MNRAS 000, 1–7 (2024) Preprint 7 June 2024 Compiled using MNRAS LATEX style file v3.3

Using Numerical Simulations to Model COVID-19 in Supermarkets

Dominika Kubanek - 32636156

7 June 2024

ABSTRACT
Understanding the effect of social distancing during the peak of the pandemic was a necessary step in developing precautions to keep people safe. We modelled the spread of COVID-19 in a supermarket as it gave restricted motion for people and meant there could, theoretically, be a moment in time where everyone is infected with COVID-19. Our objective of determining whether the 2m social distancing restrictions were necessary to keep people safe was achieved given the data we produced through our simulations.

Key words: Monte Carlo – Random Walks – Simulation

1 INTRODUCTION

Over the past five years we've heard a lot about COVID-19 and precautions necessary to control the spread of viral infections. While many people have spoken about the most efficient ways to protect oneself against infection, there have been many disputes regarding the need for any restrictions at all. Our focus was providing statistical evidence to support the need for restrictions enforced by governments around the world, especially how social distancing affected the spread of COVID-19.

In doing so we hoped to clarify the importance of following proposed restrictions and highlight the need for social distancing. Our aim was producing data which weighed the importance of masks vs social distancing and showed that they were not equal in defending from virus spread.

We decided to focus on how COVID-19 spreads around a supermarket, where people were modelled by a random walk simulation. Our model consisted of four scenarios where we went through no masks and no social distancing restrictions, masks with no social distancing restrictions, no masks with social distancing restrictions, and masks with social distancing restrictions. Our priority was developing a random walk where people maintained a 2m distance from each other to ensure we followed restrictions enforced during the peak of COVID-19.

Random walks were a fitting model for our simulations due to the unpredictability of human motion, especially in a supermarket setting where people would feel more anxious. If we instead chose to simulate the COVID-19 particles via a random walk, we would have allowed for them to travel indefinitely and would have likely faced errors due to that.

While we have heard frequently about the importance of wearing masks, people often dismissed social distancing as necessary as they believed masks were adequate barriers between COVID-19 particles and themselves. We wanted to test whether this assumption was right and whether social distancing was redundant.

In the rest of the report we focus on the supermarket model and methods used in it, as well as the results we found from the simulations. We focus on statistically analysing the data using the Mann-Whitney hypothesis test with a null hypothesis stating that 2m distancing had no effect on probability of infection and an alternative hypothesis of social distancing affecting the probability of infection.

2 METHODS

For our simulation of the spread of COVID-19 in a supermarket, we used random walks to move people. The COVID-19 particles were contained within a 2.1m radius of the people and did not have a random walk of their own. We ran the simulation 2000 times to get a probability distribution of steps for full infection and plotted the frequency against the number of steps into a histogram and fitted a normal distribution to the plot.

We decided to test on a group of 235 people which we calculated as an approximate value by taking an average supermarket size USDA (2016), subtracting the area by the amount of area taken up by obstacles (Figure 1) in our model ($4761m^2 - 1930m^2 = 2831m^2$) and dividing by 2 to accommodate for the 2m social distancing restrictions and finally divided by 6 HM (2019) to get a number of 235 people in the store at any given time. To check if this was a valid approximation we also cross-checked with what other premises did to ensure the spread of COVID-19 was contained, such as Walmart (2020), and found by using their method of 5 people per $1000ft^2$, we got a similar number ($4761m^2 \approx 50000ft^2$, when we multiplied this we got $50 * 5 = 250$) hence our approximation of 235 people was valid.

In our simulation we included four scenarios wherein the customers were unmasked with no social distancing regulations, masked with no social distancing regulations, unmasked with social distancing regulations, and masked with social distancing regulations to compare and provide statistical evidence to determine if the restrictions placed by the government were necessary. To do so we found probabilities of becoming contaminated upon contact with an infected person. For convenience and due to computational limitations we established the shop was filled with people all in the same age range of 41+.

As we had all customers in the same age range, it meant we could set a base probability of infection (P(I)) while unmasked to 0.5 Hamimes et al. (2023) for people over 41. We also found that wearing surgical masks decreased P(I) by 0.66 Andrejko et al. (2022) and that

Figure 1. The supermarket simulation set-up including all obstacles present. The total area taken up by the obstacles adds up to 1930m^2

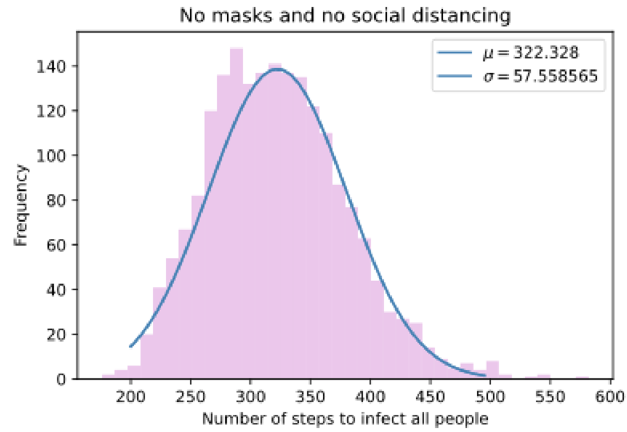

Figure 2. Histogram of steps required to infect all 235 people in the supermarket after coming in contact with a previously infected person. No masks or social distancing present, P(I) = 1.

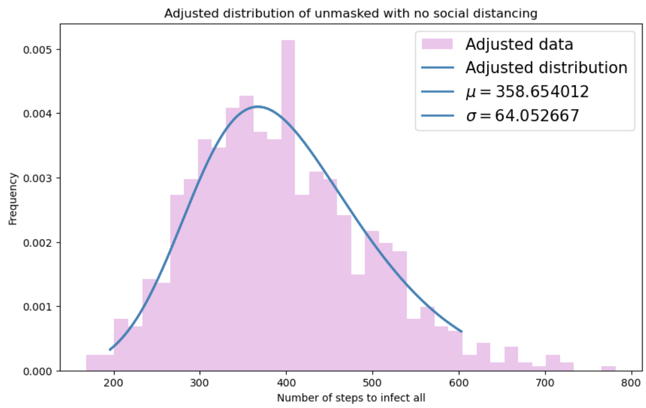

Figure 3. Histogram of steps required to infect all 235 people in the supermarket after coming in contact with a previously contaminated person. No masks worn and no social distancing enforced, P(I) = 0.5.

staying 2m apart as required during COVID-19 gave P(I) = 0.128 to catch the virus Chu et al. (2020). For each scenario we multiplied the relevant probabilities $P(A * B) = P(A) * P(B)$.

We assumed the supermarket had average ventilation, and that everyone in the shop had already entered. We initialised each person's position using a randomly generated coordinate, while ensuring they avoided the coordinates of obstacles by checking the randomly generated initial coordinate was not within the obstacle coordinates. We ensured people avoided obstacles by defining a function that measured the distance between the people and the objects and changing direction if the person was within the vertices of the objects. We used a similar idea for checking if people were within 2m for the 2m restriction simulations but checking for two people instead of people and objects. We called the object checking function in our shop simulation function which updated each persons movements for each step, while the 2m checker was implemented directly into the move function.

We also randomly selected one person to have COVID-19 and then passed the initial coordinates into the random walk for each person. For each step we tested whether uninfected people were stepping into infection range of the infected person (we specified the infection range to be 2m as we found that past 2m infection was decreased to less than 0.03 Chu et al. (2020)). Within the infection check we also specified a chance of infection for each scenario by randomly choosing a value between 0 and 1, if the value was below the specified probability, the person would become infected.

We plotted our histograms to check the frequency for the amount of steps it took for each person in the supermarket to be marked as infected and while the histograms were of a Gaussian nature, when we plotted them with a regular normal distribution fit, the plots were skewed to the side. To solve this issue, we fitted the probability distributions with a skewed Gaussian fit instead of a normal Gaussian fit, which meant to calculate the mean, μ and the standard deviation, σ we used

$$\delta = \frac{shape}{\sqrt{1 + shape^2}} \tag{1}$$

$$\mu = location + scale * \delta\sqrt{\frac{2}{\pi}} \tag{2}$$

and

$$\sigma = scale * \sqrt{1 - \frac{2 * \delta^2}{\pi}} \tag{3}$$

Equations 1, 2 and 3 Azzalini (2013) use the parameters location, scale and shape where location is the parameter which determines how the distribution shifts either to the left or the right, shape is the parameter of 'skewness' and scale is the parameter which determines how the data spreads; these parameters refer to those from the skewnorm function.

3 RESULTS

We included a test run for P(I) = 1 as seen in Figure 2 which showed a normal distribution with $\mu = 322$ and $\sigma = 58$. As normal distributions are symmetric around μ we used the empirical rule ($\mu \pm 3\sigma$) and found 99.7% of the data to have 148 as the lower boundary and 496 as the upper boundary. Reading off the histogram plotted in Figure 2 we found that the number of steps to infect all 235 people lied in that range, with outliers as expected indicating the 0.3% of data.

In each plot from Figures 3, 4, 5 and 6 we observed, as suggested from Figure 2, a Gaussian distribution, each shifted to the left horizontally indicating the probability distribution should be treated as a skew distribution. Figure 6 produced the smallest shift, while Figure 3 gave the largest. As we decreased the probability of infection, the shift became less notable.

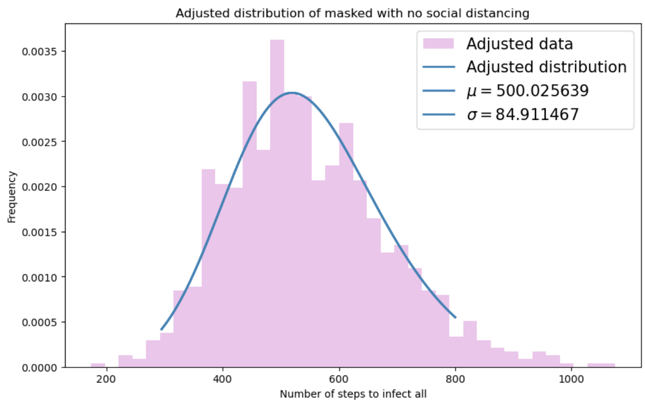

Figure 4. Histogram of steps required to infect all people in the supermarket after coming in contact with a previously infected person. No social distancing has been enforced, but masks worn by all 235, P(I) = 0.17

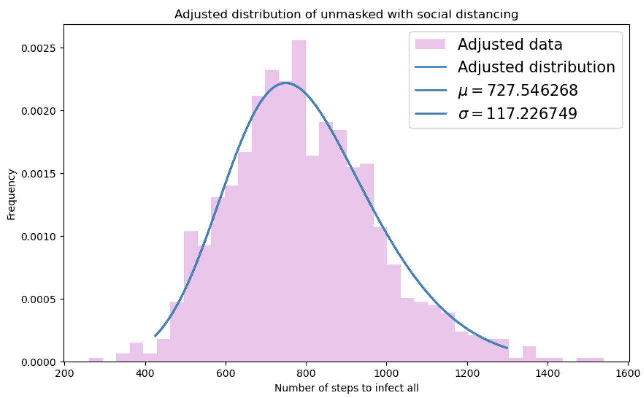

Figure 5. Histogram of steps required to infect all 235 people in the supermarket after coming in contact with a previously infected person. No masks worn but 2m social distancing enforced, P(I) = 0.064

Due to the skew Gaussian distribution in Figures 3 through 6, we had to use the cumulative distribution to find the range of the lower and upper boundaries, using the 0.0015 and 0.9985 percentiles respectively to determine where 99.7% of the data should lie. We verified this method by, on each of the plots, drawing vertical lines at each of the lower and upper bounds for the respective plots.

In Figure 3 we found the mean number of steps required to infect everyone to be $\mu = 359$, with $\sigma = 64$. We found 99.7% of the data was between 206 and 588. Using the method described above we confirmed that a majority of the data was in the range given by the percentiles.

Data from Figure 4 gave us $\mu = 500$ and $\sigma = 85$. We found the lower and upper bound of 99.7% of the data to be 284 and 794 respectively suggesting most of the data could be found in that range. As before, we plotted vertical lines and found there were outliers outside the range but the percentile data was accurate.

From Figure 5 we found $\mu = 728$ and $\sigma = 117$. We found 99.7% of data lied between 438 and 1140 steps. Plotting vertical lines to check the reliability of the method once again showed outliers indicating the 0.03% of data hence proving it reliable.

With data from Figure 6 we found $\mu = 1274$ and $\sigma = 199$. We used the upper and lower percentile to find 798 and 1985 as the bounds

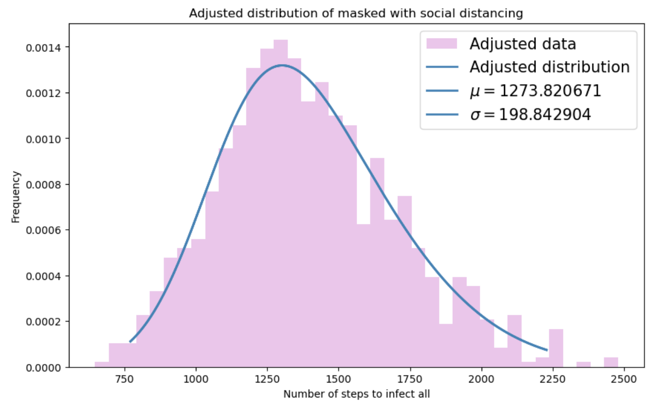

Figure 6. Histogram of steps required to infect all 235 people after coming in contact with a previously infected person. Masks and 2m distancing enforced, P(I) = 0.02176

and once again checked the vertical lines against the data to see if most of the data lied within that range. As previously, we concluded that 99.7% of the data was in that range with the exception of outliers indicating the 0.03% remainder of data.

With data now calculated, we made our null hypothesis, H_0, that keeping 2m social distancing had no effect on the probability of becoming infected. We then tested, following the Mann-Whitney U-test, each of the scenarios against the unmasked P(I) = 0.5 scenario. Our H_1 was that social distancing did affect infection rates. As our test had a larger sample size than 20 (for each case) Hessing (2019) we used μ to test our hypothesis.

Testing unmasked with no distancing against unmasked with social distancing we found that $\mu_{UM}(359) \neq \mu_{UM2m}(728)$. When we tested unmasked against masked, both with no social distancing, we found $\mu_{UM}(359) \neq \mu_{M}(500)$. Our test of masked with no social distancing against masked with social distancing produced $\mu_{M}(500) \neq \mu_{M2m}(1274)$. Finally, for unmasked against masked with 2m distancing we found $\mu_{UM}(359) \neq \mu_{M2m}(1274)$. With all tests compared, we determined it was necessary to discard H_0 and hence there was a correlation between keeping a 2m distance and the decrease in infection probability.

We calculated the percentage increase in steps between each scenario using $\frac{|A-B|}{A} * 100\%$. Going from unmasked to masked where the unmasked mean was A and the masked mean was B, we calculated the percentage increase for mean number of steps to infect all people to be 39.3%; for unmasked mean number of steps (A) to unmasked with social distancing (B) we got an increase of 102.9%; for masked (A) to masked with social distancing (B) we got an increase of 154.8%. Finally, we found the largest increase from unmasked (A) to masked with social distancing (B), at an increase of average number of steps at 254.9%.

4 DISCUSSION

As presented in section 3 we note that wearing masks gave a definite increase in number of steps required to infect everyone in the supermarket. Likewise, keeping 2m apart provided an even larger increase in number of steps required to infect everyone. We read off Figures 5 and 6 that there was a considerable difference between the mean number of steps, with 728 and 1274 respectively. From Figure 4 we saw that the mean value was 500. This suggested that social distancing played a greater part in reducing infection rates during the peak of COVID-19, as the difference between unmasked with social distancing vs unmasked without social distancing was greater than the difference between masked vs unmasked.

In section 3 we found the percentage increase for the mean number of steps between the scenarios, which emphasized how significant both masks and social distancing were to decrease infection rates. While masks caused an increase in number of steps by 39.3% our interpretation that social distancing played a greater role was once again supported by the 102.9% increase between unmasked with and without social distancing.

Unfortunately our model was not representative of the entire population as we limited our simulations to people over 41 for convenience. This meant if people were immunosuppressed due to health conditions or age, they would not be fit for following these guidelines as they may have needed more intensive precautions to ensure their safety. Similarly, for those who had a better immune system and were younger they may have been safe with less restrictions. However, despite the discrepancies we could argue the restrictions were logical as, assuming there is a probability we can use for each specific case, social distancing would still cause a decrease in infection probability overall, as would wearing masks.

Another limitation of our simulations was not taking into account vaccinations. While the vaccination wouldn't make P(I) = 0, it should ideally decrease P(I) by a significant amount. With this in mind, whether people were vaccinated or not wouldn't have been necessarily more useful as we wanted to test whether 2m distancing was necessary. With the addition of vaccinations, we would have likely seen the same results.

While we didn't test on a larger sample size, it would have been appropriate to do so in order to test whether increasing the number of people in a shop at any given time would have caused much of a difference to infection rates. If we were to guess based on our current data, having more people in the same sized building would have given less space for people to keep 2m distancing and hence would have likely caused a decreased average number of steps required to infect everyone in the shop. An additional development to our simulations would be in the form of modelling the diffusion of the COVID-19 particles as this would have given us a better understanding of why COVID-19 particles are less likely to reach someone past the 2m radius.

With all this in mind, our model did confirm that the most successful way to decrease chances of catching COVID-19 was to stay 2m apart while wearing a surgical mask, as seen through the 254.9% increase in μ between unmasked and masked with 2m distancing restrictions.

5 CONCLUSIONS

Our model for COVID-19 spread in a supermarket was suitable for analysis using a skew Gaussian distribution as it showed Gaussian curvature with a shift along the horizontal axis. This allowed for us to test our hypothesis with the help of the Mann-Whitney U-test, where H_0 stated that social distancing was not effective (H_1 stated that social distancing was effective). By running our simulations we came to the conclusion that while wearing masks did protect people from catching COVID-19, it did not affect the average number of steps as much as social distancing on its own did. While masks helped increase the average number of steps by 39.3%, we found social distancing to increase the average by 102.9%, which is much more effective and so we concluded that we could discard H_0. Furthermore, we found that wearing a mask while maintaining a 2m distance increased the number of steps by 254.9% which was more than double the increase for social distancing alone. While we didn't account for the fact that not everyone would be susceptible to infection to the same degree as others, we tried our best to generalise the model to resolve this problem by limiting our model to focus on a small age range. Even so, from our data we determined that social distancing was important in containing the spread of COVID-19 and would have been a significant contribution to the decrease of COVID-19 infection rates.

REFERENCES

Alderliesten R., 2021, Deflection by method of double integration, https://eng.libretexts.org/Bookshelves/Mechanical_Engineering/Introduction_to_Aerospace_Structures_and_Materials_(Alderliesten)/02%3A_Analysis_of_Statically_Determinate_Structures/07%3A_Deflection_of_Beams-_Geometric_Methods/7.03%3A_Deflection_by_Method_of_Double_Integration#:~:text=This%20method%20entails%20obtaining%20the,second%20integration%20gives%20the%20deflection.

Andrejko K. L., et al., 2022, MMWR. Morbidity and Mortality Weekly Report, 71, 212–216

Anisiu M. C., 2013, https://www.math.ubbcluj.ro/~didactica/pdfs/2013/didmath2013-06.pdf

Azzalini A., 2013. Cambridge University Press, p. 24–56, doi:10.1017/cbo9781139248891.003, http://dx.doi.org/10.1017/CBO9781139248891.003

Budd C., 2018, The mathematics of climate change, https://www.gresham.ac.uk/watch-now/mathematics-climate-change

Chang H., Ittner C. D., Paz M. T., 2014, Journal of Management Accounting Research, 26, 1–32

Chowdhury M., Ghosh I., Aznam S. M., Mawa S., 2021, Journal of Low Frequency Noise, Vibration and Active Control, 40, 1731–1743

Chu D. K., et al., 2020, The Lancet, 395, 1973–1987

Cohen J., 1986, Introduction to fourier transform spectroscopy. U.S. Dept. of Commerce, National Bureau of Standards; National Technical Information Service, distributor

Dangermond J., 2024, What is GIS?, https://www.esriuk.com/en-gb/what-is-gis/overview#:~:text=GIS%20connects%20data%20to%20a,science%20and%20almost%20every%20industry.

Feynman R. P., Hibbs A. R., 1975, Quantum mechanics and path integrals by R.P. Feynman, A.R. Hibbs. McGraw-Hill

Gallagher T. A., Nemeth A. J., Hacein-Bey L., 2008, American Journal of Roentgenology, 190, 1396–1405

Giacinti M., Mariani F., Recchioni M. C., Zirilli F., 2013, Intelligent Control and Automation, 04, 250–262

HM 2019, The Building Regulations 2010, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/937932/ADB_Vol2_Buildings_other_than_dwellings_2019_edition_inc_2020_amendments.pdf

Hamimes A., et al., 2023, Healthcare, 11, 1257

Hessing T., 2019, Mann-Whitney Hypothesis Test, https://sixsigmastudyguide.com/mann-whitney-non-parametric-hypothesis-test/

Hielscher A. H., Alcouffe R. E., Barbour R. L., 1998, Physics in Medicine and Biology, 43, 1285–1302

Lutter M., Peters J., 2020, Differential Equations as a Model Prior for Deep Learning and its Applications in Robotics, https://openreview.net/forum?id=_uPd3skTsj

Morris G. A., Freeman R., 1978, Journal of Magnetic Resonance (1969), 29, 433–462

Sinha D., Oct 27, 2022, How does the Shazam app recognize music? An explanation of music recognition technology used in

the Shazam app, https://www.techaheadcorp.com/blog/decoding-shazam-how-does-music-recognition-work-with-shazam-app/
Su L., Liu G., Guo Y., Zhang X., Zhu X., Wang J., 2022, Frontiers in Genetics, 13
USDA 2016, Global Agricultural Information Network
Valluri S., Dergachev V., Zhang X., Chishtie F., 2021, Physical Review D, 104
Walmart 2020, “Changes to Our Shopping Process to Encourage Social Distancing.”, https://corporate.walmart.com/news/2020/04/03/changes-to-our-shopping-process-to-encourage-social-distancing
Yang F., Rahmat-Samii Y., 2003, IEEE Transactions on Antennas and Propagation, 51, 2936
Černý A., 2004, The Journal of Derivatives, 12, 73–88

6 ADDENDUM

6.1 Integration

Integration is the process of finding the area, volume, displacement, etc. of a function by splitting it into a finite number of rectangles. The more accurate a method is, the smaller the rectangles are, until the areas become infinitely small. We can use integration in many ways.

6.1.1 In Academia

An important field of research which uses integration is via cancer research, where we use integration to analyse protein-protein interaction networks to help aid our understanding of identifying new disease-associated cells and find their importance in the formation of diseases (such as cancer) Su et al. (2022). This helps identify any potential genes which may increase the risk of cancer.

In electromagnetism, we can use integration to solve Maxwell's equations and hence compute the electric and magnetic fields Yang & Rahmat-Samii (2003). This allows us to find the coupling design of micro-strip antennas via the EM bandgap.

Another application of integration in academia can be seen in high energy physics through quantum mechanics Feynman & Hibbs (1975). Gauss quadrature can be used to compute expectation values of operators to find the probability of finding a quantum particle. We can also use integration to evaluate matrix elements in wavefunction calculations of the observable elements.

6.1.2 Outside Academia

In aerospace engineering, we see integration methods used to evaluate beam and plate bending problems. By using double integrals we first find the slope, then the deflection of the elastic curve of the beam Alderliesten (2021). By implementing the boundary conditions we obtain the solution to the problem and hence can design according to safety protocols.

In geographic information systems, integration can compute mean elevation or land cover composition from satellite imagery. This information is then processed and mapped and can be later used in other aspects of science to analyse specific parts of the data Dangermond (2024).

Another application of integration outside of academia would be in the financial market via economic modelling Chang et al. (2014). We can approximate production functions in economic models by applying integration methods which allows us to interpret how different external factors such as market growth, firm size, decentralization, etc. affect the market.

6.2 Differential Equations

There are different methods for differential equations used in numerical simulations; the most widely used method for ordinary differential equations being Runge-Kutta 4 (RK4) while for partial differential equations it is much more convenient to use Finite Differencing (FD). RK4 performs many evaluations of a function on a first order ODE to provide approximate solutions to it. RK4 is the most efficient and accurate iteration of the Runge-Kutta method. FD splits a function into a finite number of elements to obtain an approximate solution of the PDE by linearly combining the element solutions.

6.2.1 In Academia

One application of the RK4 method is through analysing dynamics of celestial objects. One of the simplest systems to solve via this method is the "three body problem". We use differential equations to compute the motion of the three bodies rotating around each other by choosing a rotation coordinate axis Anisiu (2013).

FD can be used in analysing the diffusion of particles, particularly in three-dimensional diffusion rates in homogeneous biological tissues Hielscher et al. (1998). The FD method is used continuously to approximate solutions, and the more the simulation is run, the closer the approximation is to the real solution.

Another application of RK4 in academia can be in chemical kinetics systems Chowdhury et al. (2021). In chemistry, nonlinear reactions are modelled by three-dimensional, coupled ODEs which oscillate with a measurable frequency. This means we can find the equation of motion for such systems and hence they are fit to be solved with RK4 method.

6.2.2 Outside Academia

A fun application of differential equations can be seen in computer game animation. In game animation, we can use DEs to model player motion in a realistic and physically accurate way, so the game includes friction, drag and various other forces Giacinti et al. (2013). By doing so, we can get a smoothly running game with accurate fluid physics which is translated into the graphics.

Another important application of DEs is in climate modelling. The model for weather forecasting uses PDEs for the fluid motion Budd (2018) which allows us to use numerical methods to solve the equations and predict future climate, or even climate change in general. The climate modelling with DEs can be even extended to the rise in water levels and how that will affect different countries.

Another way DEs can be used outside academia can be seen in robotics. As in the previous examples of differential equation applications, we start with an equation of motion for the robot limbs and by solving the equations using numerical methods, we can control the motion of a robot Lutter & Peters (2020).

6.3 Fourier Transforms

Fourier transforms (FT) take signals from different domains such as space or time and transform them into the frequency domain. A FT is a superposition of sine waves which processes enough cycles of the original signal to provide a complete 'cycle'.

6.3.1 In Academia

FTs are used in modulating distorted gravitational wave signal data (often distorted due to Doppler shifts etc.) in order to filter out repeated measurements which produce sharp peaks in the frequency Valluri et al. (2021). The FT then produces data which we can read and even analytically solve without large uncertainties.

Another way we can use FTs in academic research is through nuclear magnetic resonance Morris & Freeman (1978). We use FTs to analyse "frequency-selective excitation" at different resolutions, which we can do due to magnetic fields being described by sine waves. Solving nuclear magnetic resonance via FTs provides us with conclusions regarding frequency and their corresponding excitation which can lead to applications in different parts of science such as chemical exchange.

One more application of FT in academia would be through spectroscopy. In this technique, we measure the coherence of electromagnetic radiation from a source in either the time or space domain (depending on convenience) and process these signals into the frequency domain in order to analyse them Cohen (1986). By putting the signals through a FT we can find the spectroscopy of specific sources such as infrared light, chemicals, etc.

6.3.2 Outside Academia

A good example of how FTs are used in our daily lives is through phone apps such as Shazam Sinha (2022). The app has thousands of song data stored in its code which act as "fingerprints". It takes the music being played and runs the frequency through a fast FT which should, ideally, correspond to a fingerprint in its database. The more background noise there is, the harder the programme has to work to find a corresponding fingerprint by filtering out different frequencies.

Another application of FT would be through its use in the financial world. Also using fast FTs, we can make models and find real-time pricing Černý (2004). These models also take into account consistent outliers and the instability of the asset price volatility.

As a final example of FTs outside of academia we have image processing in medical environments. FTs can collect data from MRI scans Gallagher et al. (2008) and filter out all background and noise interference to reconstruct and analyse images. This is useful as it allows professionals to manipulate the data and reduce frequencies as needed.

This paper has been typeset from a TEX/LATEX file prepared by the author.