ therefore $f < n$. Also we cannot have two odd numbers (or more) among m, n, f , this is because 1,2,3 are not quadratic residue modulo 4, this means that f is odd and therefore f is even (since e and f have opposite parities).

Well, Pythagoras shouldn't have been involved at all, from $ef = mn$ we have

$$m = dm'$$

$$f = df'$$

$$n = f'n'$$

$$e = m'n'$$

which means that $\gcd(m, f) = d$ and $\gcd(e, n) = n'$, of course d or n' can be 1 but they cannot both be 1 at the same time because if they are then $e = m, f = n$ and

$$\begin{aligned}x &= e^2 - f^2 = m^2 + n^2 \\m^2 - n^2 &= m^2 + n^2 \\0 &= 2n^2 \\\rightarrow n &= 0\end{aligned}$$

which also means that $y = 0$ a trivial solution. Now for a nontrivial solution

$$\begin{aligned}x &= e^2 - f^2 = m^2 + n^2 \\e^2 - n^2 &= m^2 + f^2 \\n'^2(m'^2 - f'^2) &= (m^2 + f^2) \\n'^2(m'^2 - f'^2) \times (m^2 - f^2) &= (m^2 + f^2) \times (m^2 - f^2) \\n'^2(m'^2 - f'^2)(m'^2 - f'^2)d^2 &= (m^4 - f^4) \\n'^2(m'^2 - f'^2)^2d^2 &= (m^4 - f^4) \\\xi^2 &= m^4 - f^4\end{aligned}$$

where $\xi = n'(m'^2 - f'^2)d$, now since $x = m^2 - n^2$ and $y = 2ef$ it means that $m < x$ and $f < y$ which means that we have an infinite descent because we can repeat the whole argument from the top about $x^4 - y^4 = z^2$ by replacing $x, y, z \leftrightarrow m, f, \xi$.

In the above if $d = \gcd(m, f) > 1$ is true, we can just divide all terms m, f, ξ by d and the argument about infinite descent will still stand.

Pythagorean convolution

If we insist of deploying Pythagorean quadruples, there is a generating formula for the Pythagorean quadruples just like for the triplets, since m and n are of opposite parities (let's pick m to be odd here) and f is even, e and n are given by

$$\begin{aligned}e &= \frac{m^2 + f^2 + p^2}{2p} \\n &= \frac{m^2 + f^2 - p^2}{2p}\end{aligned}$$

which means that $e - n = p$, here the condition is that $p|m^2 + f^2$ and $p^2 < m^2 + f^2$.

If we now use the constraint that $ef = mn$

$$\begin{aligned} \left(\frac{m^2 + f^2 + p^2}{2p} \right) f &= m \left(\frac{m^2 + f^2 - p^2}{2p} \right) \\ m^2 f + f^3 + p^2 f &= m^3 + m f^2 - m p^2 \\ (m + f)p^2 &= (m^3 - f^3) - m f(m - f) \\ &= (m - f)(m^2 + m f + f^2 - m f) \\ (m + f)p^2 &= (m - f)(m^2 + f^2) \end{aligned}$$

This was where my wild goose chase began, from here on I was trying to produce all kinds of contradictions that weren't but those are for later, first the solution

$$\begin{aligned} (m + f) \times (m + f)p^2 &= (m + f) \times (m - f)(m^2 + f^2) \\ (m + f)^2 p^2 &= (m^2 - f^2)(m^2 + f^2) \\ \xi^2 &= m^4 - f^4 \end{aligned}$$

where $\xi = (m + f)p$, now since $x = m^2 - n^2$ and $y = 2ef$ it means that $m < x$ and $f < y$ which means that we have an infinite descent because we can repeat the whole argument from the top about $x^4 - y^4 = z^2$ by replacing $x, y, z \leftrightarrow m, f, \xi$.

In the above $d = \gcd(m, f) > 1$ might be true, if this is the case we can just divide all terms m, f, ξ by d and the argument about infinite descent will still stand.

Pythagorean restriction

Recruting Pythagoras is not so bad afterall, because the above shows that there's no Pythagorean quadruple $a^2 + b^2 + c^2 = d^2$ with $ad = bc$ where a is the smallest among a, b, c , a free theorem :)

Here's why this theorem is true, say we have the opposite, we have $e^2 = m^2 + n^2 + f^2$ and for now we don't care about the parities, now say that $ef = mn$ and f is the smallest one among e, m, n, f while e is the largest one.

If this is true then we can set

$$x = e^2 - f^2 = m^2 + n^2$$

$$y = 2mn = 2ef$$

$$a = m^2 - n^2$$

$$b = e^2 + f^2$$

such that

$$x^2 - y^2 = a^2$$

$$x^2 + y^2 = b^2$$

Now it doesn't matter even if m and n are not coprime (or e and f are not coprime or all of them are not coprime), we still have

$$(x^2 - y^2)(x^2 + y^2) = a^2 b^2$$

$$x^4 - y^4 = z^2$$

where $z = ab$ and if there is any common factor between the LHS and RHS we can get rid of it here, but this is a contradiction since the above solution has nontrivial solutions, *i.e.* the only trivial solutions are $(x, y, z) = (0, 0, 0)$, $(1, 1, 0)$ or $(1, 0, 1)$ and in any case it's either $y = 0$ or $z = 0$ and in this case we won't get a Pythagorean quadruple.

The errors of my way

Now let's go back to the goose I was chasing

$$(m + f)p^2 = (m - f)(m^2 + f^2)$$

at this point had m and f been coprime we would have been done because then $\gcd(m - f, m + f) = 1$ thus $m - f | p^2 \rightarrow r = \gcd(p, m - f) > 1$ but since $p | m^2 + f^2 \rightarrow r | m^2 + f^2$ and the fact that $r = \gcd(p, m - f)$ means $r | m - f \rightarrow r | m^2 - f^2 \rightarrow r | 2m^2, r | 2f^2$ and since $r | m - f$, r must be odd as $m - f$ is odd so $\gcd(r, 2) = 1$ and r is a common factor of m^2 and f^2 which is a contradiction as m and f are coprime.

But if $d = \gcd(m, f) > 1$ then

$$\mathfrak{d}(m' + f')p^2 = \mathfrak{d}(m' - f')(m'^2 + f'^2)$$

where $m = dm'$, $f = df'$. The argument will be almost the same as the above $m' - f' | p^2 \rightarrow r = \gcd(p, m' - f') > 1$ but then the final conclusion is different, $r | 2m^2$, $r | 2f^2 \rightarrow r | 2d^2m'^2$, $r | 2d^2f'^2$ and r avoids contradiction by the fact that it now divides d^2 .

I then go deeper into the rabbit hole by trying to find out more about d , starting from

$$\begin{aligned}(m + f)p^2 &= (m - f)(m^2 + f^2) \\ ((m - f) + 2f)p^2 &= (m - f)(m^2 + f^2), \quad m^2 + f^2 = Ap \\ 2f\cancel{p}p &= (m - f)\cancel{p}(A - p) \\ \rightarrow (m - f) &| 2fp \\ \cancel{d}(m' - f') &| 2\cancel{d}f'p\end{aligned}$$

again $\gcd(2, m' - f') = \gcd(f', m' - f') = 1 \rightarrow (m' - f') | p$, now from $r = \gcd(p, m' - f')$ (which also means $r | (m' - f')$) we can then express $p = rs$ where $\gcd(s, m' - f') = 1$ what we have is now

$$\begin{aligned}m' - f' &| p \\ m' - f' &| rs, \quad \gcd(m' - f', s) = 1 \\ \rightarrow \left. \begin{array}{l} m' - f' | r \\ r | m' - f' \end{array} \right\} &\Rightarrow r = m' - f'\end{aligned}$$

so what we know so far is that $r = (m' - f') | d^2$. But all this leads nowhere we are now really deep into the rabbit hole but we can still salvage this. My first not so good idea was to find another avenue to show that $\gcd(m, f)$ must be 1 somehow, but it was quite difficult.

I then tried to use the fact that $e > m > n > f$ and substituting $(m^2 + f^2 \pm p^2)/2p$ for e and n but it didn't lead anywhere either.

This is because the goal is infinite descent, but the above substitutions and manipulations were not going in this direction.

To get infinite descent we need to break down the problem into smaller ones and show that those smaller equations are of the same form but the above were just some random guesses without any clear direction although to be fair when solving problems sometimes you just don't know where to look for $\neg \setminus (\circ_o) / \neg$

One rather interesting (or not) little relationship I found in doing this was if $\gcd(p, h) = 1$ then $(p + h) \nmid ph$ and also $\gcd((p + h), ph) = 1$, we quite easily can see this, say $p = \prod p_i$, $h = \prod h_j$, if $(p + h)$ and ph are not coprime then $(p + h)$ will contain some of the p_i and/or h_j but if this is the case then say

$$p + h = p_i X$$

$$h = p_i(X - p/p_i)$$

$$p_i \mid h$$

but $\gcd(p, h) = 1$ which is a contradiction, the same can be said if $p + h$ contains some factor h_j of h (by the same token $(p + h) \nmid ph$).